

TECNOLOGICO NACIONAL DE MEXICO

Instituto Tecnológico de Cd. Cuauhtémoc

PERIODO SABATICO

Enero a julio del 2025

PROGRAMA ACADEMICO

Elaboración de apuntes de la asignatura
Probabilidad y Estadística del plan de estudio
Industrias Alimentarias

ELABORADO POR
MC Mariano Castillo Loya

Cd. Cuauhtémoc, Chih. Julio de 2025



Contenido

1. Teoría de conjuntos.....	1
1.1. Definición, notación y operaciones con conjuntos.....	1
1.2. Leyes y representación de Diagramas de Venn.	7
1.3. Análisis combinatorio Principio fundamental de conteo (diagramas de árbol).	20
1.4. Permutaciones (simples, circulares y con repetición).	23
1.5. Combinaciones y Teorema del Binomio.	30
1.6 Uso de software.....	33
2. Variables Aleatorias Discretas y sus distribuciones de Probabilidad.	34
2.1. Variable aleatoria discreta.	34
2.2. Función de probabilidad y de distribución.	35
2.2.1. Distribución Binomial.....	36
2.2.2. Distribución Hipergeométrica.	43
2.2.3. Distribución Geométrica.....	49
2.2.4. Distribución Multinomial.....	54
2.2.5. Distribución de Poisson.	57
3. Variables Aleatorias Continuas y sus distribuciones de Probabilidad.	65
3.1. Variable aleatoria continua.	65
3.2. Función de densidad y acumulativa.	66
3.3. Valor esperado.	68
3.4. Distribuciones uniforme y exponencial.	68
3.5. Distribución normal.....	79
3.5.1. Aproximación de la Distribución binomial a la normal.	91
3.6 Distribución t de student y Chi cuadrada.....	94
4. Estadística Descriptiva e Inferencial.	97
4.1. Datos no agrupados.	97
4.2. Datos agrupados.	99
4.2.1. Tablas de distribución de frecuencias y gráficas.....	100
4.2.2. Medidas de tendencia central: Medidas de dispersión Cuartiles, Deciles y Percentiles.....	112

4.3. Técnicas de Muestreo.....	120
4.4. Distribuciones muestrales.....	136
4.5. Estimación.	138
4.6. Pruebas de Hipótesis.....	179
Bibliografía consultada	247

1. Teoría de conjuntos.

1.1. Definición, notación y operaciones con conjuntos.

Definición de conjuntos.

La noción de conjunto es uno de los conceptos fundamentales en las matemáticas modernas. Un conjunto se define como una colección bien definida de objetos, llamados elementos, que comparten una característica en común o se agrupan según un criterio específico. Esta idea ha evolucionado con el tiempo, desde la intuición básica de colecciones en la vida cotidiana hasta un marco riguroso en la teoría de conjuntos moderna, base fundamental para muchas ramas matemáticas, incluida la lógica y la teoría de números (Moreno, 2022).

La definición formal de conjunto se atribuye a Georg Cantor, quien a finales del siglo XIX estableció las bases de la teoría de conjuntos. Cantor definió un conjunto como cualquier colección de objetos bien definidos, lo que significa que existe un criterio claro para determinar si un elemento pertenece o no al conjunto (Gómez & Ortega, 2021). Desde entonces, los matemáticos han explorado y refinado esta definición, lo que ha permitido la formalización de conceptos abstractos, como conjuntos vacíos, infinitos, y subconjuntos.

La definición y la teoría de conjuntos no solo son fundamentales en la matemática pura, sino que también tienen aplicaciones en la informática, la lógica y las ciencias cognitivas. A partir de la noción básica de conjunto, es posible construir conceptos como funciones, relaciones, y estructuras algebraicas, lo que convierte a los conjuntos en un pilar esencial de la matemática contemporánea.

Los dos métodos principales para expresar conjuntos son:

Notación por extensión: Enumerar los elementos del conjunto. Por ejemplo, $A = \{1,2,3,4\}$. Este método es útil para conjuntos pequeños y finitos.

Notación por comprensión: Describir los elementos a través de propiedades. Por ejemplo, $B = \{\text{Alumnos del Instituto Tecnológico de Cd. Cuauhtémoc}\}$, $B = \{x \text{ dado que } x \text{ es un número primo menor que } 100\}$, $C = \{x \text{ dado que } x \text{ es un país del mundo}\}$. Como afirma Kleiman, *"la notación por comprensión permite abstraer patrones y generalizar propiedades que serían difíciles de visualizar en conjuntos extensos o infinitos"*

A continuación, se presenta la notación utilizada en conjuntos, con ejercicios donde se aplica y explica cada uno de ellos.

Pertenencia $\in$. Como la palabra lo dice, consiste en determinar si un elemento dado pertenece o no a un cierto conjunto.

Dado el conjunto $A = \{2, 4, 6, 8\}$, determine si las siguientes expresiones son falsas o verdaderas.

- a) $2 \in A$. Como el 2 si pertenece al conjunto A, se tiene que la expresión es verdadera.
- b) $5 \in A$. Como el 5 no pertenece al conjunto A, se tiene que la expresión es falsa.
- c) $8 \in A$. Como el 8 si pertenece al conjunto A, se tiene que la expresión es verdadera
- d) $10 \in A$ Como el 10 no pertenece al conjunto A, se tiene que la expresión es falsa.

No pertenencia $\notin$. En este caso corresponde a decir si un elemento no pertenece al conjunto dado, tal como se ilustra a continuación.

Dado el conjunto $B = \{a, e, i, o, u\}$, determine si las siguientes expresiones son falsas o verdaderas.

- a) $x \notin B$. Como x no pertenece al conjunto B, se tiene que la expresión es verdadera.
- b) $1 \notin B$. Como él no pertenece al conjunto B, se tiene que la expresión es verdadera.
- c) $e \notin B$. Como la e si pertenece al conjunto B, se tiene que la expresión es falsa.
- d) $o \notin B$. Como la o si pertenece al conjunto B, se tiene que la expresión es falsa.

Subconjunto $\subseteq$. Cuando los elementos de un conjunto A se encuentran dentro de los elementos de un conjunto B, se dice que A es subconjunto de B y se expresa $A \subseteq B$

Considerando que $B = \{2, 4, 6, 8\}$, determine si las siguientes expresiones son falsas o verdaderas.

- a) Si $A = \{2, 6, 8\}$, entonces $A \subseteq B$. Esta expresión es verdadera, dado que los elementos de A si están contenidos en B.
- b) Si $C = \{4, 6, 8\}$, entonces $C \subseteq B$. Esta expresión es verdadera, dado que los elementos de C si están contenidos en B.
- c) Si $A = \{a, 6, 8\}$, entonces $A \subseteq B$. Esta expresión es falsa, dado que no todos los elementos de A están contenidos en B.
- d) Si $D = \{7, 6, 8, 9\}$, entonces $D \subseteq B$. Esta expresión es falsa, dado que no todos los elementos de D están contenidos en B.

No es un subconjunto $\not\subseteq$. Cuando los elementos de un conjunto A no se encuentran dentro de los elementos de un conjunto B, se dice que A no es subconjunto de B y se expresa $A \not\subseteq B$.

Considerando que $B = \{2, 4, 6, 8\}$, determine si las siguientes expresiones son falsas o verdaderas.

- a) Si $A = \{2, 6, 8\}$, entonces $A \not\subseteq B$. En esta expresión de está diciendo que A no es subconjunto de B, lo cual es falso.
- b) Si $C = \{4, 6, 8\}$, entonces $C \not\subseteq B$. En esta expresión de está diciendo que B no es subconjunto de B, lo cual es falso.

- c) Si $A = \{a, 6, 8\}$, entonces $A \not\subseteq B$. Esta expresión es verdadera, dado que no todos los elementos de A están contenidos en B.
- d) Si $D = \{7, 6, 8, 9\}$, entonces $D \not\subseteq B$. Esta expresión es verdadera, dado que no todos los elementos de D están contenidos en B.

Conjunto vacío $\emptyset$. Es cuando el conjunto carece de elementos, lo cual puede suceder en las operaciones de conjuntos, tal como se verá más adelante. Una expresión correspondiente a conjunto vacío, es que no existe un número que sea número natural y número negativo a la vez, lo anterior es cierto, ya que los números naturales excluyen a los valores negativos, los fraccionarios o decimales.

Conjunto Universo Ω . Corresponde a todos los elementos o resultados posibles de un evento, por ejemplo, el resultado del lanzamiento de un dado, lo cual viene a ser $U = \{1, 2, 3, 4, 5, 6\}$.

De los párrafos anteriores se puede apreciar que los conjuntos se simbolizan con letras mayúsculas (A, B, C) y que los elementos se presentan entre llaves { }

En relación a las operaciones de conjuntos, se tiene que son: unión, intersección, complemento y diferencia, cada una de ella se abordan a continuación, con su definición y un ejemplo.

Unión de conjuntos $\cup$. La unión de los conjuntos A y B corresponde al conjunto formado por los elementos de A y B sin repetirlos, tal como se muestra a continuación.

Si $A = \{1, 2, 3, 4, 5\}$ y $B = \{4, 5, 6, 7, 8\}$, entonces se tiene que $A \cup B = \{1, 2, 3, 4, 5, 6, 7, 8\}$.

Intersección de conjuntos $\cap$. La intersección de los conjuntos A y B corresponde al conjunto formado por los elementos comunes de A y B, tal como se muestra a continuación.

Si $A = \{1, 2, 3, 4, 5\}$ y $B = \{4, 5, 6, 7, 8\}$, entonces se tiene que $A \cap B = \{4, 5\}$.

Complemento A^c . Para el caso específico mencionado de A complemento, se tiene que son los elementos que le faltan al conjunto A para ser el conjunto Universo, ejemplo de ello se muestra a continuación.

Si $\Omega = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$ y $A = \{1, 2, 3, 4, 5\}$, entonces se tiene que $A^c = \{6, 7, 8, 9\}$.

Diferencia de conjuntos $A - B$. Para este caso en específico, corresponde a los elementos del conjunto A que no pertenecen al conjunto B, lo anterior se ilustra a continuación.

Si $A = \{1, 2, 3, 4, 5\}$ y $B = \{4, 5, 6, 7, 8\}$, entonces se tiene que $A - B = \{1, 2, 3\}$.

Por otro lado, se tiene o aprecia que $A - B \neq B - A$, lo cual es cierto, ya que $B - A = \{6, 7, 8\}$

Una vez que se ha presentado la notación básica de conjuntos y sus operaciones, se puede resolver el siguiente ejercicio.

Considerando que:

$$\Omega = \{1, 2, 3, 4, 5, a, e, i, o, u\}$$

$$A = \{1, 2, 3, a, u\}$$

$$B = \{1, 2, 4, 5, a, o\}$$

$$C = \{1, a, e, i, o, u\}$$

Realice las siguientes operaciones de conjuntos:

- a) $A \cup C$
- b) $B \cap C$
- c) $A \cup B \cup C$
- d) $A \cap B \cap C$
- e) A^c
- f) $A - C$
- g) $B - A$
- h) $(B \cup C) \cap A$
- i) $(A \cup B \cup C)^c$
- j) $(C - A)^c$

Solución.

$$a) A \cup C = \{1, 2, 3, a, e, i, o, u\}$$

La solución corresponde al conjunto de los elementos del conjunto A y los elementos del conjunto C, sin repetirlos.

$$b) B \cap C = \{1, a, o\}$$

La solución corresponde a los elementos comunes entre los conjuntos B y C.

$$c) A \cup B \cup C = \{1, 2, 3, 4, 5, a, e, i, o, u\}$$

La solución corresponde al conjunto de los elementos del conjunto A, los elementos del conjunto B y los elementos del conjunto C, sin repetirlos.

$$d) A \cap B \cap C = \{1, a\}$$

La solución corresponde a los elementos comunes en los conjuntos A, B y C.

$$e) A^c = \{4, 5, e, i, o\}$$

La solución corresponde a los elementos que le faltan al conjunto A para ser igual al conjunto universo.

f) $A - C = \{2, 3\}$

La solución corresponde a los elementos del conjunto A que no corresponden al conjunto C

g) $B - A = \{4, 5, o\}$

La solución corresponde a los elementos del conjunto B que no corresponden al conjunto A

h) $(B \cup C) \cap A = \{1, 2, a, u\}$

Para la solución de este inciso es conveniente establecer que $(B \cup C) = \{1, 2, 4, 5, a, e, i, o, u\}$ y luego establecer que elementos tiene en común con el conjunto A

i) $(A \cup B \cup C)^c = \text{conjunto vacío} = \emptyset$

Como la unión de los conjuntos A, B y C dan como resultado todos los elementos del conjunto universo, el complemento (elementos que faltan para ser universo) resulta ser conjunto vacío.

j) $(C - A)^c = \{1, 2, 3, 4, 5, a, u\}$

Para la solución de este inciso es conveniente establecer que $(C - A) = \{e, i, o, u\}$ (elementos de C que no pertenecen a A) y luego establecer que elementos faltan para ser conjunto universo.

Ejercicios propuestos.

Ejercicio 1. Considerando que:

$$\Omega = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$$

$$A = \{1, 3, 5, 7, 9\}$$

$$B = \{0, 2, 4, 6, 8\}$$

$$C = \{2, 3, 4, 5, 6\}$$

Realice las siguientes operaciones de conjuntos:

a) $A \cap C$

b) $B \cup C$

c) $(A \cap B) \cup C$

d) $A \cap B \cap C$

e) C^c

f) $(A \cap C)^c$

- g) $B - A$
- h) $(B \cup C) \cap A$
- i) $(A \cup B \cup C)^c$
- j) $(A - B)^c$

Ejercicio 2. Considerando que:

$$\Omega = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, x, y, z\}$$

$$A = \{0, 3, 5, 7, 9, x, y, z\}$$

$$B = \{0, 1, 2, 4, 6, 8, x\}$$

$$C = \{4, 5, 6, 7, 8, 9, x, y\}$$

Realice las siguientes operaciones de conjuntos:

- a) $B \cap A$
- b) $A \cup C$
- c) $(A \cup B) \cap C$
- d) $A \cap B \cap C$
- e) A^c
- f) $(A \cap C)^c$
- g) $C - A$
- h) $(B \cup C) \cap A$
- i) $(A \cap B \cap C)^c$
- j) $(A - C)^c$

Ejercicio 3. Considerando que:

$$\Omega = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, a, b, c, d\}$$

$$A = \{a, b, c, 3, 5, 7, 9\}$$

$$B = \{d, 2, 4, 6, 8\}$$

$$C = \{a, b, c, d, 2, 3, 4, 5, 6\}$$

Realice las siguientes operaciones de conjuntos:

- a) $B \cap C$
- b) $B \cup C$
- c) $(A \cup B) \cap C$
- d) $A \cap B \cap C$

- e) B^c
- f) $(A \cup C)^c$
- g) $B - A$
- h) $(B \cup C) \cap A$
- i) $(A \cup B \cup C)^c$
- j) $(B - C)^c$

1.2. Leyes y representación de Diagramas de Venn.

Las leyes de conjuntos son reglas fundamentales de la teoría de conjuntos que describen cómo operar y relacionar conjuntos. Estas leyes son similares en forma a las de la lógica proposicional, ya que ambos campos comparten estructuras algebraicas similares.

Principales leyes de conjuntos

1.- Leyes de identidad

$A \cup \emptyset = A$ (La unión con el conjunto vacío no cambia el conjunto).

$A \cap \Omega = A$ (La intersección con el universo no cambia el conjunto).

2.- Leyes de idempotencia

$$A \cup A = A$$

$$A \cap A = A$$

3.- Leyes de conmutación

$$A \cup B = B \cup A$$

$$A \cap B = B \cap A$$

4.- Leyes de asociación

$$(A \cup B) \cup C = A \cup (B \cup C)$$

$$(A \cap B) \cap C = A \cap (B \cap C)$$

5.- Leyes distributivas

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

6.- Leyes de absorción

$$A \cup (A \cap B) = A$$

$$A \cap (A \cup B) = A$$

7.- Leyes de complemento

$$A \cup A^c = \Omega$$

$$A \cap A^c = \emptyset$$

8.- Leyes de involución

$$(A^c)^c = A$$

9.- Leyes de De Morgan

$$(A \cup B)^c = A^c \cap B^c$$

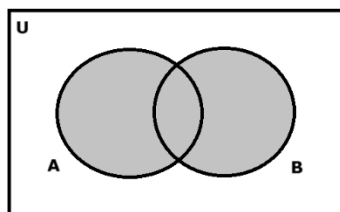
$$(A \cap B)^c = A^c \cup B^c$$

La utilidad de las leyes de conjuntos consiste en que permite la simplificación de expresiones, que se traduce a reducir expresiones complicadas en términos de conjuntos y ello facilita la solución de problemas.

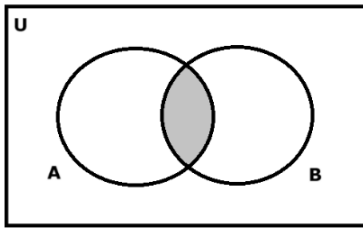
A continuación, se abordará el tema de diagramas de Venn, con la intención de posteriormente conjugarlo con las operaciones básicas de conjuntos y sus leyes.

Los diagramas de Venn son representaciones gráficas ampliamente utilizadas en matemáticas, lógica y otras disciplinas para visualizar la relación entre conjuntos. Introducidos por el lógico y filósofo británico John Venn en 1880, estos diagramas han trascendido su origen académico para convertirse en una herramienta versátil en la resolución de problemas, el análisis de datos y la enseñanza.

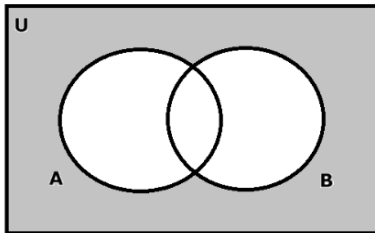
Un diagrama de Venn consiste típicamente en círculos que representan conjuntos, donde las áreas de superposición indican elementos compartidos entre ellos. Su simplicidad visual permite resumir relaciones complejas entre múltiples grupos en un formato claro y accesible. Los conceptos básicos de unión, intersección, diferencia y complemento se pueden ilustrar de forma directa con estos diagramas. Por ejemplo, la unión de dos conjuntos ($A \cup B$) corresponde al área total cubierta por los círculos de A y B, mientras que su intersección ($A \cap B$) se limita a la zona donde se solapan, empalman o traslapan. A continuación, se presentan diagramas de Venn para cada una de las operaciones básicas de conjuntos



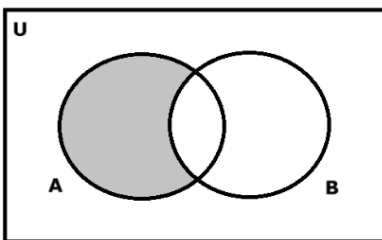
En este diagrama de Venn se puede apreciar en color gris que $A \cup B$ (A unión con B) corresponde a los elementos de A y B sin repetirlos.



En este diagrama de Venn se puede apreciar en color gris que $A \cap B$ (A intersección con B) corresponde a los elementos comunes entre A y B.



En este diagrama de Venn se puede apreciar en color gris que $(A \cap B)^c$ (A unión con B complemento) corresponde a los elementos que le faltan a la A unión con B para ser universo.



En este diagrama de Venn se puede apreciar en color gris que $A - B$ (A diferencia de B) corresponde a los elementos de A unión que no le pertenecen a B.

Uso de los diagramas de Venn.

Considere que, a un grupo de 15 personas (las cuales por razones prácticas fueron numeradas del 1 al 15), se les hizo una entrevista para conocer acerca de los deportes que les gusta practicar, teniendo así los siguientes datos:

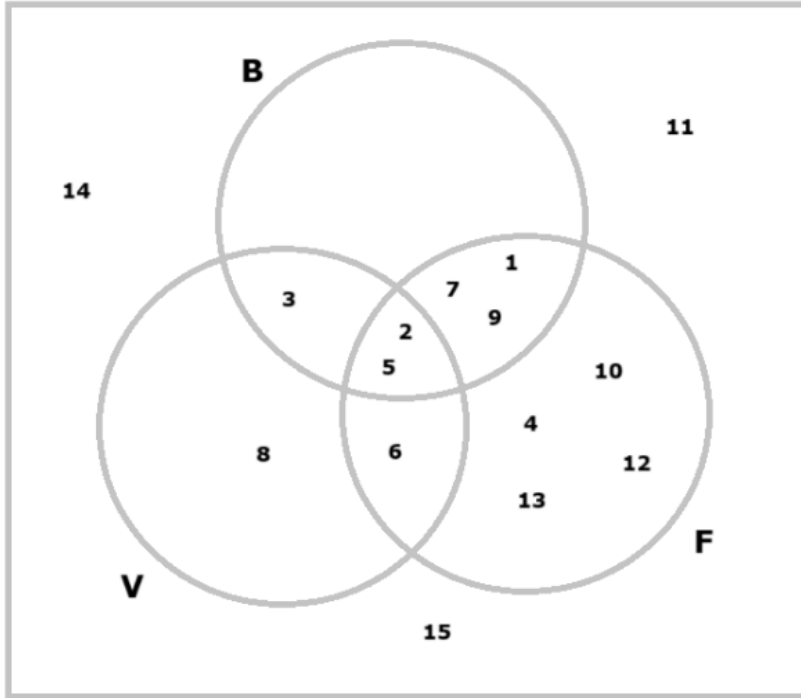
$$\Omega = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15\}$$

$$\text{Personas que les gusta el beisbol } B = \{1, 2, 3, 5, 7, 9\}$$

Personas que les gusta el futbol $F = \{1, 2, 4, 5, 6, 7, 9, 10, 12, 13\}$

Personas que les gusta el voleibol $V = \{2, 3, 5, 6, 8\}$

La representación mediante diagrama de Venn, sería la siguiente:



Mediante este diagrama de Venn se puede apreciar con facilidad aspectos como los siguientes:

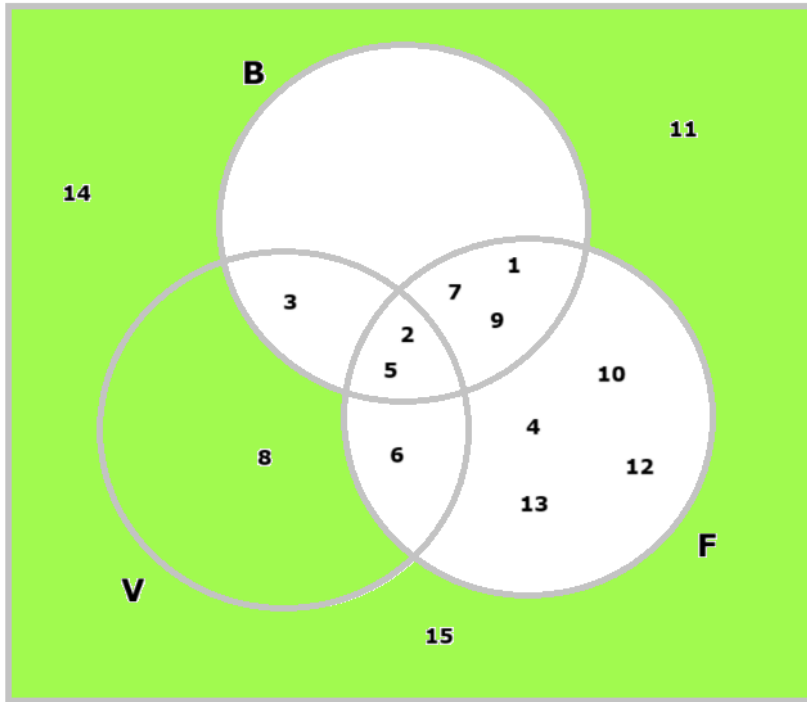
- A las personas 2 y 5 les gustan los tres deportes mencionados.
- A las personas 11, 14 y 15 no le gustan ninguno de los deportes mencionados.
- A las personas 4, 10, 12 y 13 les gusta únicamente el futbol.
- A la persona 8 le gusta únicamente el voleibol.
- A 7 personas les gusta dos o más deportes.

Ahora se realizarán operaciones de conjuntos y luego se verificará con el diagrama de Venn.

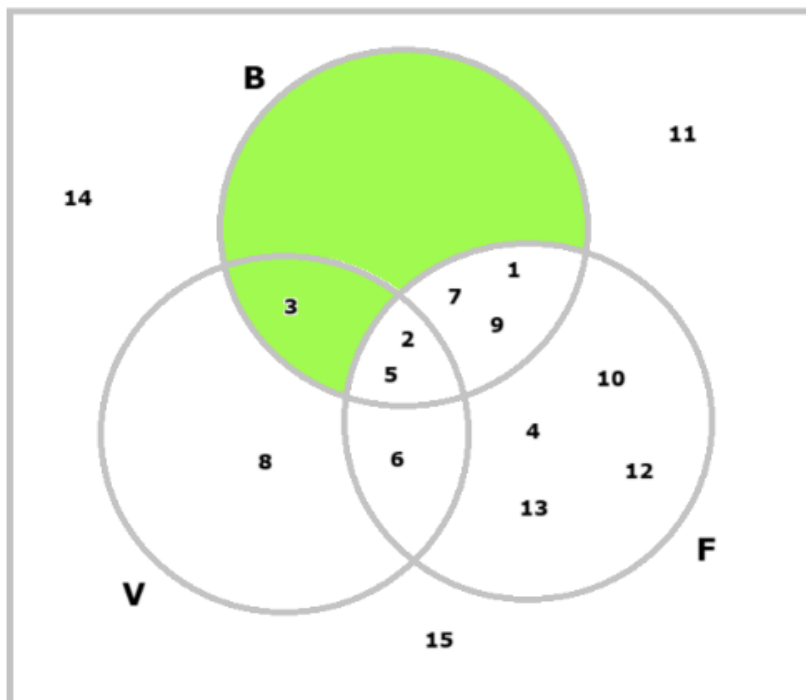
1.- $(B \cup F)^c$

Para ello se tiene que $(B \cup F) = \{1, 2, 3, 4, 5, 6, 7, 9, 10, 12, 13\}$, por lo tanto $(B \cup F)^c = \{8, 11, 14, 15\}$

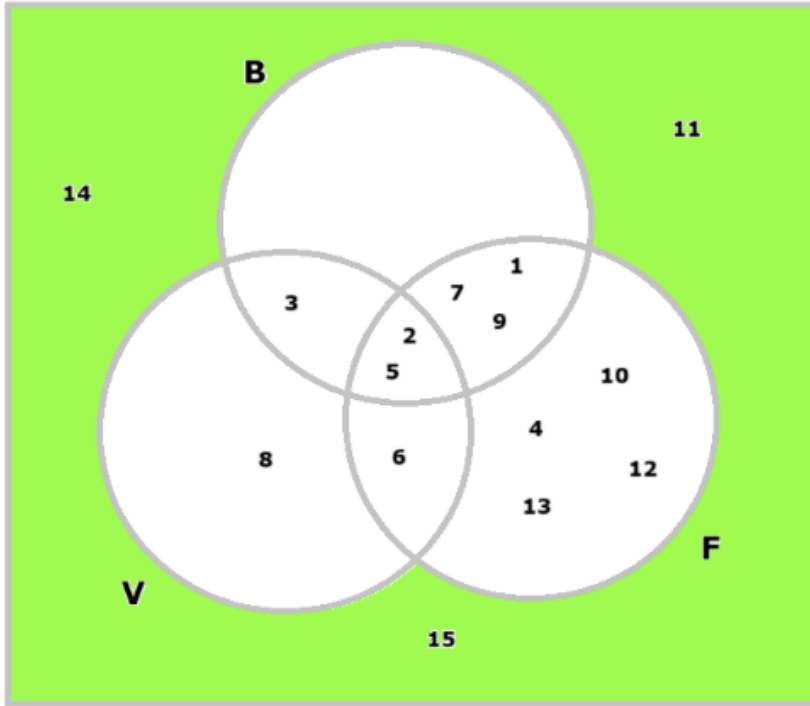
En el siguiente diagrama de Venn se presenta en color verde el área correspondiente a $(B \cup F)^c$ y se observa con facilidad que la respuesta obtenida anteriormente es correcta



2.- $(B - F) = \{3\}$ Recuerde que son los elementos de B que no pertenecen a F. Su diagrama de Venn con el área iluminada en verde sería.



3.- $(A \cup B \cup C)^c = \{11, 14, 15\}$. Esta solución corresponde a los elementos que le faltan a la unión de los tres conjuntos para ser universo. Su diagrama de Venn sería.



De lo anterior se aprecia la utilidad que tienen los diagramas de Venn en el análisis de conjuntos, sin embargo, es muy común que la cantidad de datos es mucho mayor al ejemplo proporcionado de 15 personas encuestadas y que en muchas ocasiones solo es de interés saber cuántos son los elementos de un conjunto o de una operación de conjuntos. Considerando el caso anterior se podría preguntar, a cuántas personas les gusta el beisbol, el futbol o ambos, lo que en notación de conjuntos sería $n(B \cup F)$ para lo cual la respuesta sería que son 11 personas que se encuentran en esta situación, dicha respuesta fue sencilla porque la cantidad de personas encuestadas solo fue de 15, una alternativa para obtener la respuesta anterior es aplicar la formula.

$$n(A \cup B) = n(A) + n(B) - n(A \cap B)$$

Que utilizando la nomenclatura de ejercicio que se está analizando, sería.

$$n(B \cup F) = n(B) + n(F) - n(B \cap F)$$

Del diagrama de Venn se tiene que la respuesta es de 11 personas y aplicando la formula sería.

$$n(B \cup F) = 6 + 10 - 5 = 11$$

Lo cual es coincidente con la respuesta obtenida previamente.

La fórmula correspondiente al número de elementos de la unión de 3 conjuntos sería:

$$n(A \cup B \cup C) = n(A) + n(B) + n(C) - n(A \cap B) - n(A \cap C) - n(B \cap C) + n(A \cap B \cap C)$$

Adecuando la fórmula para el caso que se está analizando sería:

$$n(B \cup F \cup V) = n(B) + n(F) + n(V) - n(B \cap F) - n(B \cap V) - n(F \cap V) + n(B \cap F \cap V)$$

$$n(B \cup F \cup V) = 6 + 10 + 5 - 5 - 5 - 3 + 2 = 12$$

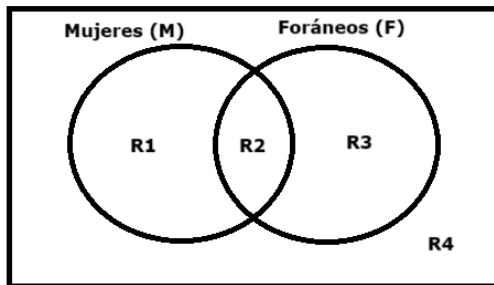
De observar el diagrama de Venn se puede verificar que efectivamente son 12 las personas que les gusta al menos uno de los tres deportes.

Con los elementos proporcionados al momento se puede resolver ejercicios de aplicación de conjuntos como el que se presenta a continuación.

1.- En entrevista realizada a 60 estudiantes del Instituto Tecnológico de Cd. Cuauhtémoc, se estableció que 35 eran mujeres, 25 eran foráneos y 10 eran foráneas y mujeres. En base a estos datos, calcule:

- A) Cuantos alumnos son hombres.
- B) Cuantas alumnas no son foráneas.
- C) Cuantos alumnos no son foráneos.

El diagrama de Venn sería el siguiente.



Solución

- A) Cuantos alumnos son hombres.

La cantidad de hombres en la muestra corresponde a $n(M^c)$ ya que hombre equivale a no ser mujer y su respuesta en regiones corresponde a la suma de las regiones R3 y R4.

$$n(M^c) = n(U) - n(M) = 60 - 35 = 25$$

Del total de los estudiantes 25 son hombres.

- B) Cuantas alumnas no son foráneas.

El termino alumna implica que es mujer y el hecho de que no sea foránea corresponde a los elementos del conjunto mujer que no corresponden al conjunto foránea, la operación de conjuntos correspondiente es $n(M) - n(M \cap F)$ y la respuesta en base a regiones viene a ser R1

$$n(M) - n(F) = 35 - 10 = 25$$

Son 25 alumnas las que no son foráneas.

C) Cuantos alumnos no son foráneos.

En esta expresión se tiene que no es mujer y que no es foráneo, por lo tanto, la solución corresponde al complemento de ambos conjuntos y la solución en regiones es R4.

$$n(M \cup F)^c = n(U) - n(M \cup F)$$

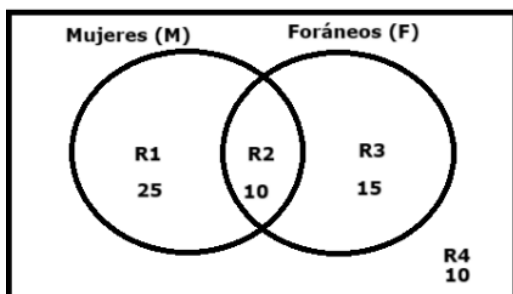
$$n(M \cup F) = n(M) + n(F) - n(M \cap F) = 35 + 25 - 10 = 50$$

$$n(M \cup F)^c = 60 - 50 = 10$$

Del total de los estudiantes encuestados, 10 son hombres que no son foráneos.

A continuación, se presenta la solución en base a regiones, para ello se procederá primeramente a calcular cada una de las regiones.

La región R2 corresponde a la intersección de ambos conjuntos ($M \cap F$), teniendo así un valor de 10, las 35 mujeres son la suma de R1 y R2, por lo que $R1 + R2 = 60$, por lo tanto, $R1 = 25$, de forma análoga se tiene que $R2 + R3 = 25$, teniendo con ello que $R3 = 15$.



El número de elementos de la unión de los conjuntos de mujeres y foráneos $n(M \cup F)$ es.

$$n(M \cup F) = n(M) + n(F) - n(M \cap F) = 35 + 25 - 10 = 50$$

Lo anterior en base a regiones corresponde a la suma de las regiones R1, R2 y R3.

$$R1 + R2 + R3 = 25 + 10 + 15 = 50$$

De lo anterior se aprecia que ambas soluciones son coincidentes.

La respuesta a los incisos de este ejercicio en base a los valores de las regiones es.

A) Cuantos alumnos son hombres.

La cantidad de hombres en la muestra corresponde a $n(M^c)$ ya que hombre equivale a no ser mujer y su respuesta en regiones corresponde a la suma de las regiones R3 y R4.

$$R3 + R4 = 15 + 10 = 25$$

Del total de los estudiantes 25 son hombres.

B) Cuantas alumnas no son foráneas.

El termino alumna implica que es mujer y el hecho de que no sea foránea corresponde a los elementos del conjunto mujer que no corresponden al conjunto foránea, la operación de conjuntos correspondiente es $n(M) - n(M \cap F)$ y la respuesta en base a regiones viene a ser R1

$$R1 = 25$$

Son 25 alumnas las que no son foráneas.

C) Cuantos alumnos no son foráneos.

En esta expresión se tiene que no es mujer y que no es foráneo, por lo tanto, la solución corresponde al complemento de ambos conjuntos y la solución en regiones es R4.

$$R4 = 10$$

Del total de los estudiantes encuestados, 10 son hombres que no son foráneos.

2.- En un estudio de mercado en Cd. Cuauhtémoc sobre el tipo de alimentos que más le gustan a la población se realizó una encuesta a 450 personas adultas seleccionadas aleatoriamente en la plaza de la ciudad, obteniéndose que 180 personas prefieren los alimentos enlatados, 165 le gustan los alimentos refrigerados, 118 prefieren alimentos deshidratados, a 58 les gustan los alimentos enlatados y los alimentos refrigerados, 39 persona prefieren alimentos enlatados y alimentos deshidratados, a 29 personas les gustan los alimentos refrigerados y los alimentos deshidratados y solo a 24 personas les gustan los tres tipos de alimentos.

Considerando lo anterior, determine a cuantas personas:

- A) Les gusta únicamente los alimentos enlatados.
- B) Les gusta dos o más de los tipos de alimentos mencionados.
- C) Les gusta al menos uno de los tipos de alimentos mencionados.
- D) Les gusta solamente los alimentos enlatados y los alimentos deshidratados.
- E) No les gusta ninguno de los tipos de alimentos mencionados.

Datos:

$$n(U) = 450$$

$$n(E) = 180$$

$$n(R) = 165$$

$$n(D) = 118$$

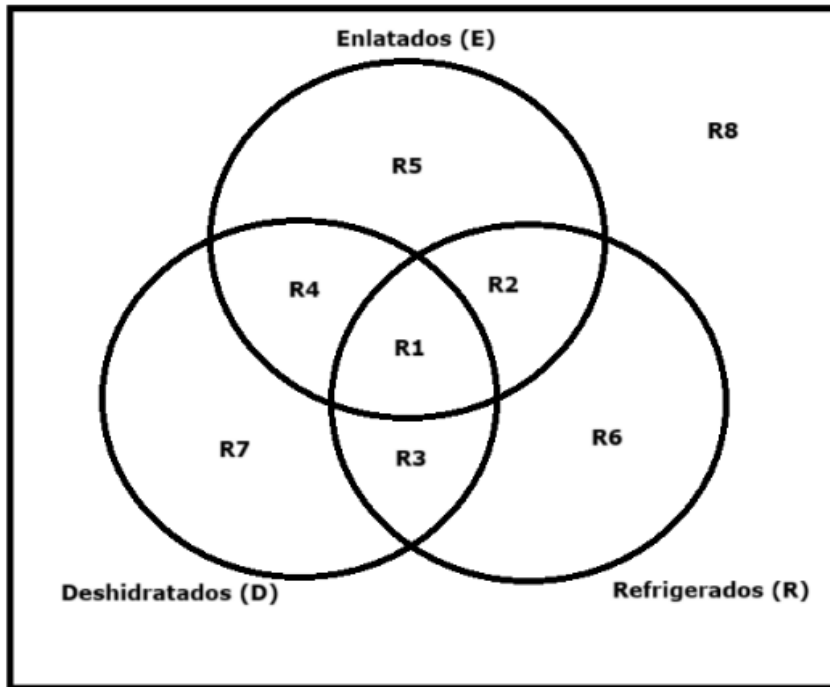
$$n(E \cap R) = 58$$

$$n(E \cap D) = 39$$

$$n(R \cap D) = 29$$

$$n(E \cap R \cap D) = 24$$

El diagrama de Venn para este caso seria



En este diagrama se aprecia que existen 8 regiones, lo cual facilita los planteamientos necesarios para responder a las preguntas planteadas.

Respuestas.

A) Les gusta únicamente los alimentos enlatados.

De observar el diagrama se aprecia que la respuesta a esta pregunta corresponde a la región R5, lo cual equivale decir que son los elementos de la unión de los tres conjuntos, que no pertenecen a los elementos de la unión de los conjuntos R y D, que expresado en operaciones de conjuntos sería $n(E \cup R \cup D) - n(R \cup D)$.

$$n(E \cup R \cup D) = 180 + 165 + 118 - 58 - 39 - 29 + 24 = 361$$

$$n(R \cup D) = 165 + 118 - 29 = 254$$

Por lo tanto, el resultado sería

$$n(E \cup R \cup D) - n(R \cup D) = 361 - 254 = 107$$

De lo anterior se desprende que a 107 personas les gusta únicamente los alimentos enlatados.

B) Les gusta dos o más de los tipos de alimentos mencionados.

Las regiones correspondientes a que a las personas les guste 2 o 3 de los tipos de alimentos en cuestión corresponde a la suma de las regiones R1, R2, R3 y R4, que expresado en operaciones de conjuntos sería $n(E \cap R) + n(E \cap D) + n(R \cap D) - 2n(E \cap R \cap D)$.

$$n(E \cap R) + n(E \cap D) + n(R \cap D) - 2n(E \cap R \cap D) = 58 + 39 + 29 - 2(24) = 78$$

A 78 personas les gusta 2 o 3 de los tipos de alimentos en cuestión.

C) Les gusta al menos uno de los tipos de alimentos mencionados.

Las regiones correspondientes a que a las personas les guste 1, 2 o 3 de los tipos de alimentos en cuestión corresponde a la suma de las regiones R1, R2, R3, R4, R5, R6 y R7 que expresado en operaciones de conjuntos viene a ser $n(E \cup R \cup D)$.

$$(E \cup R \cup D) = 180 + 165 + 118 - 58 - 39 - 29 + 24 = 361$$

A 361 personas les gusta al menos uno de los tipos de alimentos en mención.

D) Les gusta solamente los alimentos enlatados y los alimentos deshidratados.

La región correspondiente a que les gusta solamente los alimentos enlatados y los alimentos deshidratados corresponde a la región R4, que expresado en operaciones de conjuntos corresponde a $n(E \cap D) - n(E \cap R \cap D)$.

$$n(E \cap D) - n(E \cap R \cap D) = 39 - 24 = 15$$

A 15 personas les gusta solamente los alimentos enlatados y los alimentos deshidratados.

E) No les gusta ninguno de los tipos de alimentos mencionados.

La región correspondiente a que no les gusta ninguno de los tipos de alimentos mencionados corresponde a la región R8, que expresado en operaciones de conjuntos corresponde a $n(U) - n(E \cup R \cup D)$.

$$n(U) - n(E \cup R \cup D) = 450 - 361 = 89$$

A 89 personas no les gusta ninguno de los tipos de alimentos mencionados.

La solución de estos incisos en base a los planteamientos de las regiones a que corresponde cada uno de los incisos, implica calcular cada una de las 8 regiones y posteriormente hacer las sumas correspondientes. El cálculo de cada una de las regiones se presenta a continuación,

R1 corresponde a la intersección de los 3 conjuntos y es igual a 24

La intersección entre E y R corresponde a 58, por lo que $R1 + R2 = 58$, de tal forma que $R2 = 34$

La intersección entre E y D corresponde a 39, por lo que $R1 + R4 = 39$, de tal forma que $R4 = 15$.

La intersección entre R y D corresponde a 29, por lo que $R1 + R3 = 29$, de tal forma que $R3 = 5$

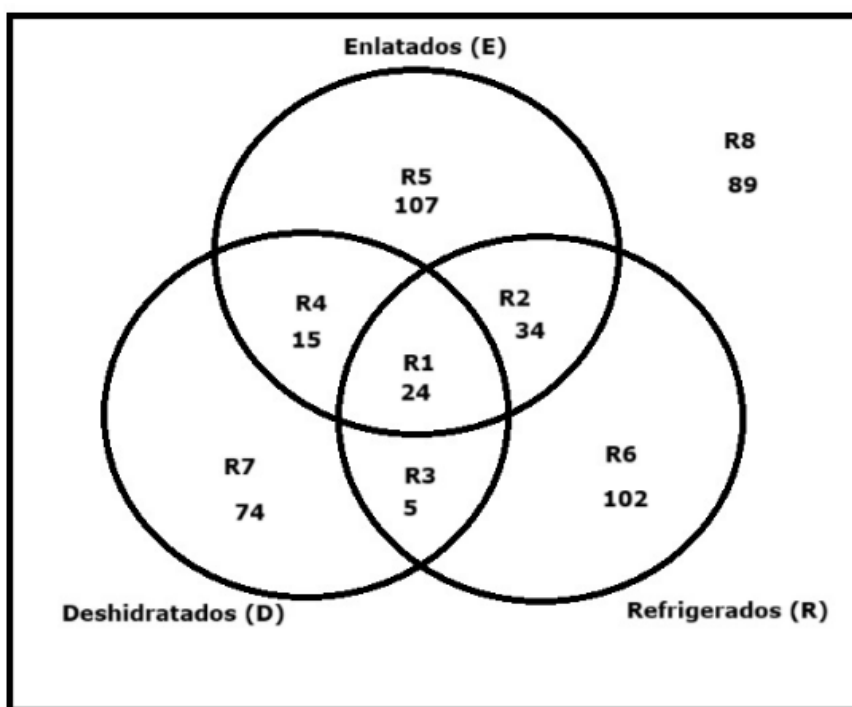
La cantidad de personas que les gusta los alimentos enlatados es de 180, por lo que $R1 + R2 + R4 + R5 = 180$, como ya se conocen los valores de R1, R2 y R4, se establece que $R5 = 107$

De forma análoga se calcula los valores de R6 y R7, resultando en que $R6 = 102$ y $R7 = 74$.

De aplicar la fórmula de la cantidad de elementos correspondiente a la unión de los tres conjuntos resulto en un valor de 361 personas, lo cual coincide perfectamente con la suma de las regiones R1 a la R7.

El valor de R8 resulta de restarle al total de 450 personas los valores de las regiones R1 a la R7 y da como resultado 89.

Los valores de cada una de las regiones se pueden observar en el siguiente diagrama de Venn.



La verificación de las respuestas obtenidas anteriormente, mediante el uso de los valores calculados de cada una de las 8 regiones se presenta a continuación

A) Les gusta únicamente los alimentos enlatados.

La respuesta a esta pregunta corresponde a la región R5 y es igual a 107.

B) Les gusta dos o más de los tipos de alimentos mencionados.

Las regiones correspondientes a que a las personas les guste 2 o 3 de los tipos de alimentos en cuestión corresponde a la suma de las regiones R1, R2, R3 y R4

$$R1, R2, R3 \text{ y } R4 = 24 + 34 + 5 + 15 = 78.$$

C) Les gusta al menos uno de los tipos de alimentos mencionados.

Las regiones correspondientes a que a las personas les guste 1, 2 o 3 de los tipos de alimentos en cuestión corresponde a la suma de las regiones R1, R2, R3, R4, R5, R6 y R7

$$R1 + R2 + R3 + R4 + R5 + R6 + R7 = 24 + 34 + 5 + 15 + 107 + 102 + 74 = 361.$$

Les gusta solamente los alimentos enlatados y los alimentos deshidratados.

La región correspondiente a que les gusta solamente los alimentos enlatados y los alimentos deshidratados corresponde a la región R4 y es igual a 15.

D) No les gusta ninguno de los tipos de alimentos mencionados.

La región correspondiente a que no les gusta ninguno de los tipos de alimentos mencionados corresponde a la región R8 y es igual a 89.

Comparando ambas formas de solución se aprecia que son equivalentes.

Problemas propuestos.

1.- En una investigación a los empresarios de la industria alimentaria de la región de Cuauhtémoc, se les pregunto que si que consideraban más importante en los alimentos, la calidad, la inocuidad o ambos. De 95 encuestas realizadas 55 contestaron que la calidad, 26 que la inocuidad y 4 consideraron que ninguna era importante. En base a los datos proporcionados, calcule:

- A) Cuantos empresarios consideran que la calidad y la inocuidad es importante. (Respuesta 30 empresarios)
- B) Cuantos empresarios consideran que únicamente la calidad es importante. (Respuesta 5 empresarios)
- C) Cuantos empresarios consideran que la calidad, la inocuidad o ambos es importante. (Respuesta 91 empresarios).

2.- En estudio realizado a 68 alumnos de ingeniería en industrias alimentarias del ITCC se les pregunto qué tipo de materias les gustaba más, 25 dijeron que las del área de investigación, 31 opinaron que las del área de tecnología y 12 opinaron que ambas, considerando los datos obtenidos, calcule:

- A) Cuantos estudiantes tienen el gusto por la investigación, la tecnología o ambos. (Respuesta 44 estudiantes.
- B) Cuantos estudiantes tienen el gusto por la investigación y la tecnología. (Respuesta 12 estudiantes.
- C) A cuantos estudiantes no les gusta ninguna de las dos. (Respuesta 24 estudiantes).

3.- En un estudio realizado en un centro comercial de la ciudad de chihuahua se analizaron los ingredientes de la etiqueta de 78 alimentos procesados, encontrando que además de los ingredientes principales: 35 contenían saborizantes (S), 25 contenían acidulantes (A), 40 contenían conservadores (C), 15 contenían saborizantes y acidulantes, 12 contenían saborizantes y conservadores, 8 contenían acidulantes y conservadores, y 5 contenían los tres ingredientes. En base a los datos obtenidos, se desea calcular:

- A) Cuantos alimentos contenían únicamente conservador Les gusta únicamente los alimentos enlatados.
- B) Además de los ingredientes principales cuantos contenían solamente acidulantes. (Respuesta 7 alimentos).

- C) Además de los ingredientes principales cuantos contenían dos de los ingredientes en mención. (Respuesta 20 alimentos).
- D) Además de los ingredientes principales cuantos contenían uno o más de los ingredientes en mención. (Respuesta 70 alimentos).
- E) Además de los ingredientes principales cuantos contenían acidulantes, conservadores o ambos. (Respuesta 45 alimentos).
- F) Además de los ingredientes principales cuantos no contenían ninguno de los ingredientes en mención. (Respuesta 8 alimentos).

4.- Unos estudiantes de la carrera de ingeniería en industrias alimentarias desean emprender con la idea de una golosina para niños enriquecida en proteínas y para el diseño del producto desean establecer cuál es el sabor de frutas que más les gusta a los niños, por lo que deciden realizar una investigación de mercado, encuestando a niños en edad de primaria. Después de diseñar la encuesta, establecer el tamaño de muestra requerido y llevar a cabo dicho estudio, se obtuvieron los siguientes datos: fueron un total de 580 niños encuestados, a 324 de ellos les gusta el sabor naranja, a 395 les gusta el de fresa, a 322 les gusta el de uva, a 185 les gusta naranja y fresa, a 179 les gusta naranja y uva, a 322 les gusta fresa y uva y solo a 13 niños no les gusta ninguno de los sabores. En base a estos datos establezca:

- A) A cuántos niños les gusta los tres sabores. (Respuesta a 13 niños)
- B) A cuántos niños les gusta únicamente el sabor de uva. (Respuesta a 33 niños)
- C) A cuántos niños les 2 o más de los sabores mencionados. (Respuesta a 472 niños)
- D) A cuántos niños el sabor de fresa, el de uva o ambos. (Respuesta a 512 niños)
- E) A cuántos niños les gusta el sabor de naranja y fresa, pero no el de uva. (Respuesta a 90 niños)
- F) A cuántos niños les gusta al menos uno de los sabores en cuestión. (Respuesta a 567 niños)

1.3. Análisis combinatorio Principio fundamental de conteo (diagramas de árbol).

El objetivo central del análisis combinatorio es responder preguntas del tipo: ¿De cuántas maneras se puede organizar un conjunto de elementos? o ¿Cuántas combinaciones diferentes se pueden formar bajo ciertas restricciones? Estas preguntas, aparentemente simples, pueden volverse extremadamente complejas a medida que aumenta el tamaño de los conjuntos o las restricciones (González, 2012).

Por ejemplo, al estudiar la disposición de las letras de una palabra, el análisis combinatorio permite calcular las permutaciones (donde el orden importa) o las combinaciones (donde el orden no importa). Si se agregan condiciones, como que ciertas letras deben estar juntas, el problema se vuelve más intrincado, pero las herramientas combinatorias siguen siendo capaces de resolverlo (López & Torres, 2010).

El Principio Fundamental de Conteo. Este principio establece que, si un proceso se compone de varias etapas independientes, y cada etapa puede realizarse de un número específico de maneras, el número total de formas en que puede completarse todo el proceso es el producto de las maneras en que puede realizarse cada etapa (Ballesteros & Hernández, 2017).

Por ejemplo, si una persona quiere armar un conjunto de ropa eligiendo entre 3 tipos de camisas, 2 tipos de pantalones y 4 tipos de zapatos, el número total de combinaciones posibles será: $3 \times 2 \times 4 = 24$ formas diferentes en que se puede vestir esta persona.

Dentro del análisis combinatorio se encuentran las combinaciones y permutaciones, las cuales se abordarán a detalle en los temas posteriores, así mismo se verá cómo es una herramienta de amplio uso en el área de la probabilidad.

Como se veía anteriormente, el principio multiplicativo es de utilidad para establecer la cantidad de ordenaciones que se pueden obtener combinando los elementos de dos conjuntos o más, sin embargo, en ocasiones es requerido no solo saber la cantidad, sino cuáles son, para lo cual es de utilidad la elaboración de diagramas de árbol o tablas de entrada múltiple, ejemplo de ello puede ser lo siguiente.

1.- Para el lanzamiento de dos dados establezca de cuantas formas diferentes se puede obtener el resultado y cuáles son.

Por principio multiplicativo se obtiene la cantidad de formas diferentes en que puede caer el lanzamiento de los dos dados, correspondiendo esto a $6 \times 6 = 36$ formas diferentes.

Para saber cuáles son esas 36 formas diferentes en que puede caer el lanzamiento de los dos dados se puede elaborar el diagrama de árbol y la tabla de entrada múltiple, que se presentan a continuación.

Diagrama de árbol

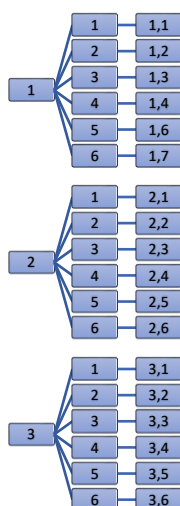
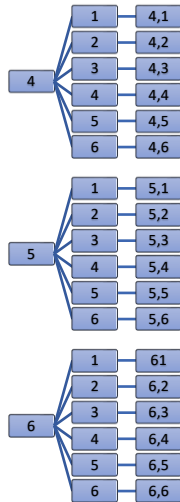


Tabla de entrada múltiple

		Segundo dado					
		1	2	3	4	5	6
Primer dado	1	1, 1	2, 1	3, 1	4, 1	5, 1	6, 1
	2	1, 2	2, 2	3, 2	4, 2	5, 2	6, 2
	3	1, 3	2, 3	3, 3	4, 3	5, 3	6, 3
	4	1, 4	2, 4	3, 4	4, 4	5, 4	6, 4
	5	1, 5	2, 5	3, 5	4, 5	5, 5	6, 5
	6	1, 6	2, 6	3, 6	4, 6	5, 6	6, 6



2.- Para el lanzamiento de tres monedas establezca de cuantas formas diferentes se puede obtener el resultado y cuales son.

Diagrama de árbol

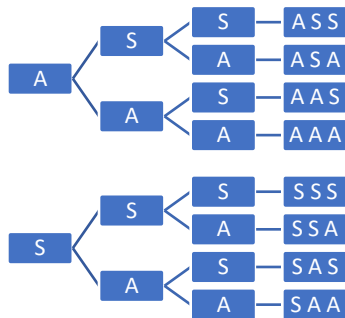


Tabla de entrada múltiple

		Tercer Moneda			
		A		S	
Segunda Moneda	A	AAA	ASA	AAS	ASS
	S	SAA	SSA	SAS	SSS

En ambos casos se puede apreciar que la cantidad de formas diferentes en que puede caer el lanzamiento de 3 monedas es de 8, lo cual coincide perfectamente con el principio multiplicativo ($2 \times 2 \times 2 = 8$), adicional a ello se puede observar cuáles son esas 8 formas diferentes.

Ejercicios propuestos.

1.- Elabore un diagrama de árbol o una tabla de entrada múltiple, para las formas diferentes en que puede servirse un platillo en un restaurante si: la sopa puede ser arroz o fideos, el guisado puede ser barbacoa, chile colorado o mole y las tortillas pueden ser de maíz o de harina. Verifique que la cantidad de combinaciones logradas coincide con el resultado del principio multiplicativo.

2.- Elabore un diagrama de árbol o una tabla de entrada múltiple, para las formas diferentes en que una persona puede vestirse si tiene 3 camisetas, 4 pantalones y 2 pares de zapatos.

Verifique que la cantidad de combinaciones logradas coincide con el resultado del principio multiplicativo.

3.- Elabore un diagrama de árbol o una tabla de entrada múltiple, para las formas diferentes en que puede realizarse un experimento de deshidratación de manzana si tiene 3 opciones de espesor de las rodajas (2, 3 y 4 mm), 3 diferentes temperaturas (50, 60 y 70 °C) y dos equipos diferentes de secado (Eq1 y Eq2). Verifique que la cantidad de combinaciones logradas coincide con el resultado del principio multiplicativo.

1.4. Permutaciones (simples, circulares y con repetición).

Tal como se mencionaba, dentro de las ordenaciones se tiene que cuando el orden de los elementos tiene importancia, se dice que es una permutación, las cuales pueden ser simples, circulares y con repetición, a continuación, se describirá cada una de ellas.

Permutaciones simples

Este tipo de permutaciones corresponde a cuando los elementos no pueden ser repetidos, a continuación, se presentan unos ejemplos representativos de este tipo de permutación.

Ejemplo 1.

Establezca de cuantas formas diferentes puede suceder el resultado de primer, segundo y tercer lugar en resolver correctamente un ejercicio de probabilidad, en un grupo que está constituido por 30 alumnos.

De la lectura de este ejercicio se tiene que, si corresponde a permutaciones, ya que el orden si es importante, por el hecho de que por lo general a las personas les gusta ocupar el primer lugar, luego el segundo y así sucesivamente.

Aplicando el principio multiplicativo se tiene que el primer lugar puede ser ocupado por cualquiera de los 30 alumnos, luego el segundo por 29 alumnos y el tercero por 28 alumnos, teniendo así que el resultado corresponde a $30 \times 29 \times 28 = 24360$ formas diferentes.

La otra alternativa de solución es aplicar la formula.

$${}_nP_x = \frac{n!}{(n-x)!}$$

Donde:

n = Total de elementos

x = Cantidad de objetos tomados.

n! = Factorial de n

Ejemplos de factorial:

$$5! = 5 \times 4 \times 3 \times 2 \times 1 = 120$$

$$8! = 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1 = 40320$$

$$0! = 1$$

Aplicando la fórmula para el ejemplo que se está analizando, se tiene que $n = 30$ y $x = 3$, llegando así a lo siguiente.

$${}_{30}P_3 = \frac{n!}{(n-x)!} = \frac{30!}{(30-3)!} = \frac{30!}{27!} = 24360$$

24360 formas diferentes de que se ocupen el primer, segundo y tercer lugar.

Lo anterior coincide perfectamente con el resultado obtenido por principio multiplicativo.

Ejemplo 2.

Con los dígitos del 0 al 9, cuantos números diferentes sin repetir los dígitos se pueden obtener si:

- A) Se toman 4 de los dígitos.
- B) Se toman 7 de los dígitos.
- C) Se toman los 10 dígitos.

Como el orden en que se acomoden los dígitos es importante y se indica que los dígitos no se pueden repetir, este ejercicio si corresponde a permutaciones simples.

- A) ${}_{10}P_4 = \frac{n!}{(n-x)!} = \frac{10!}{(10-4)!} = \frac{10!}{6!} = 5040$ números diferentes de 4 dígitos.
- B) ${}_{10}P_7 = \frac{n!}{(n-x)!} = \frac{10!}{(10-7)!} = \frac{10!}{3!} = 604800$ números diferentes de 7 dígitos.
- C) ${}_{10}P_{10} = \frac{n!}{(n-x)!} = \frac{10!}{(10-10)!} = \frac{10!}{0!} = 3628800$ números diferentes de 10 dígitos.

Recuerde que, factorial de cero es igual a uno ($0! = 1$).

Ejercicios propuestos.

1.- Para abrir el portón de acceso a un fraccionamiento privado es mediante una clave de 4 dígitos que deben ser tecleados en un teclado. De cuantas formas diferentes puede ser la clave de acceso si:

- A) Los dígitos no se pueden repetir.
- B) Los dígitos no se pueden repetir y son puros dígitos pares.
- C) Los dígitos no se pueden repetir y la clave es un número impar.

D) Los dígitos no se pueden repetir, la clave empieza con número par y termina en número impar.

2.- Cuantos números de teléfono celular se pueden ofrecer en la región de ciudad Cuauhtémoc, considerando que el número de celular está compuesto por 10 dígitos, los dígitos no se pueden repetir, su lada es 625 y:

A) Termina en número par.

B) Después de la lada 625 se sabe que los siguientes dígitos son 106.

C) Son puros dígitos pares.

D) Ninguno de los dígitos es el 4.

3.- Se sabe que la escuela primaria Abraham González de Cd. Cuauhtémoc cuenta con 250 padres de familia. De cuantas formas diferentes se puede formar el comité de padres de familia constituido por presidente, secretario y tesorero.

4.- Considerando que un grupo de 20 amigos organizó una rifa entre ellos, para lo cual cada uno aportó cien pesos, juntando así \$2000 pesos, que distribuyeron en un premio de primer lugar de \$1000 pesos, un premio de segundo lugar de \$600 pesos y un premio de tercer lugar de \$400 pesos. De cuantas formas diferentes puede suceder la premiación.

5.- A un maratón que se realizó en la Sierra Tarahumara asistieron 50 canadienses, 80 estadounidenses, 200 mexicanos y 50 sudamericanos. De cuantas formas pueden entregarse las medallas de primero, segundo y tercer lugar, si:

A) Se considera a todos los participantes.

B) El primer lugar lo obtuvo un mexicano.

C) El segundo y tercer lugar lo obtuvieron mexicanos.

D) El tercer lugar lo obtuvo un canadiense.

Permutaciones con repetición.

En este tipo de ordenaciones se tiene que el orden de los elementos también tiene importancia y adicional a ello se tiene que los elementos se pueden repetir. A continuación, se presenta ejemplo de ello.

Ejemplo 1. Con los dígitos del 0 al 9, considerando que los dígitos se pueden repetir, cuantos números diferentes se pueden obtener si:

A) Se toman 4 de los dígitos.

B) Se toman 7 de los dígitos.

C) Se toman los 10 dígitos.

Solución

A) La solución se obtiene fácilmente aplicando el principio multiplicativo, teniendo que para cada una de las casillas correspondiente a los cuatro dígitos del número la respuesta se logra mediante la multiplicación de $10 \times 10 \times 10 \times 10 = 10^4 = 10,000$ formas diferentes.

B) De forma análoga al inciso anterior, la solución sería $10 \times 10 \times 10 \times 10 \times 10 \times 10 \times 10 = 10^7 = 10,000,000$ formas diferentes.

C) $10 \times 10 \times 10 \times 10 \times 10 \times 10 \times 10 \times 10 \times 10 \times 10 = 10^{10} = 10,000,000,000$ formas diferentes.

Ejercicios propuestos.

1.- Para abrir el portón de acceso a un fraccionamiento privado es mediante una clave de 4 dígitos que deben ser tecleados en un teclado. De cuantas formas diferentes puede ser la clave de acceso si:

A) Los dígitos se pueden repetir.

B) Los dígitos se pueden repetir y son puros dígitos pares.

C) Los dígitos se pueden repetir y la clave es un número impar.

D) Los dígitos se pueden repetir, la clave empieza con número par y termina en número impar.

2.- Cuantos números de teléfono celular se pueden ofrecer en la región de ciudad Cuauhtémoc, considerando que el número de celular está compuesto por 10 dígitos, los dígitos se pueden repetir, su lada es 625 y:

A) Termina en número par.

B) Después de la lada 625 se sabe que los siguientes dígitos son 106.

C) Son puros dígitos pares.

D) Ninguno de los dígitos es el 4.

3.- Calcular la cantidad de formas diferentes de establecer la clave con que puede ser abierto un portafolio con candado de 4 perillas, usted sabe que la clave termina en el número cero y los dígitos se pueden repetir.

4.- Considerando las letras a, b, c, d, e, f, g. Calcule.

A) Cuantas palabras de 4 letras aun sin sentido se pueden hacer, si las letras se pueden repetir.

B) Cuantas palabras de 6 letras aun sin sentido se pueden hacer, si las letras se pueden repetir.

C) Cuantas palabras de 3 letras aun sin sentido se pueden hacer, si las letras se pueden repetir y la palabra termina con la letra a.

5.- A un maratón que se realizó en la Sierra Tarahumara asistieron 50 canadienses, 80 estadounidenses, 200 mexicanos y 50 sudamericanos. De cuantas formas pueden entregarse las medallas de primero, segundo y tercer lugar, si:

- A) Se considera a todos los participantes.
- B) El primer lugar lo obtuvo un mexicano.
- C) El segundo y tercer lugar lo obtuvieron mexicanos.
- D) El tercer lugar lo obtuvo un canadiense.

Permutaciones circulares.

Las permutaciones circulares se aplican cuando los elementos se disponen en un círculo, como en una mesa redonda. En este caso, las rotaciones no generan nuevas permutaciones. Por ejemplo, si tenemos las posiciones A, B, C, y D en un círculo, las disposiciones (A,B,C,D) y (B,C,D,A) se consideran equivalentes.

El número de permutaciones circulares de n elementos se calcula como: $(n-1)!$

Esto se debe a que, al fijar un elemento como referencia (para evitar contar rotaciones redundantes), los $n - 1$ elementos pueden permutarse libremente.

Un ejemplo de permutaciones circulares se presenta a continuación.

1.- De cuantas formas diferentes pueden sentarse en una mesa redonda Juan, Pedro y Manuel.

Como el número de elementos es igual a 3, las formas diferentes en que pueden sentarse estas personas en una mesa redonda es igual a $(3 - 1)! = 2! = 2$, las cuales corresponden a:

2.- De cuantas formas diferentes pueden sentarse en una mesa redonda Juan, Pedro, Luisa y Manuel.

Como el número de elementos es igual a 4, las formas diferentes en que pueden sentarse estas personas en una mesa redonda son igual a $(4 - 1)! = 3! = 6$, las cuales corresponden a:

Ejercicios propuestos

1.- De cuantas formas diferentes se pueden subir los niños a un carrusel de 7 caballitos.

2.- De cuantas formas diferentes se puede hacer una ruleta de 10 colores.

3.- De cuantas formas diferentes de pueden acomodar 6 platillos en una mesa redonda.

4.- De cuantas formas diferentes se pueden acomodar 8 persona en una mesa redonda.

5.- De cuantas formas diferentes se pueden acomodar las cuatro velas de una corona de adviento.

Permutaciones distinguibles.

Este tipo de permutaciones se aplica cuando dentro del conjunto de elementos a ordenar, se encuentran elementos iguales.

Ejemplo 1. Cuantas y cuales palabras aun sin sentido se pueden hacer con las letras de la palabra SOS.

Aplicando la fórmula de permutaciones distinguibles.

$$nP_{n_1 n_2 \dots n_k} = \frac{n!}{n_1! n_2! \dots n_k!}$$

Se tiene que la cantidad de palabras diferentes son:

$$3P_{2,1} = \frac{3!}{2! 1!} = 3$$

En el denominador se puede apreciar que son 2 elementos diferentes en la palabra SOS y que la letra “S” aparece en dos ocasiones.

Las palabras diferentes son: SOS, SOO, OOS. Lo cual coincide perfectamente con el resultado de la fórmula de permutaciones distinguibles.

Ejemplo 2. Cuantas y cuales palabras aun sin sentido se pueden hacer con las letras de la palabra CASA.

Aplicando la fórmula de permutaciones distinguibles.

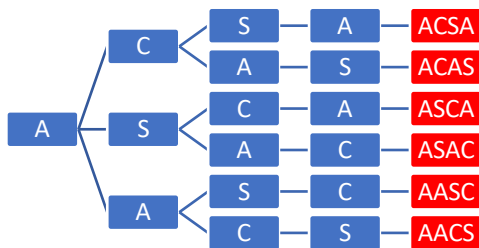
$$nP_{n_1 n_2 \dots n_k} = \frac{n!}{n_1! n_2! \dots n_k!}$$

Se tiene que la cantidad de palabras diferentes son:

$$4P_{2,1,1} = \frac{4!}{2! 1! 1!} = 12$$

En el denominador se puede apreciar que son 3 elementos diferentes en la palabra CASA y que la letra “a” aparece en dos ocasiones.

Las doce palabras diferentes que se pueden hacer con la palabra CASA se observan en el siguiente diagrama de árbol, en el cual se colocaron en color rojo las palabras repetidas.



1.5. Combinaciones y Teorema del Binomio.

Las combinaciones son ordenaciones donde el orden no es importante, ejemplo de ello es un cono con dos bolas de nieve de diferente sabor, seleccionar a tres representantes de un grupo de estudiantes,

A continuación, se presentan ejemplos de combinaciones.

Ejemplo 1. De cuantas formas diferentes puede usted solicitar un cono con dos bolas de nieve diferentes si en la nevería cuentan con 5 sabores de nieve de yogurt (frutos secos, zarzamora, nuez, kiwi y pistache).

Solución usando una tabla de entrada múltiple.

		Segunda bola de nieve				
		F	Z	N	K	P
Primer bola de nieve	F	F, F	F, Z	F, N	F, K	F, P
	Z	Z, F	Z, Z	Z, N	Z, K	Z, P
	N	N, F	N, Z	N, N	N, K	N, P
	K	K, F	K, Z	K, N	K, K	K, P
	P	P, F	P, Z	P, N	P, K	P, P

En esta tabla se puede observar que las combinaciones en amarillo deben descartarse porque son conos con un solo sabor y que las combinaciones en verde son iguales a las azules porque el orden en que sirvan las bolas de nieve es indiferente, por lo que zarzamora – Frutos secos es lo mismo de Frutos secos – Zarzamora. Por lo anterior se tiene que son 10 formas diferentes de pedir un cono con dos bolas de nieve de yogurt de diferente sabor.

Solución con formula.

$$nC_x = \frac{n!}{(n-x)!x!}$$

Donde:

n = Total de elementos

x = Cantidad de objetos tomados.

n! = Factorial de n

Recuerde que de 5 sabores (n = 5) quiere tomar 2 (x = 2)

$$5C_2 = \frac{n!}{(n-x)!x!} = \frac{5!}{(5-2)!2!} = \frac{5!}{3!2!} = 10$$

10 formas diferentes de solicitar un cono de dos bolas de nieve de diferente sabor, lo cual coincide perfectamente con la solución obtenida mediante tabla de entrada múltiple.

Ejemplo 2. En el Tacazo cuentan con 4 guisos diferentes (Barbacoa, Tripitas, Colorado y Pasado). ¿De cuantas formas diferentes puede pedir un taco, si es posible que lo pida con 1 solo guiso o combinado con 2, 3 o los 4 guisos?

Combinaciones correspondientes de pedir un solo guiso de los 5 con que cuenta el Tacazo

$${}^5C_1 = \frac{n!}{(n-x)!x!} = \frac{5!}{(5-1)!1!} = \frac{5!}{4!1!} = 5$$

Combinaciones correspondientes de pedir un taco combinado con 2 guisos.

$${}^5C_2 = \frac{n!}{(n-x)!x!} = \frac{5!}{(5-2)!2!} = \frac{5!}{3!2!} = 10$$

Combinaciones correspondientes de pedir un taco combinado con 3 guisos.

$${}^5C_3 = \frac{n!}{(n-x)!x!} = \frac{5!}{(5-3)!3!} = \frac{5!}{2!3!} = 10$$

Combinaciones correspondientes de pedir un taco combinado con 4 guisos.

$${}^5C_4 = \frac{n!}{(n-x)!x!} = \frac{5!}{(5-4)!4!} = \frac{5!}{1!4!} = 5$$

Combinaciones correspondientes de pedir un taco combinado con 5 guisos.

$${}^5C_5 = \frac{n!}{(n-x)!x!} = \frac{5!}{(5-5)!5!} = \frac{5!}{0!5!} = 1$$

La suma de las diferentes combinaciones es la respuesta a la pregunta planteada, siendo así, 31 formas diferentes de pedir un taco.

Ejercicios propuestos.

1.- Un grupo de alumnos del ITCC cuenta con 34 alumnos y un día llega el presidente de la sociedad de alumnos para solicitarles que nombren a 3 de ellos para entregarles un boleto a cada uno de los seleccionados para un concierto en la ciudad de Chihuahua. ¿De cuantas formas diferentes pueden ser entregados los boletos para el concierto?

Respuesta 5984 formas diferentes.

2.- En una frutería hay 10 tipos de frutas diferentes: manzanas, peras, naranjas, plátanos, uvas, sandías, melones, duraznos, fresas y cerezas. Si solo puedes llevar 4 frutas diferentes, ¿de cuántas maneras puedes elegir las?

Respuesta 210 formas diferentes.

3.- Un entrenador tiene 18 jugadores disponibles, pero solo puede seleccionar 11 para que formen el equipo titular. ¿De cuántas maneras puede elegir a los 11 jugadores?

Respuesta 31,824 formas diferentes.

4.- En un concurso de talentos participan 15 personas. De estas, se seleccionarán 5 finalistas para competir en la última etapa. ¿De cuántas formas se pueden elegir a los 5 finalistas?

Respuesta 3,003 formas diferentes.

5.- Tienes 12 libros diferentes, pero solo puedes escoger 6 para colocarlos en un estante. Si el orden no importa, ¿cuántas maneras hay de elegir los libros?

Respuesta 924 formas diferentes.

El teorema del binomio con amplia aplicación en algebra y probabilidad, ofrece una fórmula general para expandir potencias de un binomio $(a + b)^n$, la cual es.

$$(a + b)^n = \sum_{x=1}^k nC_x a^{n-x} b^x$$

En esta fórmula se puede apreciar que implica el uso de combinaciones y se verá en páginas posteriores que es de utilidad en el tema de probabilidad binomial.

Ejemplo 1. Considerando la formula general del binomio, encuentre el resultado para $(x + y)^4$

$$(x + y)^4 = 4C_0 x^4 y^0 + 4C_1 x^3 y^1 + 4C_2 x^2 y^2 + 4C_3 x^1 y^3 + 4C_4 x^0 y^4$$

$$(x + y)^4 = 1x^4 y^0 + 4x^3 y^1 + 6x^2 y^2 + 4x^1 y^3 + 1x^0 y^4$$

$$(x + y)^4 = x^4 + 4x^3 y + 6x^2 y^2 + 4xy^3 + y^4$$

Ejemplo 2. Considerando la formula general del binomio, encuentre el resultado para $(2x + 3y)^4$

$$(2x + 3y)^4 = 4C_0 (2x)^4 (3y)^0 + 4C_1 (2x)^3 (3y)^1 + 4C_2 (2x)^2 (3y)^2 + 4C_3 (2x)^1 (3y)^3 + 4C_4 (2x)^0 (3y)^4$$

$$(2x + 3y)^4 = 16x^4 + 96x^3y + 216x^2y^2 + 216xy^3 + 81y^4$$

Ejercicios propuestos.

1.- Considerando la formula general del binomio, encuentre el resultado para $(x + 2y)^4$

Respuesta $(x + 2y)^4 = x^4 + 8x^3y + 24x^2y^2 + 32xy^3 + 16y^4$

2.- Considerando la formula general del binomio, encuentre el resultado para $(x + y)^5$

Respuesta $(x + y)^5 = x^5 + 5x^4y + 10x^3y^2 + 10x^2y^3 + 5xy^4 + y^5$

3.- Considerando la formula general del binomio, encuentre el resultado para $(2x - y)^6$

Respuesta $(x + y)^6 = x^6 + 6x^5y + 15x^4y^2 + 20x^3y^3 + 15x^2y^4 + 6xy^5 + y^6$

4.- Considerando la formula general del binomio, encuentre el resultado para $(2x - y)^7$

Respuesta $(x + y)^7 = x^7 + 7x^6y + 21x^5y^2 + 35x^4y^3 + 35x^3y^4 + 21x^2y^5 + 7xy^6 + y^7$

5.- Considerando la formula general del binomio, encuentre el resultado para $(x - 3y)^4$

Respuesta $(x - 3y)^4 = x^4 - 12x^3y + 54x^2y^2 - 108xy^3 + 81y^4$

1.6 Uso de software.

A continuación, se presenta una serie de videos en los cuales se explica la aplicación de Excel en la solución de ejercicios representativos de esta unidad.

Permutaciones simples <https://youtu.be/ExajvAe0cw4>

Permutaciones con repetición. <https://youtu.be/w9TMsRM8zS8>

Permutaciones circulares <https://youtu.be/aUEHeGIM06U>

Permutaciones distinguibles <https://youtu.be/522uXHzeSMc>

Combinaciones <https://www.youtube.com/watch?v=sA3Ldh-diic>

A continuación, se presenta una sencilla evaluación de esta unidad de teoría de conjuntos y combinaciones, mediante el siguiente código QR



2. Variables Aleatorias Discretas y sus distribuciones de Probabilidad.

2.1. Variable aleatoria discreta.

En el ámbito de la ciencia de los alimentos, la probabilidad y la estadística desempeñan un papel fundamental en la toma de decisiones y en la comprensión de fenómenos aleatorios. Uno de los conceptos clave en este contexto es el de variable aleatoria discreta. Esta variable, que puede tomar un número finito o contable de valores, es esencial para modelar y analizar diversos aspectos relacionados con la producción, procesamiento y consumo de alimentos.

Una variable es una característica o atributo que puede variar entre diferentes elementos de un conjunto. En el contexto de la estadística, una variable aleatoria es aquella cuyo valor numérico depende del resultado de un experimento realizado completamente al azar. Una variable aleatoria es discreta cuando los valores que puede tomar son aislados y contables, es decir, no hay valores intermedios entre ellos, ejemplo de este tipo de variable son:

1. Número de bacterias por gramo de un alimento: Esta variable puede tomar valores enteros no negativos y varía de acuerdo al tipo de alimento, las condiciones de almacenamiento y otros factores.
2. Número de piezas defectuosas en una línea de producción de alimentos: Esta variable también toma valores enteros no negativos y es crucial para evaluar la calidad del proceso de producción.
3. Número de consumidores que prefieren un nuevo producto: Esta variable es útil para medir la aceptación de un producto en el mercado y puede tomar valores enteros desde cero hasta el número total de consumidores encuestados.
4. Número de días que tarda un alimento en descomponerse: Esta variable puede variar dependiendo de las condiciones de almacenamiento y las características del alimento.
5. Número de ingredientes utilizados en una receta: Esta variable es importante para la formulación de nuevos productos y para el análisis nutricional de los alimentos.

En cada uno de los ejemplos presentados se aprecia claramente que este tipo de variable solo puede tomar valores enteros, lo cual hace la diferencia de una variable aleatoria continua, la cual puede asumir valores fraccionarios, lo cual se verá más a detalle en la unidad 3.

Las distribuciones de probabilidad describen la probabilidad de que una variable aleatoria tome cada uno de sus posibles valores. Algunas de las distribuciones de probabilidad discretas más comunes utilizadas en el ámbito de los alimentos son: Distribución Binomial, Distribución Hipergeométrica, Distribución Geométrica, Distribución Multinomial y Distribución de Poisson, las cuales se verán a detalle en los siguientes subtemas.

2.2. Función de probabilidad y de distribución.

Función de Probabilidad

La función de probabilidad asigna a cada posible valor de una variable aleatoria discreta una probabilidad específica. En otras palabras, nos indica la probabilidad de que la variable aleatoria tome exactamente ese valor (Walpole, Myers, Myers, & Ye, 2012).

Características:

- No negatividad: La probabilidad asignada a cada valor siempre es mayor o igual a cero.
- Suma de probabilidades igual a 1: La suma de todas las probabilidades asignadas a todos los posibles valores de la variable aleatoria debe ser igual a 1.
- Definida para valores discretos: Solo se aplica a variables aleatorias que pueden tomar un conjunto contable de valores.
- Notación: Suele representarse por $P(X = x)$, donde:

P: Probabilidad

X: Variable aleatoria

x: Valor específico que puede tomar la variable aleatoria

Función de distribución

La función de distribución de probabilidad, también conocida como función de distribución acumulada (FDA), nos indica la probabilidad de que una variable aleatoria tome un valor menor o igual a un determinado valor. Es decir, acumula las probabilidades de todos los valores menores o iguales al valor en cuestión (Montgomery & Runger, 2016).

Características:

- No decreciente: La función de distribución siempre es no decreciente, es decir, a medida que aumenta el valor de la variable aleatoria, la probabilidad acumulada también aumenta o permanece igual.
- Valores entre 0 y 1: La función de distribución siempre toma valores entre 0 y 1.
- Límite en el infinito: El límite de la función de distribución cuando la variable tiende a infinito es 1.
- Notación: Suele representarse por $F(x)$, donde:

F: Función de distribución

x: Valor específico

La relación entre ambas funciones es que la función de distribución se construye a partir de la función de probabilidad, acumulando las probabilidades de los valores menores o iguales.

2.2.1. Distribución Binomial.

La distribución binomial es una herramienta estadística fundamental en diversas disciplinas, incluyendo la ciencia de los alimentos. Esta distribución modela el número de éxitos en un número fijo de ensayos de Bernoulli, donde cada ensayo tiene solo dos posibles resultados: éxito o fracaso. En el contexto de los alimentos, la distribución binomial puede ser utilizada para modelar una amplia variedad de situaciones, desde la probabilidad de encontrar bacterias en una muestra de alimentos hasta la proporción de consumidores que prefieren un determinado producto.

Características de la Distribución Binomial

- Número fijo de ensayos (n): El número total de veces que se repite el experimento.
- Probabilidad de éxito constante (p): La probabilidad de obtener un éxito en cada ensayo.
- Ensayos independientes: El resultado de un ensayo no afecta el resultado de los demás.
- Dos posibles resultados: Éxito o fracaso.

Aplicación en el Área de Alimentos

La distribución binomial encuentra numerosas aplicaciones en el control de calidad de alimentos. Por ejemplo, se puede utilizar para:

- Estimar la proporción de productos defectuosos: Al inspeccionar un lote de productos, se puede calcular la probabilidad de encontrar un cierto número de productos con defectos.
- Evaluar la eficacia de un nuevo proceso de producción: Al comparar la proporción de productos que cumplen con las especificaciones antes y después de implementar un nuevo proceso, se puede evaluar su efectividad.
- Determinar el tamaño de muestra necesario: Para estimar una proporción con un determinado nivel de confianza, se puede utilizar la distribución binomial para calcular el tamaño de muestra adecuado.

La fórmula o función de la distribución de probabilidad binomial es la siguiente.

$$P(X = x) = nC_x p^x q^{n-x} \text{ donde:}$$

X es la variable aleatoria

x es el valor específico que puede tomar la variable aleatoria

n es la cantidad de ensayos realizados o tamaño de muestra.

p es la probabilidad de éxitos

q es la probabilidad de fracaso $q = 1 - p$

Ejemplo 1. Supongamos que un fabricante de yogurt afirma que el 95% de sus productos tienen una vida útil de al menos 14 días. Si se toma una muestra aleatoria de 20 yogurts.

A) ¿Cuál es la probabilidad de que exactamente 18 yogurts tengan una vida útil de al menos 14 días?

- B) ¿Cuál es la probabilidad de que 17 o más yogurts tengan una vida útil de al menos 14 días?
 C) ¿Cuál es la probabilidad de que 17 o 19 yogurts tengan una vida útil de al menos 14 días?
 D) ¿Cuál es la probabilidad de que menos de 19 yogurts tengan una vida útil de al menos 14 días?

Solución.

- A) ¿Cuál es la probabilidad de que exactamente 18 yogurts tengan una vida útil de al menos 14 días?

$n = 20$ (número de yogurts en la muestra)

$p = 0.95$ (probabilidad de que un yogurt tenga una vida útil de al menos 14 días)

$x = 18$ (cantidad de yogurts con una vida útil de al menos 14 días)

$$P(X = x) = nC_x p^x q^{n-x}$$

$$P(X = 18) = 20C_{18} 0.95^{18} 0.05^{20-18} = 190 * 0.95^{18} * 0.05^2 = 0.1887$$

Interpretación. Existe un 18.87 % de probabilidades que suceda que 18 yogurts tengan una vida útil de al menos 14 días.

- B) ¿Cuál es la probabilidad de que 17 o más yogurts tengan una vida útil de al menos 14 días?

$n = 20$ (número de yogurts en la muestra)

$p = 0.95$ (probabilidad de que un yogurt tenga una vida útil de al menos 14 días).

$x \geq 17$ (cantidad de yogurts con una vida útil de al menos 14 días)

$$P(X = x) = nC_x p^x q^{n-x}$$

$$P(X = 17) = 20C_{17} 0.95^{17} 0.05^3 = 0.0596$$

$$P(X = 18) = 20C_{18} 0.95^{18} 0.05^2 = 0.1887$$

$$P(X = 19) = 20C_{19} 0.95^{19} 0.05^1 = 0.3774$$

$$P(X = 20) = 20C_{20} 0.95^{20} 0.05^0 = 0.3585$$

$$P(x \geq 17) = P(X = 17) + P(X = 18) + P(X = 19) + P(X = 20) = 0.9841$$

Interpretación. Existe un 98.41 % de probabilidades que suceda que 17 o más yogurts tengan una vida útil de al menos 14 días.

- C) ¿Cuál es la probabilidad de que 17 o 19 yogurts tengan una vida útil de al menos 14 días?

$$P(X = 17 \text{ o } X = 19) = P(X = 17) + P(X = 19) = 0.0596 + 0.3774 = 0.4369$$

Interpretación. Existe un 43.69 % de probabilidades que suceda que 17 o 19 yogurts tengan una vida útil de al menos 14 días.

D) ¿Cuál es la probabilidad de que menos de 19 yogurts tengan una vida útil de al menos 14 días?

Como la suma de las probabilidades de $x = 0$ hasta $X = 20$ es 1 (100%), la respuesta a este inciso es.

$$P(X < 19) = 1 - P(X = 20) - P(X = 19) = 1 - 0.3585 - 0.3774 = 0.2642$$

Interpretación. Existe un 26.4 % de probabilidades que suceda que menos de 19 yogurts tengan una vida útil de al menos 14 días.

La solución de este ejercicio se puede observar en el siguiente link <https://youtu.be/xiDosEwPHmo>

Ejemplo 2. Un fabricante de galletas afirma que el 10% de sus galletas contienen más de 300 calorías. Si se selecciona una muestra aleatoria de 9 galletas

- A) ¿Cuál es la probabilidad de que ninguna galleta contenga más de 300 calorías?
- B) ¿Cuál es la probabilidad de que 10 galletas contengan más de 300 calorías?
- C) ¿Cuál es la probabilidad de que 3 galletas contengan más de 300 calorías?
- D) ¿Cuál es la probabilidad de que 2 a 5 galletas contengan más de 300 calorías?
- E) ¿Cuál es la probabilidad de que menos de 3 galletas contengan más de 300 calorías?
- F) ¿Cuál es la probabilidad de que más de 2 galletas contengan más de 300 calorías?

Solución.

- A) ¿Cuál es la probabilidad de que ninguna galleta contenga más de 300 calorías?

$$P(X = 0) = {}^9C_0 0.1^0 0.9^9 = 0.3874$$

Existe 3874 % de probabilidades de que ninguna galleta contenga más de 300 calorías

- B) ¿Cuál es la probabilidad de que 10 galletas contengan más de 300 calorías?

Para este caso la probabilidad es 0.00, por el hecho de que el tamaño de muestra es de 9 y resulta imposible que 10 galletas contengan más de 300 calorías

- C) ¿Cuál es la probabilidad de que 3 galletas contengan más de 300 calorías?

$$P(X = 3) = {}^9C_3 0.1^3 0.9^6 = 0.0446$$

Existe 4.46 % de probabilidades de que 3 galletas contengan más de 300 calorías

- D) ¿Cuál es la probabilidad de que 2 a 5 galletas contengan más de 300 calorías?

$$P(2 \leq X \leq 5) = P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5)$$

$$P(X = 2) = {}^9C_2 0.1^2 0.9^7 = 0.1722$$

$$P(X = 3) = {}^9C_3 0.1^3 0.9^6 = 0.0446$$

$$P(X = 4) = {}^9C_3 0.1^4 0.9^5 = 0.0074$$

$$P(X = 5) = {}^9C_3 0.1^5 0.9^4 = 0.0008$$

$$P(2 \leq X \leq 5) = 0.1722 + 0.0446 + 0.0074 + 0.0008 = 0.2251$$

Existe 22.51 % de probabilidades de que 2 a 5 galletas contengan más de 300 calorías

E) ¿Cuál es la probabilidad de que menos de 3 galletas contengan más de 300 calorías?

$$P(X < 3) = P(X = 0) + P(X = 1) + P(X = 2)$$

$$P(X = 0) = {}^9C_0 0.1^0 0.9^9 = 0.3874$$

$$P(X = 1) = {}^9C_1 0.1^1 0.9^8 = 0.3874$$

$$P(X = 2) = {}^9C_2 0.1^2 0.9^7 = 0.1722$$

$$P(X < 3) = 0.387 + 0.3874 + 0.1722 = 0.9470$$

Existe 94.70 % de probabilidades de que menos de 3 galletas contengan más de 300 calorías

F) ¿Cuál es la probabilidad de que más de 2 galletas contengan más de 300 calorías?

$$P(X > 2) = 1 - P(X = 0) + P(X = 1) + P(X = 2)$$

$$P(X > 2) = 1 - 0.9470 = 0.0530$$

Existe 5.30 % de probabilidades de que más de 2 galletas contengan más de 300 calorías

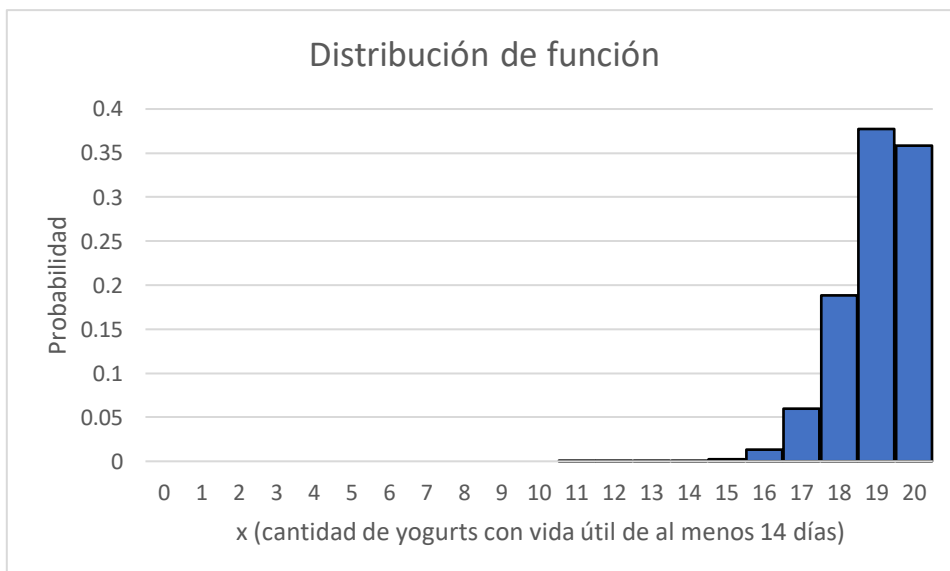
Elaboración de la distribución de función y la distribución de función acumulada.

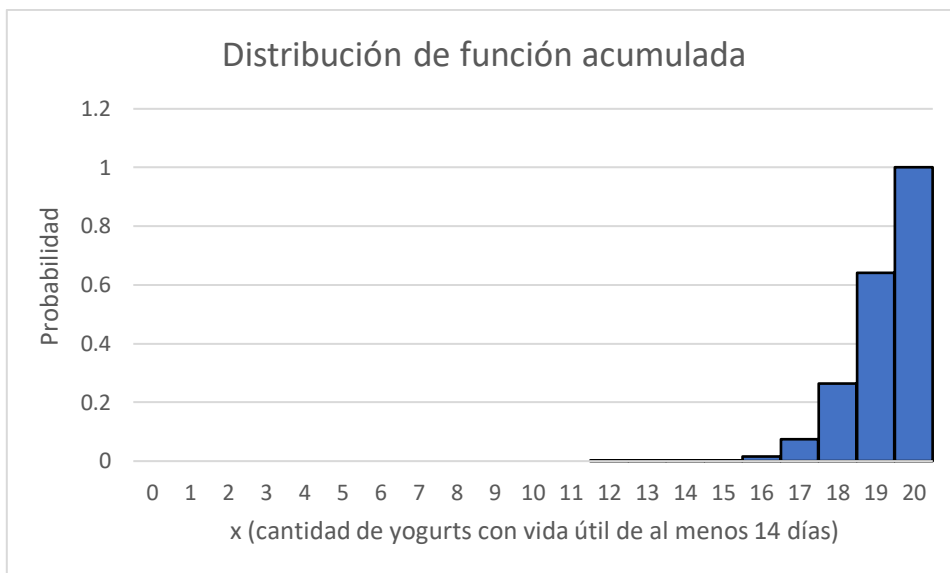
El primer paso consiste en calcular la probabilidad para los valores de $x = 0$ hasta $x = n$, lo cual es conveniente realizarlo de forma tabular, luego obtener la probabilidad acumulada para cada valor de x , posteriormente elaborar la gráfica de las probabilidades marginales para obtener la distribución de función y finalmente graficar las probabilidades acumuladas para obtener la distribución de función acumulada.

Ejemplo 1. Elabore la distribución de función y la distribución de función acumulada para el ejercicio correspondiente al yogurt.

El procedimiento consiste en calcular cada una de las probabilidades marginales desde x igual a cero hasta n , luego graficar las probabilidades marginales obtenidas para obtener así la distribución de función. Posteriormente se calculan las probabilidades acumuladas para cada valor de x , luego se grafican los valores obtenidos y se obtiene así la distribución de función acumulada. Los resultados de ello se presentan a continuación.

x	P(x)
0	9.53674E-27
1	3.62396E-24
2	6.54125E-22
3	7.45703E-20
4	6.02155E-18
5	3.6611E-16
6	1.73902E-14
7	6.60829E-13
8	2.04031E-11
9	5.16878E-10
10	1.08028E-08
11	1.86593E-07
12	2.65895E-06
13	3.10893E-05
14	0.000295348
15	0.002244646
16	0.013327586
17	0.059582148
18	0.188676801
19	0.377353603
20	0.358485922





El procedimiento para la realización de estas distribuciones de probabilidad en Excel se puede observar en el link <https://youtu.be/E4Q3gBAAtqO4>

Lo anterior con GeoGebra se puede observar en el link <https://youtu.be/RNdDnz-sDMU>

Ejercicios propuestos.

La solución de estos ejercicios en Excel se puede consultar dando clic en el link [ejercicios de probabilidad binomial.](#)

Ejercicio 1: Producción de Galletas

Una máquina produce galletas, y se sabe que el 5% de ellas salen defectuosas. Se toma una muestra aleatoria de 10 galletas.

- ¿Cuál es la probabilidad de que exactamente 2 galletas estén defectuosas? Respuesta: 0.0746
- ¿Cuál es la probabilidad de que a lo más 1 galleta esté defectuosa? Respuesta: 0.9139
- ¿Cuál es la probabilidad de que al menos 3 galletas estén defectuosas? Respuesta: 0.0115
- ¿Cuál es la probabilidad de que ninguna galleta esté defectuosa? Respuesta: 0.5987
- Elaborar la función de probabilidad y la función de distribución.

Ejercicio 2: Empaquetado de Frutas

Un empaquetador de frutas afirma que el 90% de las bolsas de manzanas contienen al menos 5 manzanas rojas. Se seleccionan al azar 8 bolsas.

- ¿Cuál es la probabilidad de que exactamente 7 bolsas cumplan con el requisito? Respuesta: 0.3826
- ¿Cuál es la probabilidad de que todas las bolsas cumplan con el requisito? Respuesta: 0.4305

- c) ¿Cuál es la probabilidad de que menos de 6 bolsas cumplan con el requisito? Respuesta: 0.0381
- d) ¿Cuál es la probabilidad de que 6 o 7 bolsas cumplan con el requisito? Respuesta: 0.5314
- e) Elaborar la función de probabilidad y la función de distribución.

Ejercicio 3: Control de Calidad en una Embotelladora

Una embotelladora de jugo asegura que el 98% de sus botellas están correctamente selladas. Se selecciona una muestra aleatoria de 20 botellas.

- a) ¿Cuál es la probabilidad de que exactamente 19 botellas estén bien selladas? Respuesta: 0.2725
- b) ¿Cuál es la probabilidad de que todas las botellas estén bien selladas? Respuesta: 0.6676
- c) ¿Cuál es la probabilidad de que a lo más 18 botellas estén bien selladas? Respuesta: 0.0599
- d) ¿Cuál es la probabilidad de que al menos 3 botellas no estén bien selladas? Respuesta: 0.0071
- e) Elaborar la función de probabilidad y la función de distribución.

Ejercicio 4: Preferencias de Consumidores

Una encuesta revela que el 60% de los consumidores prefiere el yogurt natural sin azúcar. Se selecciona una muestra aleatoria de 12 consumidores.

- a) ¿Cuál es la probabilidad de que exactamente 8 consumidores prefieran el yogurt natural? Respuesta: 0.2128
- b) ¿Cuál es la probabilidad de que menos de 5 consumidores prefieran el yogurt natural? Respuesta: 0.0573
- c) ¿Cuál es la probabilidad de que más de 10 consumidores prefieran el yogurt natural? Respuesta: 0.0196
- d) ¿Cuál es la probabilidad de que al menos 7 consumidores prefieran el yogurt natural? Respuesta: 0.6652
- e) Elaborar la función de probabilidad y la función de distribución.

Ejercicio 5: Semillas de Tomate

Un paquete de semillas de tomate indica que tiene una tasa de germinación del 85%. Se siembran 15 semillas.

- a) ¿Cuál es la probabilidad de que exactamente 12 semillas germinen? Respuesta: 0.2184
- b) ¿Cuál es la probabilidad de que menos de 10 semillas germinen? Respuesta: 0.0168
- c) ¿Cuál es la probabilidad de que más de 13 semillas germinen? Respuesta: 0.3186
- d) ¿Cuál es la probabilidad de que al menos 11 semillas germinen? Respuesta: 0.9383
- e) Elaborar la función de probabilidad y la función de distribución.

2.2.2. Distribución Hipergeométrica.

La distribución hipergeométrica es una herramienta estadística que encuentra aplicación en diversas áreas, incluyendo la industria de alimentos, donde puede emplearse para el control de calidad y la evaluación de lotes de productos (Benítez, 2018). A diferencia de la distribución binomial, la hipergeométrica se utiliza cuando se realiza un muestreo sin reemplazo, lo que es común en escenarios donde los recursos son limitados o se desea mantener la integridad del lote examinado.

En el sector alimentario, asegurar la calidad de los productos es crucial para satisfacer las normativas regulatorias y las expectativas de los consumidores (García, 2021). Supongamos que un inspector de calidad revisa un lote de 100 paquetes de harina de trigo, de los cuales se sabe que 10 contienen contaminantes. Si el inspector selecciona 15 paquetes al azar sin reemplazo, la distribución hipergeométrica permite calcular la probabilidad de encontrar un número específico de paquetes contaminados en la muestra. Este enfoque facilita la toma de decisiones sobre la aceptación o rechazo de un lote.

La función de probabilidad de esta distribución se define como:

$$P(X = x) = \frac{K C_x (N - k) C_{(n-x)}}{N C_n}$$

donde:

N es el tamaño total de la población (en el ejemplo, 100 paquetes).

K es el número de éxitos en la población (10 paquetes contaminados).

n es el tamaño de la muestra (15 paquetes seleccionados).

x es el número de éxitos observados en la muestra.

En esta distribución la proporción de éxito p es: $p = \frac{K}{N}$ y la probabilidad de fracaso q es $q = 1 - p$

Adicional a lo anterior se tiene que la media, la varianza y la desviación estándar son:

$$\mu = n \frac{K}{N}, \quad \sigma^2 = npq \frac{N - n}{N - 1}, \quad \sigma = \sqrt{npq \frac{N - n}{N - 1}}$$

Continuando con el ejemplo, podría surgir la pregunta ¿Cuál es la probabilidad de que, en la muestra de 15 paquetes, 4 de ellos estén contaminados?

Aplicando la formula, la respuesta sería.

$$P(X = 4) = \frac{10 C_4 (100 - 10) C_{(15-4)}}{100 C_{15}} = \frac{10 C_4 90 C_{11}}{100 C_{15}} = \frac{10 C_4 90 C_{11}}{100 C_{15}} = 0.0345$$

Existe un 3.45% de probabilidades de que, en la muestra de 15 paquetes de harina, 4 de ellos estén contaminados.

Ejemplos adicionales de la distribución de probabilidad hipergeométrica se presentan a continuación.

Ejemplo 1.- Don Juan, el dueño de la Frutería La favorita, al estar limpiando unas cajas de naranjas encontró que, de las 135 naranjas, 20 de ellas estaban podridas. Si usted hubiera tomado al azar 10 naranjas.

- ¿Cuál es la probabilidad de que exactamente 3 naranjas estuvieran podridas?
- ¿Cuál es la probabilidad de que a lo más 2 naranjas estuvieran podridas?
- ¿Cuál es la probabilidad de que al menos 4 naranjas estuvieran podridas?
- Calcular la media, la varianza y la desviación estándar.
- Elaborar la distribución de función y la distribución de función acumulada.

Solución.

- ¿Cuál es la probabilidad de que exactamente 3 naranjas estuvieran podridas?

$$P(X = x) = \frac{KC_x (N - k)C_{(n-x)}}{NC_n} = \frac{20C_3 (135 - 20)C_{(10-3)}}{135C_{10}} = \frac{20C_3 115C_7}{135C_{10}} = 0.1268$$

Existe 12.68 % de probabilidades de que de las 10 naranjas tomadas al azar 3 de ellas estuvieran podridas.

- ¿Cuál es la probabilidad de que a lo más 2 naranjas estuvieran podridas?

Considerando que, a lo más 2 naranjas podridas, implica que dentro del conjunto solución se encuentran el valor de 0, 1 y 2 naranjas, por lo que, es necesario calcular cada uno de esos valores, luego sumarlos y llegar así a la solución.

$$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)$$

$$P(X = 0) = \frac{KC_x (N - k)C_{(n-x)}}{NC_n} = \frac{20C_0 115C_{10}}{135C_{10}} = 0.1893$$

$$P(X = 1) = \frac{KC_x (N - k)C_{(n-x)}}{NC_n} = \frac{20C_1 115C_9}{135C_{10}} = 0.3571$$

$$P(X = 2) = \frac{KC_x (N - k)C_{(n-x)}}{NC_n} = \frac{20C_2 115C_8}{135C_{10}} = 0.2854$$

$$P(X \leq 2) = 0.1893 + 0.3571 + 0.2854 = 0.8318$$

Existe 83.18 % de probabilidades de que a lo más 2 naranjas estuvieran podridas.

c) ¿Cuál es la probabilidad de que al menos 4 naranjas estuvieran podridas?

La expresión al menos 4 implica los valores de 4 hasta 10, por lo tanto, la solución es que a 1 (100%), se le reste las probabilidades de x igual a 0, 1, 2 y 3.

$$P(X \geq 4) = 1 - (P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3))$$

$$P(X \geq 4) = 1 - (0.1893 + 0.3571 + 0.2854 + 0.1268)$$

$$P(X \geq 4) = 1 - (0.9586) = 0.0414$$

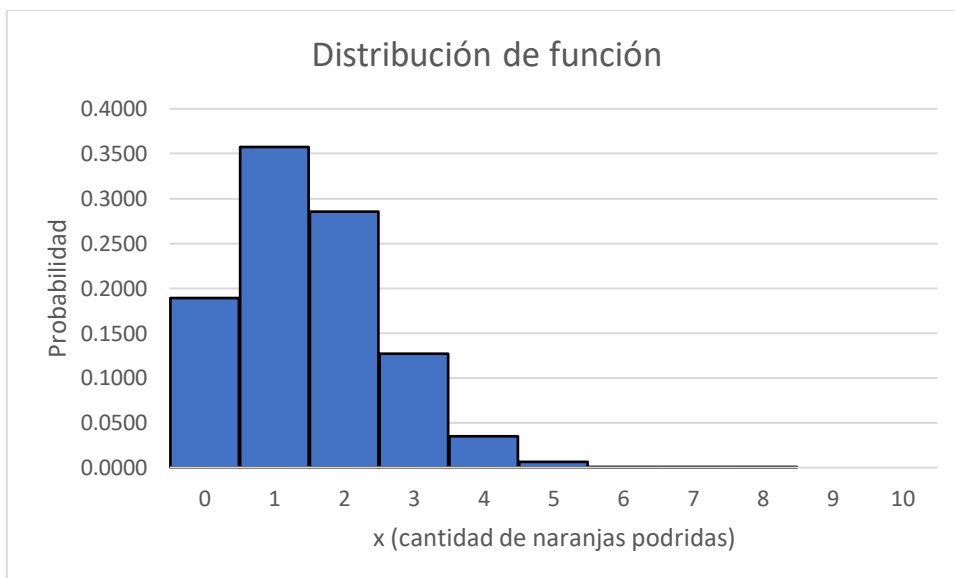
Existe 4.14 % de probabilidades de que al menos 4 naranjas estuvieran podridas.

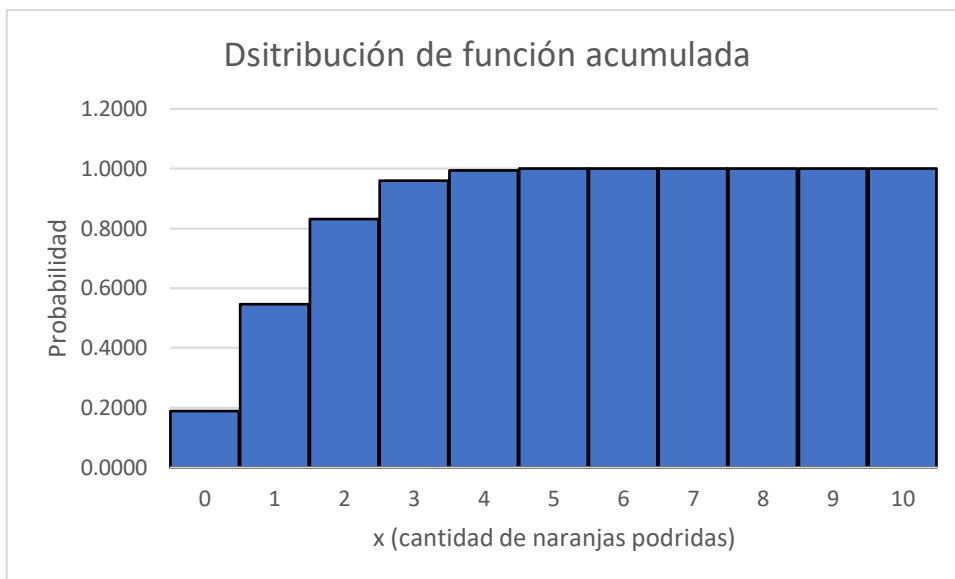
d) Calcular la media, la varianza y la desviación estándar.

Media	1.4815
Varianza	1.1772
Desviación estándar	1.0850

Ver la solución de estos incisos en Excel en el link solución de <https://youtu.be/H7gAtt10LnY>

e) Elaborar la función de probabilidad y la función de distribución.





Ver procedimiento de elaboración de estas distribuciones de función en el link <https://youtu.be/Aoyx9Zn8TIM>.

Las distribuciones anteriores realizadas con GeoGebra pueden observarse en el link <https://youtu.be/iRKm8KEssnE>

Ejemplo 2. Considerando que una caja contiene 60 galletas, de las cuales 7 están rotas. Si se seleccionan al azar 8 galletas.

- ¿Cuál es la probabilidad de que ninguna galleta esté rota?
- ¿Cuál es la probabilidad de que exactamente 2 galletas estén rotas?
- ¿Cuál es la probabilidad de que las 8 galletas estén rotas?
- Calcular la media, la varianza y la desviación estándar.
- Elaborar la función de probabilidad y la función de distribución.

Solución.

- ¿Cuál es la probabilidad de que ninguna galleta esté rota?

$$P(X = 0) = \frac{K C_x (N - k) C_{(n-x)}}{N C_n} = \frac{7 C_0 53 C_8}{60 C_8} = 0.3464$$

Existe 34.64 % de probabilidades de que ninguna galleta esté rota.

- ¿Cuál es la probabilidad de que exactamente 2 galletas estén rotas?

$$P(X = 2) = \frac{K C_x (N - k) C_{(n-x)}}{N C_n} = \frac{7 C_2 53 C_6}{60 C_8} = 0.1884$$

Existe 18.84 % de probabilidades de que exactamente 2 galletas estén rotas.

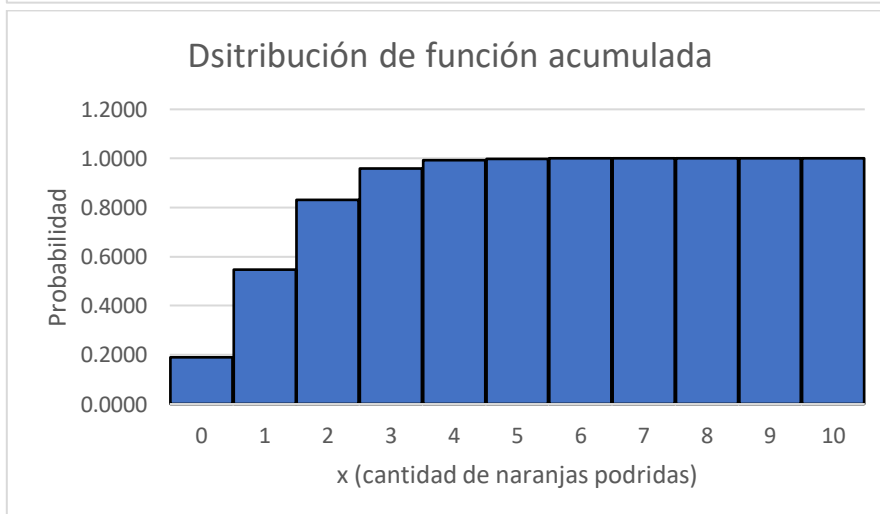
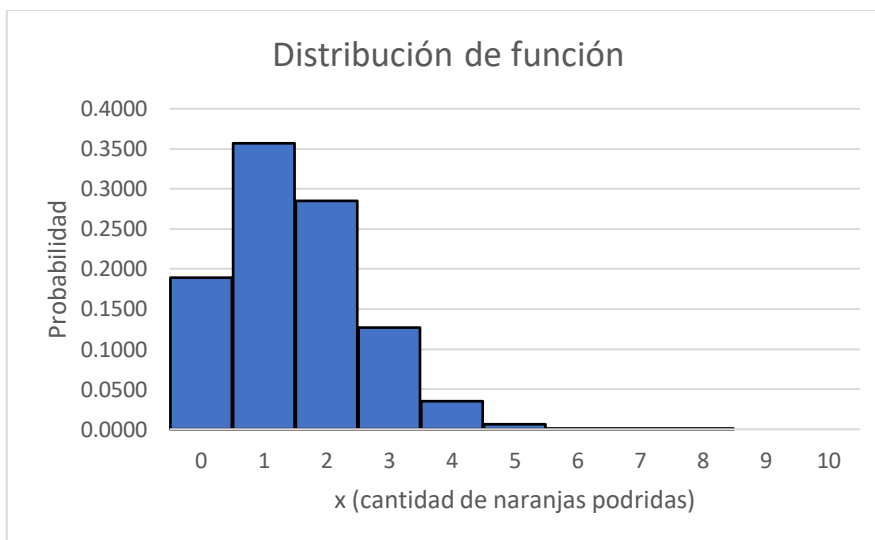
c) ¿Cuál es la probabilidad de que las 8 galletas estén rotas?

Considerando que en la caja de galletas solo hay 7 galletas rotas, es imposible que en la muestra extraída de 8 galletas, las 8 estén rotas, por lo tanto, es innecesario aplicar la formula, por lo cual se dice que existe 0 % de probabilidades de que 8 galletas estén rotas.

d) Calcular la media, la varianza y la desviación estándar.

Media	0.9333
Varianza	0.8226
Desviación estándar	0.9070

e) Elaborar la función de probabilidad y la función de distribución.



Problemas propuestos.

Ver la solución de estos ejercicios en Excel dando clic en el link [ejercicios de probabilidad hipergeométrica](#).

1.- Un lote de 500 latas de atún contiene 5 latas con una abolladura. Se seleccionan al azar 25 latas para una inspección.

- a) ¿Cuál es la probabilidad de que ninguna lata tenga una abolladura? Respuesta: 0.7730
- b) ¿Cuál es la probabilidad de que exactamente una lata tenga una abolladura? Respuesta: 0.2051
- c) ¿Cuál es la probabilidad de que al menos 2 latas tengan una abolladura? Respuesta: 0.0219
- d) Calcular la media, la varianza y la desviación estándar.
- e) Elaborar la función de probabilidad y la función de distribución.

2.- Un lote de 200 piezas de carne contiene 5 piezas contaminadas. Se seleccionan al azar 10 piezas para un análisis.

- a) ¿Cuál es la probabilidad de que una pieza esté contaminada? Respuesta: 0.2075
- b) ¿Cuál es la probabilidad de que 3 o 4 piezas estén contaminadas? Respuesta: 0.0009
- c) ¿Cuál es la probabilidad de que menos 2 piezas estén contaminadas? Respuesta: 0.9792
- d) Calcular la media, la varianza y la desviación estándar.
- e) Elaborar la función de probabilidad y la función de distribución.

3.- Un paquete de 50 semillas contiene 4 semillas que no germinarán. Se siembran 12 semillas.

- a) ¿Cuál es la probabilidad de que todas las semillas germinen? Respuesta: 0.3205
- b) ¿Cuál es la probabilidad de que exactamente una semilla no germine? Respuesta: 0.4396
- c) ¿Cuál es la probabilidad de que al menos 2 semillas no germinen? Respuesta: 0.2399
- d) Calcular la media, la varianza y la desviación estándar.
- e) Elaborar la función de probabilidad y la función de distribución.

4.- Un lote de 150 chocolates contiene 10 chocolates con un relleno diferente. Se seleccionan al azar 20 chocolates.

a) ¿Cuál es la probabilidad de que ninguno de los chocolates tenga un relleno diferente?
Respuesta: 0.2278

b) ¿Cuál es la probabilidad de que exactamente 2 chocolates tengan un relleno diferente?
Respuesta: 0.2639

c) ¿Cuál es la probabilidad de que al menos 3 chocolates tengan un relleno diferente?
Respuesta: 0.1319

d) Calcular la media, la varianza y la desviación estándar.

e) Elaborar la función de probabilidad y la función de distribución.

5.- Un lote de 300 botes de conserva contiene 8 caducos. Se seleccionan al azar 15 botes.

a) ¿Cuál es la probabilidad de que ninguno de los botes este caduco? Respuesta: 0.6601

b) ¿Cuál es la probabilidad de que exactamente un bote este caduco? Respuesta: 0.2849

c) ¿Cuál es la probabilidad de que menos de 2 botes este caduco? Respuesta: 0.9451

d) Calcular la media, la varianza y la desviación estándar.

e) Elaborar la función de probabilidad y la función de distribución.

2.2.3. Distribución Geométrica.

La distribución geométrica, una herramienta fundamental en la teoría de probabilidad, encuentra un amplio espectro de aplicaciones en la industria alimentaria, particularmente en el ámbito del control de calidad. Esta distribución discreta modela la probabilidad de que el primer "éxito" (como encontrar un producto defectuoso) ocurra en el ensayo número k , bajo la condición de que cada ensayo sea independiente y la probabilidad de éxito (p) se mantenga constante.

Una de las aplicaciones más directas de la distribución geométrica en la industria alimentaria es la evaluación de la calidad de lotes de productos. Al examinar una línea de producción, por ejemplo, se puede modelar el número de unidades que deben inspeccionarse antes de encontrar la primera con algún defecto. Esta información es crucial para establecer planes de muestreo eficientes y garantizar que los productos que llegan al consumidor cumplan con los estándares de calidad.

Otro escenario común es el análisis de la vida útil de los alimentos. La distribución geométrica puede utilizarse para modelar el número de días que transcurren hasta que el primer producto de un lote se eche a perder. Esto es especialmente relevante para productos perecederos como frutas y verduras, donde la frescura es un factor determinante para el consumidor.

La fórmula de la distribución de probabilidad geométrica es.

$$P(X = k) = (1 - p)^{k-1} p \text{ donde:}$$

k corresponde en número de ensayo en que se presenta el primer éxito.

p es la probabilidad de éxito.

Los parámetros de media, varianza y desviación estándar para este tipo de distribución son:

$$\mu = \frac{1}{p} \quad \sigma^2 = \frac{1-p}{p^2} \quad \sigma = \sqrt{\frac{1-p}{p^2}} \text{ respectivamente.}$$

La media (o valor esperado) en una distribución de probabilidad geométrica se interpreta como el número promedio de intentos necesarios para obtener el primer éxito en un experimento de Bernoulli repetido de manera independiente, con una probabilidad constante de éxito en cada intento.

Si p es alta, el éxito es más probable en los primeros intentos, y la media será menor. Por ejemplo, si $p = 0.5$, en promedio se necesitarán 2 intentos para obtener el primer éxito.

Si p es baja, el éxito es menos probable en cada intento, por lo que se necesitarán más intentos en promedio. Por ejemplo, si $p=0.1$, se esperaría un éxito aproximadamente cada 10 intentos.

Por otra parte, se tiene que la varianza mide la dispersión de los resultados en torno a la media, es decir, qué tan variables son los intentos necesarios para obtener el primer éxito.

Si p es alta, la varianza será menor, lo que significa que el número de intentos hasta el primer éxito estará más concentrado cerca de la media. Por ejemplo, si $p = 0.8$, la probabilidad de éxito en cada intento es alta, y habrá poca variabilidad en el número de intentos necesarios.

Si p es baja, la varianza será mayor, lo que indica una mayor dispersión en el número de intentos necesarios para obtener el primer éxito. Por ejemplo, si $p=0.1$, los intentos necesarios pueden variar ampliamente, ya que el éxito es mucho menos probable en cada intento.

En una distribución de probabilidad geométrica, la desviación estándar al igual que la varianza, es una medida de dispersión que indica cuánto varían los resultados (el número de intentos necesarios para obtener el primer éxito) respecto a la media.

La desviación estándar σ (sigma) tiene la misma unidad que la media, lo que facilita la interpretación práctica en comparación con la varianza.

Una desviación estándar más grande indica una mayor dispersión en los intentos necesarios para obtener el primer éxito.

Para valores altos de p (mayor probabilidad de éxito en cada intento), σ es pequeña, lo que indica que los intentos hasta el primer éxito tienden a estar más concentrados cerca de la media.

Para valores bajos de p (menor probabilidad de éxito), σ es grande, lo que refleja que los intentos están más dispersos y varían considerablemente.

A continuación, se presentan ejemplos de la distribución de probabilidad geométrica.

Ejemplo 1. Una fábrica de chocolates produce barras de chocolate con almendras. Según los registros históricos, el 2% de las barras presentan menos almendras de las indicadas en la etiqueta. Si se inspeccionan las barras de chocolate de manera aleatoria.

- a) ¿Cuál es la probabilidad de que la tercera barra de chocolate inspeccionada tenga menos almendras que las declaradas en la etiqueta?
- b) ¿Cuál es la probabilidad de que la quinta barra de chocolate inspeccionada tenga menos almendras que las declaradas en la etiqueta?
- c) Calcule los parámetros media, varianza y desviación estándar.
- d) Elabore la función de probabilidad y la función de distribución.

Solución.

- a) ¿Cuál es la probabilidad de que la tercera barra de chocolate inspeccionada tenga menos almendras que las declaradas en la etiqueta?

Aplicando el principio multiplicativo analizado en la unidad anterior se tendría que la primera y segunda barra de chocolate si cumplen con la especificación de la cantidad de almendras y que la tercera barra de chocolate no cumple.

Aplicando la formula

$$P(X = k) = (1 - p)^{k-1} p$$

$$P(X = 3) = (1 - .02)^{3-1} .02 = .98^2 .02 = 0.0192$$

Existe un 1.92 % de probabilidades de que la tercera barra no cumpla con las especificaciones de la cantidad de almendras.

De lo anterior se desprende que el uso de la fórmula es confiable y sencilla de aplicar.

- b) ¿Cuál es la probabilidad de que la quinta barra de chocolate inspeccionada tenga menos almendras que las declaradas en la etiqueta?

$$P(X = 5) = .98^4 .02 = 0.0184$$

Existe un 1.84 % de probabilidades de que la quinta barra no cumpla con las especificaciones de la cantidad de almendras.

- c) Calcule los parámetros media, varianza y desviación estándar.

$$\mu = \frac{1}{p} = \frac{1}{0.02} = 50$$

$$\sigma^2 = \frac{1-p}{p^2} = \frac{1-0.02}{0.02^2} = 2450$$

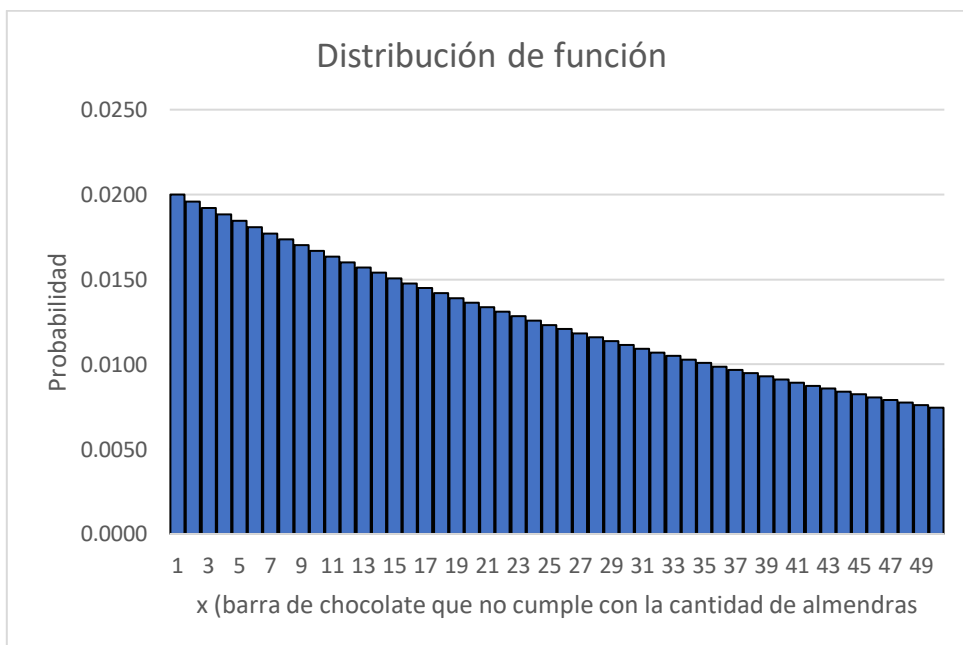
$$\sigma = \sqrt{\frac{1-p}{p^2}} = \sqrt{2450} = 49.5$$

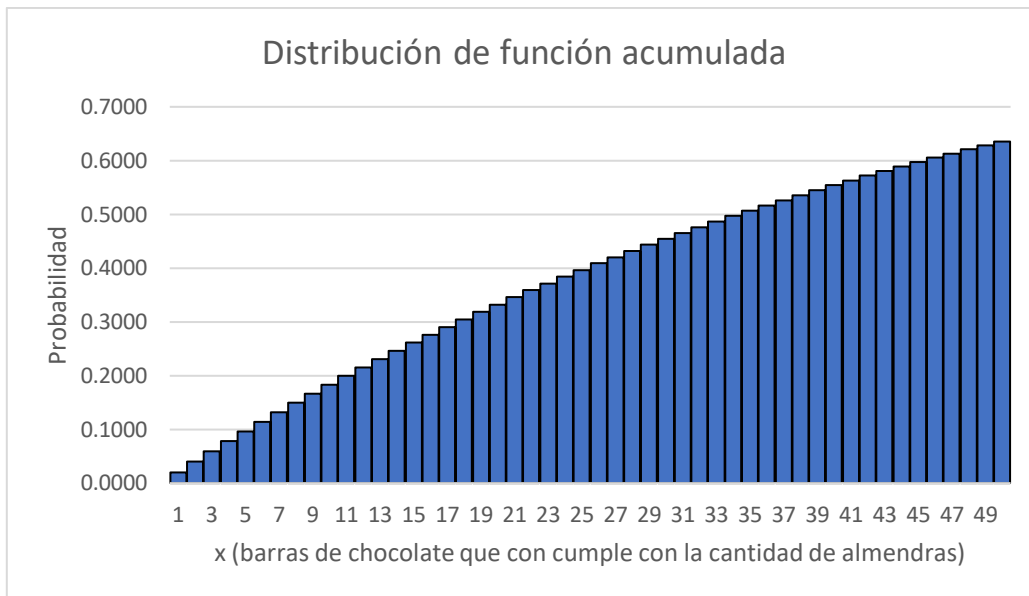
La media de 50 indica que es el número promedio de intentos necesarios para obtener el primer éxito.

La varianza y la desviación estándar de 2450 y 49.5 respectivamente indican la variabilidad en cuanto al número de intentos necesarios para obtener el primer éxito, con respecto a la media.

La solución de este ejercicio con Excel se puede observar en el video del siguiente link <https://youtu.be/dyktQaKQV5U>

d) Elabore la función de probabilidad y la función de distribución.





El procedimiento de elaboración de estas distribuciones en Excel se puede observar en el link <https://youtu.be/csUOdgkiyLU>

Problemas propuestos.

Ver la solución de estos ejercicios en Excel dando clic en el link [ejercicios de probabilidad geométrica](#).

Ejercicio 1.- En una fábrica de alimentos enlatados, cada lata de sopa tiene una probabilidad de $p = 0.95$ de cumplir con los estándares de calidad. Los inspectores revisan las latas una por una hasta encontrar la primera que cumple con los estándares.

- ¿Cuál es la probabilidad de que la primera lata que cumple con los estándares sea la tercera revisada?
- Calcule e interprete la media, la varianza y la desviación estándar.
- Elabore la función de probabilidad y la función de distribución.

Ejercicio 2.- Un panificador está ajustando su máquina para hornear barras de pan integral. Cada barra horneada tiene una probabilidad de $p = 0.9$ de salir con el peso correcto. Hornean una barra a la vez y la pesan hasta encontrar la primera con el peso adecuado.

- Si el panadero encuentra la primera barra con el peso adecuado en el quinto intento, ¿cuál es la probabilidad de que esto ocurra?
- Calcule e interprete la media, la varianza y la desviación estándar.
- Elabore la función de probabilidad y la función de distribución.

Ejercicio 3.- En una línea de empaquetado, cada paquete de galletas tiene una probabilidad de $p = 0.85$ de estar sellado correctamente. Los inspectores revisan paquetes hasta encontrar el primero con un sello adecuado.

- a) ¿Cuál es la probabilidad de que el primer paquete correctamente sellado se encuentre en el cuarto intento?
- b) Calcule e interprete la media, la varianza y la desviación estándar.
- c) Elabore la función de probabilidad y la función de distribución.

2.2.4. Distribución Multinomial.

En el sector de alimentos, la toma de decisiones basada en datos probabilísticos es esencial para garantizar la calidad, optimizar procesos y satisfacer las preferencias de los consumidores. Una herramienta matemática relevante en este contexto es la distribución de probabilidad multinomial, que extiende la binomial para analizar situaciones donde los resultados pueden clasificarse en más de dos categorías.

La distribución multinomial describe la probabilidad de observar un conjunto de resultados categóricos en n ensayos independientes, donde cada resultado pertenece a una de las k categorías posibles con probabilidades fijas $p_1, p_2, \dots, p_k$ (Triola, 2018).

Situaciones reales de la aplicación de la distribución de probabilidad relacionadas con la industria de alimentaria pueden ser.

Control de Calidad en Procesos de Producción. En el control de calidad, la distribución multinomial permite clasificar productos en múltiples categorías, como "sin defectos", "defectos menores" y "defectos graves". Por ejemplo, en una línea de producción de galletas, cada galleta inspeccionada puede clasificarse según su estado. Si la proporción esperada de galletas sin defectos es $p_1 = 0.9$, con defectos menores $p_2 = 0.07$, y con defectos graves $p_3 = 0.03$, la distribución multinomial puede modelar las probabilidades de obtener ciertos conteos de cada categoría en una muestra (Montgomery, 2020).

Análisis de Preferencias del Consumidor. En estudios de mercado, las preferencias de los consumidores se agrupan en diferentes categorías, como "dulce", "salado", o "agridulce". La distribución multinomial es adecuada para modelar los resultados de encuestas donde los encuestados eligen su sabor preferido. Por ejemplo, una encuesta a 100 consumidores sobre un nuevo producto puede revelar que 40 prefieren "dulce", 35 "salado" y 25 "agridulce". Este análisis permite ajustar la producción para satisfacer la demanda.

Una de las principales ventajas de la distribución multinomial es su capacidad para manejar múltiples categorías simultáneamente, proporcionando una visión completa del fenómeno estudiado. Sin embargo, requiere que los ensayos sean independientes y que las

probabilidades de las categorías sean constantes, lo que puede ser un desafío en situaciones reales donde estas condiciones no siempre se cumplen.

La función de probabilidad para esta distribución se expresa como:

$$P(X_1 = x_1, X_2 = x_2, \dots, X_k = x_k) = \frac{n!}{x_1! x_2! \dots x_k!} p_1^{x_1} p_2^{x_2} \dots p_k^{x_k}$$

Donde:

n Es el número total de ensayos.

x_i Es el número de veces que ocurre la categoría i .

p_i Es la probabilidad de la categoría i .

Las fórmulas para el cálculo de la media, la varianza y la desviación de una distribución de probabilidad multinomial son:

Media (promedio esperado de ocurrencia de la categoría i en n ensayos) $\mu_i = np_i$

Varianza (variabilidad en el número de ocurrencias en la categoría i en n ensayos) $\sigma_i^2 = np_i(1 - p_i)$.

Desviación estándar (Dispersión promedio alrededor de la media para la categoría i) $\sigma = \sqrt{np_i(1 - p_i)}$.

A continuación, se presentan ejemplos ilustrativos de la distribución de probabilidad multinomial.

Ejemplo 1.- Clasificación de Frutas en una Planta Procesadora. En una planta procesadora de frutas, las naranjas que pasan por una máquina clasificadora se dividen en tres categorías: calidad alta, calidad media y calidad baja. Considerando que la proporción de ellas son: 0.6, 0.3 y 0.1 respectivamente, si se procesa una cantidad de 100 naranjas.

a) ¿Cuál es la probabilidad de que se clasifiquen 60 como calidad alta, 30 como calidad media, y 10 como calidad baja?

b) ¿Calcule e interprete la media, la varianza y la desviación estándar del número de naranjas clasificadas como calidad alta?

Solución.

a) ¿Cuál es la probabilidad de que se clasifiquen 60 como calidad alta, 30 como calidad media, y 10 como calidad baja?

$$P(X_1 = x_1, X_2 = x_2, \dots, X_k = x_k) = \frac{n!}{n_1! n_2! \dots n_k!} p_1^{x_1} p_2^{x_2} \dots p_k^{x_k}$$

$$P(X_1 = 60, X_2 = 30, X_3 = 10) = \frac{100!}{60! 30! 10!} 0.6^{60} 0.3^{30} 0.1^{10} = 0.0117$$

Existe un 1.17 % de probabilidades de que 60 naranjas sean de calidad alta, 30 de calidad media y 10 de calidad baja

b) ¿Calcule e interprete la media, la varianza y la desviación estándar del número de naranjas clasificadas como calidad alta?

Media $\mu_i = np_i = 100 * .6 = 60$. Este valor indica que después de realizar el muestreo de 60 piezas por repetidas ocasiones, el valor esperado o promedio será de 60 naranjas de alta calidad

$$\text{Varianza } \sigma_i^2 = np_i(1 - p_i) = 60 * .6 * (1 - .6) = 24$$

$$\text{Desviación estándar } \sigma = \sqrt{np_i(1 - p_i)} = \sqrt{24} = 4.9$$

Tanto la varianza como la desviación estándar indican la variabilidad de las mediciones de un muestreo a otro, con respecto a la media.

Ver solución del anterior ejercicio con Excel en el link https://youtu.be/rog3IoUA_eA

Ejemplo 2.- Preferencias de sabores en una encuesta. Una empresa de bebidas realiza una encuesta con 200 consumidores sobre su preferencia entre tres nuevos sabores: mango, piña y fresa, resultando que el 50 % prefieren mango, el 35 % prefieren piña y el restante 15 % prefiere fresa. De una muestra de 20 personas cual es la probabilidad de que:

a) 10 consumidores prefieran mango, 6 prefieran piña y 4 prefieran fresa.

b) En promedio, ¿cuántos consumidores prefieren piña, y cuál es la desviación estándar de este grupo?

Solución.

a) 10 consumidores prefieran mango, 6 prefieran piña y 4 prefieran fresa.

$$P(X_1 = 10, X_2 = 6, X_3 = 4) = \frac{20!}{10! 6! 4!} 0.50^{10} 0.35^6 0.15^4 = 0.0353$$

Existe un 3.53 % de probabilidad de que 10 consumidores prefieran mango, 6 prefieran piña y 4 prefieran fresa.

b) En promedio, ¿cuántos consumidores prefieren piña, y cuál es la desviación estándar de este grupo?

Media o promedio $\mu_i = np_i = 20 * .35 = 7$. Este valor indica que después de realizar el muestreo de 20 piezas por repetidas ocasiones, el valor esperado o promedio será que 7 personas prefieran el sabor de piña.

Problemas propuestos.

Ver la solución de estos ejercicios en Excel dando clic en el link [ejercicios de probabilidad multinomial](#).

Ejercicio 1.- Defectos en barras de granola. En una línea de producción de barras de granola, los defectos detectados se dividen en tres tipos: falla en el empaque, falla en el tamaño y falla en el contenido. Considerando que las probabilidades en las fallas mencionadas son de 0.2, 0.3 y 0.5 respectivamente, en una muestra de 30 barras de granola.

a) ¿Cuál es la probabilidad de que haya 3 fallas en el empaque, 7 fallas en el tamaño y 10 fallas en el contenido?

b) Calcule la media y la desviación estándar de las fallas en el tamaño.

Ejercicio 2.- Tipos de Salsas Producidas. Una fábrica de alimentos produce tres tipos de salsas en cada lote: salsa muy picante, salsa picante, salsa dulce y salsa agridulce. Considerando que la proporción de producción son 0.30, 0.35, 0.25 y 0.10 respectivamente, de una muestra de 20 salsas.

a) ¿Cuál es la probabilidad de que 6 sean de salsa muy picante, 6 sean de salsa picante, 4 sean de salsa dulce y 4 sean de salsa agridulce?

b) ¿Cuál es la media y la desviación estándar del número de salsas agridulces producidas?

Ejercicio 3.- Revisión de Ingredientes en alimentos congelados. Una empresa productora de alimentos congelados clasifica las irregularidades detectadas en sus productos, en tres categorías: ingredientes incompletos, ingredientes en exceso e ingredientes correctos. Si en base a sus datos históricos, establece que los porcentajes de cada uno de ellos es 25%, 15 % y 60% respectivamente.

a) ¿Cuál es la probabilidad de que, en una muestra de 50 productos congelados, 12 tengan ingredientes incompletos, 8 tengan ingredientes en exceso y 30 tengan ingredientes correctos?

b) Calcule la media y la desviación estándar de los productos con ingredientes incompletos.

2.2.5. Distribución de Poisson.

La distribución de Poisson, una distribución de probabilidad discreta, encuentra un amplio campo de aplicación en diversas áreas, entre ellas la industria alimentaria.

La distribución de Poisson se emplea para modelar el número de ocurrencias de un evento en un intervalo de tiempo o espacio dado, cuando estos eventos son independientes y ocurren a una tasa promedio constante. En otras palabras, esta distribución es ideal para describir fenómenos aleatorios donde el interés radica en contar el número de veces que sucede algo en un determinado período.

En el ámbito de la industria alimentaria, la distribución de Poisson tiene múltiples aplicaciones. Una de las más comunes es el control de calidad. Por ejemplo, se puede utilizar para modelar el número de bacterias presentes en una muestra de alimento. Al conocer la tasa promedio de contaminación bacteriana, se puede calcular la probabilidad de encontrar un determinado número de bacterias en una muestra y establecer los límites de aceptación para garantizar la seguridad alimentaria (Montgomery, Runger, & Hubele, 2018).

Otra aplicación importante es en la gestión de inventarios. Al modelar la demanda de un producto como un proceso de Poisson, se puede optimizar los niveles de inventario y reducir los costos asociados a la escasez o al exceso de stock. Por ejemplo, en una fábrica de helados, se puede utilizar la distribución de Poisson para estimar la demanda diaria de helados de un sabor específico y así ajustar la producción (Walpole, Myers, Myers, & Ye, 2012).

Asimismo, la distribución de Poisson es útil en el análisis de riesgos. Por ejemplo, en la industria cárnica, se puede emplear para modelar el número de casos de contaminación por una determinada bacteria en un lote de productos. Esto permite evaluar la probabilidad de que ocurra un brote y tomar las medidas preventivas necesarias.

En conclusión, la distribución de Poisson es una herramienta estadística fundamental en la industria alimentaria. Su capacidad para modelar eventos aleatorios y contar el número de ocurrencias en un intervalo de tiempo la convierte en una herramienta valiosa para el control de calidad, la gestión de inventarios y el análisis de riesgos.

La fórmula de una distribución de probabilidad de Poisson es.

$$p(X = x) = \frac{\lambda^x e^{-\lambda}}{x!} \text{ Donde:}$$

x es el número de éxitos que se desea que ocurran

λ es la media o promedio de éxitos por unidad de tiempo, área o producto.

e es la constante de Euler, con valor de 2.718

Para este tipo de distribución se tiene que la media es igual a λ , la varianza es igual a la media o promedio ($\sigma^2 = \lambda$) y la desviación estándar es igual a la raíz cuadrada de la varianza ($\sigma = \sqrt{\lambda}$).

A continuación, se presentan unos ejemplos de aplicación de la distribución de probabilidad de Poisson.

Ejemplo1.- Producción de Galletas. Considerando que una máquina produce galletas a una tasa promedio de 120 galletas por hora.

- a) ¿Cuál es la probabilidad de que en un minuto se produzcan exactamente 2 galletas?
- b) ¿Cuál es la probabilidad de que en 5 minutos se produzcan menos de 4 galletas?
- c) ¿Cuál es la probabilidad de que en 10 minutos se produzcan de 18 a 21 galletas?
- d) Para el inciso a calcule la media, la desviación estándar.
- e) Para el inciso c elabore la distribución de función y la distribución de función acumulada.

Solución

- a) ¿Cuál es la probabilidad de que en un minuto se produzcan exactamente 2 galletas?

Como la media proporcionada es de 120 galletas por hora y esta pregunta es en relación a un minuto, por regla de tres se tiene que la media ajustada corresponde a 2 galletas por minuto, de tal forma que la respuesta a este inciso es.

$$p(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

$$p(X = 2) = \frac{2^2 e^{-2}}{2!} = 0.2707$$

Existe un 27.07 % de probabilidades de que en un minuto se produzcan exactamente 2 galletas

- b) ¿Cuál es la probabilidad de que en 5 minutos se produzcan menos de 4 galletas?

Como la maquina produce 120 galletas en 60 minutos, la media ajustada para 5 minutos corresponde 10 galletas.

La probabilidad de que sea menos de 4 galletas corresponde a la suma de las probabilidades de 0, 1, 2 y 3 galletas.

$$p(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

$$p(X = 0) = \frac{10^0 e^{-10}}{0!} = 0.0000$$

$$p(X = 1) = \frac{10^1 e^{-10}}{1!} = 0.0005$$

$$p(X = 2) = \frac{10^2 e^{-10}}{2!} = 0.0023$$

$$p(X = 3) = \frac{10^3 e^{-10}}{3!} = 0.0076$$

$$p(X < 4) = p(X = 0) + p(X = 1) + p(X = 2) + p(X = 3) = 0.0000 + 0.0005 + 0.0023 + 0.0076 = 0.0103$$

Existe un 1.03 % de probabilidades de que en 5 minutos se produzcan menos de 4 galletas.

c) ¿Cuál es la probabilidad de que en 10 minutos se produzcan de 18 a 21 galletas?

La media ajustada para 10 minutos corresponde a 20 galletas.

La solución de este inciso corresponde a la suma de las probabilidades de 18, 19, 20 y 21 galletas.

$$p(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

$$p(X = 18) = \frac{20^{18} e^{-20}}{18!} = 0.0844$$

$$p(X = 19) = \frac{20^{19} e^{-20}}{19!} = 0.0888$$

$$p(X = 20) = \frac{20^{20} e^{-20}}{20!} = 0.0888$$

$$p(X = 21) = \frac{20^{21} e^{-20}}{21!} = 0.0846$$

$$p(18 \leq X \leq 21) = p(X = 18) + p(X = 19) + p(X = 20) + p(X = 21) = 0.0844 + 0.0888 + 0.0888 + 0.0846 = 0.3467$$

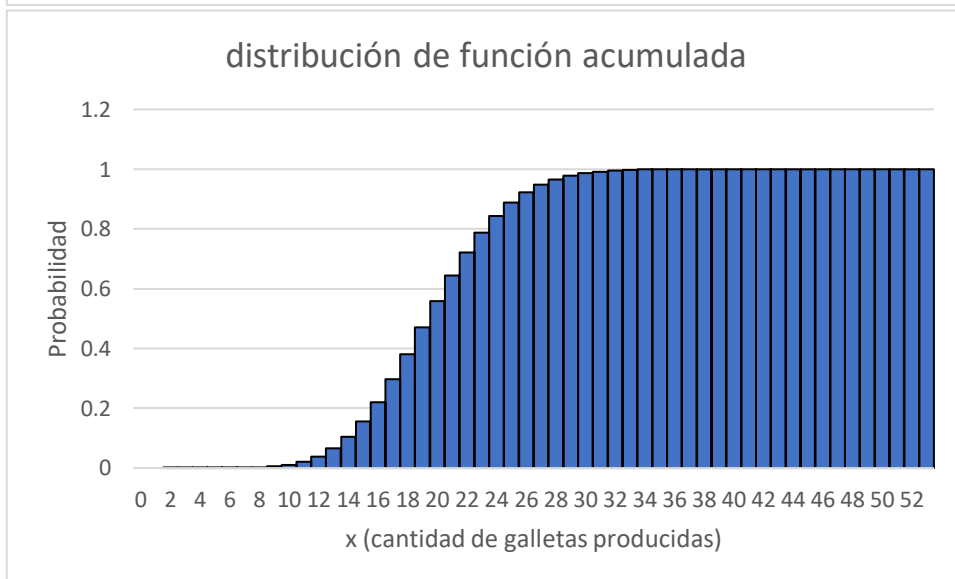
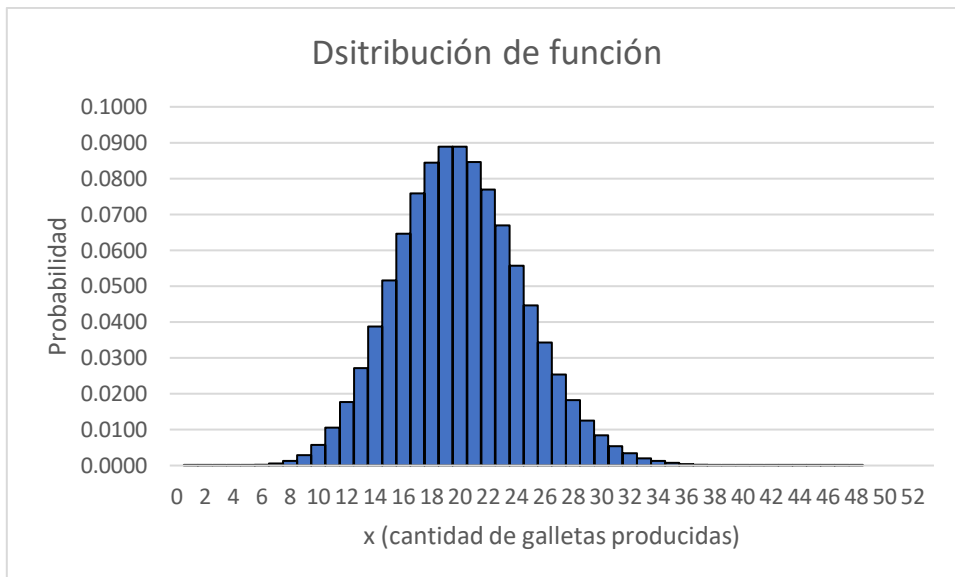
Existe un 34.67 % de probabilidades de que en 10 minutos se produzcan de 18 a 21 galletas.

d) Para el inciso a calcule la media, la desviación estándar.

La media (λ) es de 2 galletas y la desviación estándar ($\sqrt{\lambda}$) es de 1.41.

Ver solución de este ejercicio dando clic al link <https://youtu.be/nY0ZUEJgS9E>

Para el inciso c elabore la distribución función y la distribución función acumulada



La elaboración de estas distribuciones en Excel se puede observar dando clic al link <https://youtu.be/DUE3A4zTChs>

Ejemplo 2.- Demanda de Productos en un Supermercado. En base a las ventas de los últimos meses en el supermercado Las Américas, se pudo establecer que se vende en promedio 50 yogurts por día.

- ¿Cuál es la probabilidad de que en un día determinado se vendan exactamente 60 yogurts?
- ¿Cuál es la probabilidad de que en una semana se vendan 380 yogurts?
- ¿Cuál es la probabilidad de que en un día se vendan 52 o 54 yogurts?
- Para el inciso a calcule la media, la desviación estándar.
- Para el inciso c elabore la función de probabilidad y la función de distribución.

Solución.

a) ¿Cuál es la probabilidad de que en un día determinado se vendan exactamente 60 yogurts?

$$p(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

$$p(X = 60) = \frac{50^{60} e^{-50}}{60!} = 0.0201$$

Existe 2.01 % de probabilidades de que en un día determinado se vendan exactamente 60 yogurts.

b) ¿Cuál es la probabilidad de que en una semana se vendan 380 yogurts?

La media ajustada para este caso sería de 350 yogurts por semana.

$$p(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

$$p(X = 380) = \frac{350^{380} e^{-350}}{380!} = 0.0059$$

Existe 0.59 % de probabilidades de que en una semana se vendan 380 yogurts

c) ¿Cuál es la probabilidad de que en un día se vendan 52 o 54 yogurts?

La media es de 50 yogurts por día.

La respuesta a este inciso implica la suma de las probabilidades de vender 52 y 54 yogurts.

$$p(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

$$p(X = 52) = \frac{50^{52} e^{-50}}{52!} = 0.0531$$

$$p(X = 54) = \frac{50^{54} e^{-50}}{54!} = 0.0464$$

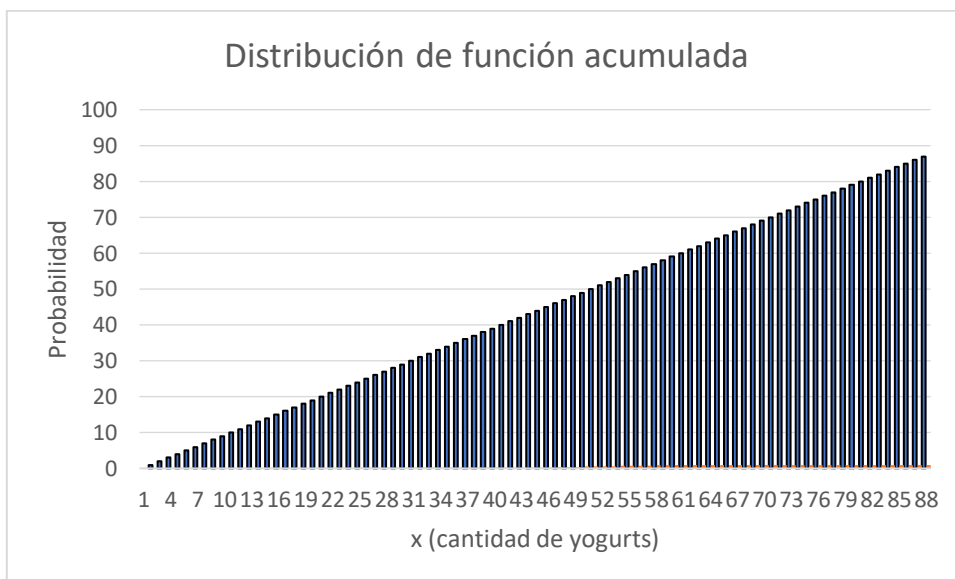
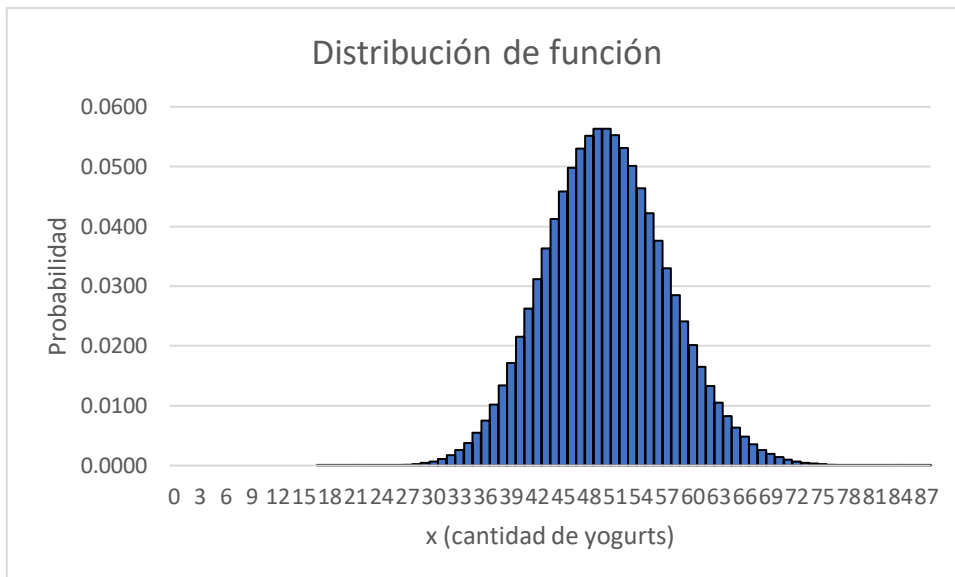
$$p(X = 52 \text{ o } X = 54) = p(X = 52) + p(X = 54) = 0.0531 + 0.0464 = 0.0995$$

Existe un 9.95 % de probabilidades de que en un día se vendan 52 o 54 yogurts

d) Para el inciso a calcule la media, la desviación estándar.

La media (λ) es de 50 yogurts por día y la desviación estándar ($\sqrt{\lambda}$) es de 7.07 yogurts por día

e) Para el inciso c elabore la función de probabilidad y la función de distribución.



Problemas propuestos.

Las respuestas en Excel se pueden observar dando clic al link [distribución de probabilidad de Poisson](#).

Ejercicio 1.- Microorganismos en Alimentos En un proceso de producción de queso, se estima que hay un promedio de 2 bacterias por gramo de producto.

- ¿Cuál es la probabilidad de que en una muestra de 1 gramo no haya ninguna bacteria?
- ¿Cuál es la probabilidad de que en una muestra de 2 gramos haya más de 3 bacterias?
- Para el inciso a calcule la media, la desviación estándar.
- Para el inciso a elabore la función de probabilidad y la función de distribución.

Ejercicio 2.- Llegada de Clientes a una Cafetería. Una cafetería recibe un promedio de 30 clientes por hora.

- a) ¿Cuál es la probabilidad de que en un intervalo de 6 minutos no llegue ningún cliente?
- b) ¿Cuál es la probabilidad de que en una hora lleguen más de 40 clientes?
- c) ¿Cuál es la probabilidad de que en 10 minutos lleguen exactamente 6 clientes?
- d) Para el inciso a calcule la media, la desviación estándar.
- e) Para el inciso c elabore la función de probabilidad y la función de distribución.

Ejercicio 3.- Defectos en Empaques de Snacks. Una máquina empaqa bolsas de papas fritas a una tasa de 500 bolsas por minuto. Se sabe que el 2% de las bolsas tienen un cierre defectuoso.

- a) ¿Cuál es la probabilidad de que en un minuto se produzcan exactamente 5 bolsas con cierre defectuoso?
- b) ¿Cuál es la probabilidad de que en dos minutos se produzcan exactamente 20 bolsas con cierre defectuoso?
- c) Para el inciso b calcule la media, la desviación estándar.
- d) Para el inciso b elabore la función de probabilidad y la función de distribución.

Ejercicio 4.- Reclamos por Alergenos. Si una empresa de alimentos recibe un promedio de 2 reclamos por semana relacionados con alérgenos no declarados en sus productos.

- a) ¿Cuál es la probabilidad de que en una semana no se reciba ningún reclamo por alérgenos?
- b) ¿Cuál es la probabilidad de que en una semana reciba más de 2 reclamos por alérgenos?
- c) Para el inciso b calcule la media, la desviación estándar.
- d) Para el inciso b elabore la función de probabilidad y la función de distribución.

Ejercicio 5.- Contaminación Bacteriana en Productos Lácteos. Considerando que se estima que en un lote de queso hay un promedio de 3 bacterias coliformes por cada 100 gramos.

- a) ¿Cuál es la probabilidad de que en 200 gramos de queso se encuentren 5 bacterias?
- b) ¿Cuál es la probabilidad de que en 100 gramos de queso se encuentren más de 2 bacterias?
- c) Para el inciso b calcule la media, la desviación estándar.
- d) Para el inciso b elabore la función de probabilidad y la función de distribución.

Una evaluación de esta unidad puede ser realizada escaneando el código QR siguiente.



3. Variables Aleatorias Continuas y sus distribuciones de Probabilidad.

3.1. Variable aleatoria continua.

Las variables aleatorias continuas son elementos fundamentales en la gestión y el control de calidad dentro de la compleja industria alimentaria. A diferencia de las variables discretas, que representan valores contables y finitos (como el número de microorganismos en una muestra), las variables continuas pueden asumir cualquier valor dentro de un intervalo específico.

La temperatura es un claro ejemplo de variable aleatoria continua en la industria alimentaria. Durante tratamientos térmicos como la pasteurización, la esterilización y la cocción, el control preciso de la temperatura es vital para la inactivación de microorganismos patógenos, la preservación de las propiedades nutricionales y organolépticas, y la prolongación de la vida útil del producto. La temperatura, medida en grados Celsius o Fahrenheit, fluctúa continuamente dentro de un rango operativo y su monitoreo constante es un pilar de la seguridad alimentaria. Desviaciones de los parámetros de temperatura establecidos pueden acarrear consecuencias graves, desde la proliferación de bacterias nocivas hasta la alteración de la textura y el sabor del alimento, generando pérdidas económicas significativas y poniendo en riesgo la salud pública (Fellows, 2017).

El peso y el volumen de los productos alimentarios constituyen otras variables continuas de gran relevancia. En las líneas de envasado y empaquetado, la precisión en el peso y el volumen del contenido es esencial para cumplir con las normativas de etiquetado, garantizar la equidad comercial y satisfacer las expectativas del consumidor. Las máquinas de llenado automatizadas incorporan sensores y sistemas de control que miden ininterrumpidamente estas variables, realizando ajustes dinámicos para minimizar la variabilidad y asegurar la uniformidad del producto final. El análisis estadístico de estas mediciones, mediante distribuciones de probabilidad continuas como la distribución normal (o Gaussiana), permite a las empresas optimizar los procesos de envasado, reducir el desperdicio de material y producto, y mejorar la eficiencia global de la producción (Montgomery & Runger, 2010).

La viscosidad, que mide la resistencia de un fluido a fluir, es una variable continua de suma importancia en la elaboración de productos líquidos y semilíquidos como salsas, aderezos, néctares, purés y cremas. La viscosidad influye directamente en la textura, la consistencia, la fluidez y, por ende, en la aceptabilidad sensorial del producto. Los viscosímetros, instrumentos de medición especializados, permiten determinar la viscosidad de forma continua, proporcionando a los fabricantes información valiosa para controlar la calidad y la consistencia de sus productos. Variaciones en la viscosidad pueden afectar la experiencia del consumidor, la estabilidad del producto durante su almacenamiento y su comportamiento durante el procesamiento (Barbosa-Cánovas et al., 2005).

El pH, una medida de la acidez o alcalinidad de una sustancia, es otra variable continua crítica en la industria alimentaria. El pH influye en la actividad microbiana, las reacciones enzimáticas, la estabilidad de las proteínas y la calidad sensorial de los alimentos. El control preciso del pH es fundamental en la elaboración de productos fermentados, como el yogur y el queso, y en la conservación de alimentos mediante acidificación. Los medidores de pH permiten monitorear esta variable de forma continua, asegurando el cumplimiento de los estándares de calidad y seguridad alimentaria.

El contenido de humedad, expresado como la cantidad de agua presente en un alimento, es una variable continua que afecta significativamente la textura, la actividad de agua (a_w), la estabilidad microbiológica y la vida útil de los productos. El control preciso del contenido de humedad es esencial en la producción de alimentos deshidratados, productos horneados y productos cárnicos. Los analizadores de humedad permiten medir esta variable de forma continua o intermitente, permitiendo a los fabricantes ajustar los procesos de secado, horneado o cocción para obtener productos con las características deseadas.

Además de las variables ya mencionadas, otras variables continuas relevantes en la industria alimentaria incluyen la concentración de azúcares (medida con refractómetros), la actividad de agua (a_w), el color (medido con colorímetros) y la textura (medida con texturómetros). El análisis estadístico avanzado, que incluye el análisis de varianza, el diseño de experimentos y el análisis de regresión, permite a las empresas investigar la interrelación entre estas variables y las características del producto, optimizando los procesos de producción y desarrollando nuevos productos con atributos específicos (Walpole et al., 2012). La implementación de sistemas de control estadístico de procesos basados en el monitoreo continuo de estas variables permite la detección temprana de desviaciones y la implementación de acciones correctivas, minimizando la variabilidad y garantizando la calidad constante del producto.

En conclusión. La correcta medición, análisis e interpretación, de las variables aleatorias continuas, mediante el uso de técnicas estadísticas y tecnologías de medición avanzadas, permiten a las empresas tomar decisiones informadas, optimizar procesos, garantizar la seguridad alimentaria, mejorar la calidad de los productos y satisfacer las demandas de un mercado cada vez más exigente. El dominio de estos conceptos y su aplicación práctica son cruciales para los profesionales de la industria alimentaria que buscan la innovación, la eficiencia y la excelencia.

3.2. Función de densidad y acumulativa.

Tal como se mencionaba anteriormente, las variables aleatorias continuas, a diferencia de las discretas que toman valores contables, pueden asumir cualquier valor dentro de un intervalo dado. Para describir la probabilidad asociada a estas variables, se utilizan dos funciones

fundamentales: la función de densidad de probabilidad (FDP) y la función de distribución acumulativa (FDA).

Función de Densidad de Probabilidad (FDP)

La función de densidad de probabilidad, denotada generalmente como $f(x)$, describe la probabilidad relativa de que una variable aleatoria continua tome un valor dado. A diferencia de las variables discretas, donde se asigna una probabilidad a cada valor específico, en las variables continuas la probabilidad de que la variable tome un valor *exacto* es infinitesimalmente pequeña, prácticamente cero. En cambio, la FDP permite calcular la probabilidad de que la variable se encuentre dentro de un *intervalo* determinado.

Matemáticamente, la FDP se define de tal manera que el área bajo la curva de $f(x)$ dentro de un intervalo dado representa la probabilidad de que la variable aleatoria se encuentre en ese intervalo. Formalmente:

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

Donde:

- X es la variable aleatoria continua.
- a y b son los límites del intervalo.
- $\int_a^b f(x)dx$ representa la integral de $f(x)$ desde a hasta b , que geométricamente corresponde al área bajo la curva entre estos límites.

Función de Distribución Acumulativa (FDA)

La función de distribución acumulativa, denotada generalmente como $F(x)$, representa la probabilidad de que la variable aleatoria X tome un valor menor o igual a un valor específico x . Se define como la integral de la FDP desde $-\infty$ hasta x :

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(t)dt$$

Donde:

- t es una variable de integración.

La FDA es una función creciente que varía entre 0 y 1. Sus propiedades principales son:

- $F(-\infty) = 0$
- $F(\infty) = 1$

- Si $a \leq b$, entonces $F(a) \leq F(b)$
- $P(a \leq X \leq b) = F(b) - F(a)$

3.3. Valor esperado.

El concepto de valor esperado, también conocido como esperanza matemática o media, es fundamental en la teoría de la probabilidad y la estadística. Proporciona una medida del "promedio" de una variable aleatoria, ponderando cada posible valor por su probabilidad de ocurrencia. Si bien este concepto es relativamente intuitivo para variables discretas, su extensión a variables continuas requiere un tratamiento más formal mediante el cálculo integral.

Una variable aleatoria continua, a diferencia de una discreta, puede tomar cualquier valor dentro de un intervalo dado. Su comportamiento se describe mediante una función de densidad de probabilidad (FDP), denotada por $f(x)$. La FDP cumple con la condición de que su integral sobre todo el espacio muestral es igual a 1, lo que asegura que la probabilidad total es 100%.

El valor esperado de una variable aleatoria continua X , denotado como $E[X]$ o μ , se define mediante la siguiente integral:

$$E[X] = \int_{-\infty}^{\infty} xf(x)dx$$

Esta integral pondera cada valor posible x por su "densidad" de probabilidad $f(x)$ y suma (integra) estos productos sobre todo el rango de valores posibles. Es importante notar que, a diferencia de las variables discretas donde sumamos probabilidades puntuales, aquí integramos una densidad, lo que implica que la probabilidad de que la variable tome un valor *exacto* específico es cero.

El valor esperado no es necesariamente un valor que la variable pueda tomar realmente. Más bien, representa el valor promedio que se esperaría obtener si se repitiera el experimento un gran número de veces. En el contexto de distribuciones continuas, podemos visualizarlo como el centro de masa de la distribución de probabilidad.

3.4. Distribuciones uniforme y exponencial.

Distribución uniforme.

La distribución uniforme es un modelo de probabilidad continua donde cada valor dentro de un intervalo específico tiene la misma probabilidad de ocurrencia. Esta característica de equiprobabilidad la convierte en una herramienta útil para modelar situaciones donde se desconoce la distribución subyacente o donde se asume una variabilidad aleatoria sin sesgos.

Si bien puede parecer un modelo simplista, su aplicación en diversos campos, incluyendo la ingeniería en industria alimentaria, ofrece perspectivas interesantes para el análisis y la optimización de procesos.

Definición y Características Matemáticas

Una variable aleatoria X sigue una distribución uniforme continua en el intervalo $[a, b]$ si su función de densidad de probabilidad (FDP) está definida por:

$$f(x) = \frac{1}{b-a} \text{ para } a \leq x \leq b, f(x) = 0 \text{ para cualquier otro valor de } x$$

Donde a es el límite inferior y b es el límite superior del intervalo. La probabilidad de que X tome un valor dentro de cualquier subintervalo de $[a, b]$ es proporcional a la longitud de ese subintervalo. El valor esperado (media), la varianza y la desviación estándar de una variable con distribución uniforme, son respectivamente:

$$E[X] = \mu = \frac{a+b}{2}, \sigma^2 = \frac{(b-a)^2}{12}, \sigma = \sqrt{\frac{(b-a)^2}{12}}$$

Estas fórmulas proporcionan medidas centrales y de dispersión que permiten caracterizar la distribución.

Para el cálculo de la probabilidad de un intervalo contenido dentro del intervalo total de la distribución de probabilidad uniforme se obtiene mediante la formula.

$$P(c \leq x \leq d) = \frac{d - c}{b - a}$$

Donde el intervalo $[c, d]$ está contenido dentro del intervalo $[a, b]$.

En la ingeniería en industria alimentaria, la distribución uniforme puede ser útil en diversas situaciones, aunque su aplicación directa puede ser menos frecuente que otras distribuciones como la normal o la exponencial. Sin embargo, en escenarios donde se asume una variabilidad aleatoria sin información previa sobre la distribución, la uniforme puede servir como un primer acercamiento o como un modelo simplificado. A continuación, se presentan algunos ejemplos:

- Distribución de tamaños de partículas: En procesos de molienda o tamizado, si se asume que los tamaños de las partículas se distribuyen uniformemente dentro de un cierto rango, se puede utilizar este modelo para estimar la proporción de partículas dentro de un tamaño específico. Esto es relevante para el control de calidad de harinas, azúcares y otros ingredientes pulverizados. Los principios de probabilidad y estadística, incluyendo la comprensión de distribuciones como la uniforme, son fundamentales para el análisis de datos en la industria alimentaria (Walpole, Myers, Myers, & Ye, 2012).

- Tiempo de llenado de envases: Si se considera que el tiempo que tarda una máquina en llenar un envase varía uniformemente dentro de un intervalo conocido, se puede utilizar la distribución uniforme para calcular la probabilidad de que el tiempo de llenado esté dentro de ciertos límites. Esto es importante para la optimización de líneas de envasado y el cumplimiento de estándares de producción. La optimización de procesos, como el llenado de envases, es un aspecto clave en la tecnología de procesamiento de alimentos (Fellows, 2017).
- Variación de la temperatura en un proceso corto: En procesos térmicos cortos, donde se busca mantener una temperatura constante, se podría modelar la pequeña variación de la temperatura alrededor del valor objetivo como una distribución uniforme. Esto ayudaría a evaluar la estabilidad del proceso y el posible impacto en la calidad del alimento.

A continuación, se presentan unos ejemplos de aplicación de la distribución de probabilidad uniforme.

Ejemplo 1: Llenado de Botellas de Aceite

En base a los datos históricos de una máquina automática que llena botellas de aceite de oliva, se estableció que el volumen de llenado varía uniformemente entre 980 ml y 1020 ml. Atendiendo esto, calcule lo siguiente:

- ¿Cuál es la probabilidad de que una botella contenga entre 990 ml y 1010 ml de aceite?
- Si se seleccionan 100 botellas al azar, ¿cuántas se espera que contengan menos de 995 ml de aceite?
- La media, la varianza y la desviación estándar.

Solución.

- ¿Cuál es la probabilidad de que una botella contenga entre 990 ml y 1010 ml de aceite?

En este caso el intervalo de 990 a 1010 ml. $[c, d]$ está contenido dentro del intervalo 980 a 1020 $[a, b]$ y de aplicar la formula.

$$P(c \leq x \leq d) = \frac{d - c}{b - a}$$

$$P(990 \leq x \leq 1010) = \frac{1010 - 990}{1020 - 980} = \frac{20}{40} = 0.5$$

50 % de probabilidades que la botella de aceite de oliva contenga de 990 a 1010 ml.

- Si se seleccionan 100 botellas al azar, ¿cuántas se espera que contengan menos de 995 ml de aceite?

Para este caso menos de 995 ml. Corresponde al intervalo comprendido de 980 a 995 ml., y su respectiva probabilidad sería.

$$P(c \leq x \leq d) = \frac{d - c}{b - a}$$

$$P(990 \leq x \leq 1010) = \frac{980 - 995}{1020 - 980} = \frac{15}{40} = 0.375$$

37.5 % de probabilidades que la botella de aceite de oliva contenga menos de 995 ml., por lo que, de una muestra de 100 botellas, 37.5 de ellas se espera que tengan menos de 995 ml.

Lo anterior tendría más sentido si se dijera que de 1000 botellas, se espera que 375 de ellas contengan menos de 995 ml.

c) La media, la varianza y la desviación estándar.

$$\text{Media } E[X] = \mu = \frac{a+b}{2} = \frac{980+1020}{2} = 1000$$

$$\text{Varianza } \sigma^2 = \frac{(b-a)^2}{12} = \frac{(1020-980)^2}{12} = 133.33$$

$$\text{Desviación estándar } \sigma = \sqrt{\frac{(b-a)^2}{12}} = \sqrt{\frac{(1020-980)^2}{12}} = 11.55$$

Ver solución usando Excel dando clic en el link <https://youtu.be/cNeO2scUvD8>

Ejemplo 2: Peso de Paquetes de Galletas

Considerando que el peso de los paquetes de galletas producidos por una fábrica se distribuye uniformemente entre 190 g y 210 g.

a) Calcula la probabilidad de que un paquete de galletas pese más de 205 g.

b) Calcula la media, la varianza y la desviación estándar.

Solución.

a) Calcula la probabilidad de que un paquete de galletas pese más de 205 g.

En este caso el intervalo de 205 a 210 g. [c, d] está contenido dentro del intervalo 190 a 210 g. [a, b] y de aplicar la formula.

$$P(c \leq x \leq d) = \frac{d - c}{b - a}$$

$$P(205 \leq x \leq 210) = \frac{210 - 205}{210 - 190} = \frac{5}{20} = 0.25$$

25 % de probabilidades de que un paquete de galletas pese más de 205 g.

b) Calcula la media, la varianza y la desviación estándar.

$$\text{Media } E[X] = \mu = \frac{a+b}{2} = \frac{190+210}{2} = 200$$

$$\text{Varianza } \sigma^2 = \frac{(b-a)^2}{12} = \frac{(210-190)^2}{12} = 33.33$$

$$\text{Desviación estándar } \sigma = \sqrt{\frac{(b-a)^2}{12}} = \sqrt{\frac{(21-19)^2}{12}} = 5.77$$

Ejercicios propuestos.

Ver ejercicios resueltos con Excel dando clic en el link [ejercicios de distribución uniforme](#).

Ejercicio 1. Tiempo de Cocción de Lotes de Verduras Congeladas

Si el tiempo de cocción óptimo para un lote de verduras congeladas en una línea de producción varía uniformemente entre 8 y 12 minutos.

- ¿Cuál es la probabilidad de que un lote de verduras se cocine entre 9 y 10.5 minutos?
- Si se procesan 50 lotes, ¿cuántos se espera que requieran un tiempo de cocción superior a 10 minutos?
- Calcule la media, la varianza y la desviación estándar.

Ejercicio 2. Nivel de Acidez del Vinagre

El departamento de control de calidad de la empresa Eva dedicada a la producción de vinagre de manzana determinó que el nivel de acidez del vinagre producido se distribuye uniformemente entre 4.5% y 5.5%.

- Calcule la probabilidad de que una muestra de vinagre tenga un nivel de acidez entre 4.8% y 5.2%.
- Calcule la media, la varianza y la desviación estándar.

Ejercicio 3. Diámetro de las Manzanas

El gerente del empaque de Huertas El Gemelo ubicada en Cd. Cuauhtémoc calculó en base a los datos de los últimos lotes procesados, que el diámetro de las manzanas en una línea de clasificación automática se distribuye uniformemente entre 7 cm y 9 cm.

- ¿Cuál es la probabilidad de que una manzana seleccionada al azar tenga un diámetro inferior a 8 cm?
- Si se miden 200 manzanas, ¿cuántas se espera que tengan un diámetro entre 8 cm y 8.5 cm?
- Calcule la media, la varianza y la desviación estándar.

Distribución exponencial

La distribución de probabilidad exponencial es una herramienta estadística fundamental para modelar el tiempo entre eventos sucesivos en procesos donde estos ocurren de forma continua, independiente y a una tasa promedio constante. A diferencia de la distribución de Poisson, que se centra en el número de eventos que ocurren en un intervalo de tiempo fijo, la distribución exponencial se enfoca en el *tiempo* que transcurre entre esos eventos. Esta característica la convierte en una herramienta valiosa en diversos campos, incluyendo la ingeniería de alimentos, donde el tiempo entre ciertos sucesos impacta directamente en la producción, la calidad y la seguridad alimentaria.

La distribución exponencial se define por un único parámetro, λ (lambda), que representa la tasa promedio de ocurrencia de los eventos. Matemáticamente, su función de densidad de probabilidad (FDP) se expresa como:

$$f(t) = \lambda e^{-\lambda t} \text{ para } x \geq 0$$

Donde:

- t es la variable aleatoria que representa el tiempo entre eventos.
- λ es la tasa promedio de ocurrencia de eventos (eventos por unidad de tiempo).
- e es la base del logaritmo natural (aproximadamente 2.718).

Una propiedad crucial de la distribución exponencial es la "falta de memoria". Esta propiedad implica que la probabilidad de que un evento ocurra en el futuro no depende del tiempo que haya transcurrido desde el último evento. En otras palabras, el sistema no "recuerda" cuánto tiempo ha pasado. Por ejemplo, si se modela el tiempo entre fallas de una máquina envasadora, la probabilidad de que la máquina falle en la próxima hora es la misma, independientemente de si ha estado funcionando sin problemas durante una hora o durante cien horas. Como señalan Montgomery y Runger (2010), la distribución exponencial se caracteriza por esta propiedad de "falta de memoria", lo que la hace adecuada para modelar el tiempo hasta la ocurrencia de un evento en sistemas donde el pasado no influye en el futuro inmediato.

En la industria alimentaria, la distribución exponencial encuentra diversas aplicaciones. Un ejemplo relevante es el modelado del tiempo entre fallas de equipos en líneas de producción. Las máquinas utilizadas en el procesamiento y envasado de alimentos, como llenadoras, pasteurizadoras o etiquetadoras, pueden experimentar paradas debido a fallos mecánicos, eléctricos o de software. El tiempo entre estas paradas puede modelarse mediante una distribución exponencial. Conocer la tasa de fallas (λ) permite a los ingenieros de mantenimiento programar intervenciones preventivas, optimizar el inventario de repuestos y minimizar el tiempo de inactividad de la producción.

Otro ámbito de aplicación importante es el estudio de la vida útil de los alimentos. Si bien el deterioro de los alimentos es un proceso complejo influenciado por múltiples factores (temperatura, humedad, composición del alimento, etc.), bajo ciertas condiciones controladas, el tiempo que tarda un producto en alcanzar un nivel de deterioro inaceptable puede aproximarse mediante una distribución exponencial. Esta información es crucial para la gestión de inventarios, la planificación de la producción y la determinación de fechas de caducidad. No obstante, es importante reconocer que esta aproximación es válida bajo condiciones específicas y que, en muchos casos, modelos más complejos como la distribución de Weibull pueden ser más adecuados para modelar la vida útil de los alimentos, como se discute en Evans, Hastings y Peacock (2000), donde se exploran diferentes distribuciones de probabilidad aplicables a problemas de confiabilidad y vida útil.

Es fundamental destacar que la aplicación de la distribución exponencial requiere la validación de sus supuestos, principalmente la independencia de los eventos y la constancia de la tasa λ . En la práctica, estos supuestos pueden no cumplirse estrictamente. Por ejemplo, el mantenimiento preventivo puede influir en la tasa de fallas de una máquina, haciendo que no sea constante. De igual manera, las fluctuaciones de temperatura durante el almacenamiento pueden afectar la tasa de deterioro de un alimento. Por lo tanto, es crucial realizar un análisis cuidadoso para determinar si la distribución exponencial es un modelo adecuado para el fenómeno en estudio.

En conclusión, la distribución de probabilidad exponencial es una herramienta valiosa en la industria alimentaria, permitiendo modelar el tiempo entre eventos en diversos procesos, desde fallas de equipos hasta el deterioro de alimentos. Su comprensión y aplicación correcta contribuyen a mejorar la eficiencia de la producción, la gestión de la calidad y la seguridad alimentaria. Sin embargo, es esencial considerar las limitaciones del modelo y validar sus supuestos antes de aplicarlo en un contexto específico.

Para el cálculo de la probabilidad de que el evento se presente en un tiempo en específico o antes se usa la formula.

$$P(X \leq t) = 1 - e^{-\lambda t}$$

Para el cálculo de la probabilidad de que el evento se presente después de un tiempo en específico se usa la formula.

$$P(X \geq t) = e^{-\lambda t}$$

Donde:

- t es la variable aleatoria que representa el tiempo entre eventos.
- λ es la tasa promedio de ocurrencia de eventos (eventos por unidad de tiempo).

- e es la base del logaritmo natural (aproximadamente 2.718).

Las fórmulas para calcular la media o promedio (μ), la varianza (σ^2) y la desviación estándar (σ) de falla, respectivamente son:

$$\mu = \frac{1}{\lambda} \quad \sigma^2 = \frac{1}{\lambda^2} \quad \sigma = \sqrt{\frac{1}{\lambda^2}}$$

Donde:

- λ es la tasa promedio de ocurrencia de eventos (eventos por unidad de tiempo).

A continuación, se presentan ejemplos de aplicación de la distribución de probabilidad exponencial.

Ejemplo 1.- Fallas en una máquina envasadora de leche

El gerente de la empresa Quesería Dos Lagunas ha podido determinar que la máquina envasadora de leche presenta 0.2 fallas por hora. Asumiendo que el tiempo entre fallas sigue una distribución exponencial:

- ¿Cuál es el tiempo promedio, la varianza y la desviación estándar entre fallas de la máquina?
- ¿Cuál es la probabilidad de que la máquina falle en 3 horas o menos?
- ¿Cuál es la probabilidad de que la máquina falle después de 4 horas?
- ¿Cuál es la probabilidad de que la máquina falle entre 1 y 2 horas?

Solución.

- ¿Cuál es el tiempo promedio, la varianza y la desviación estándar entre fallas de la máquina?

$\mu = \frac{1}{\lambda} = \frac{1}{0.2 \frac{\text{fallas}}{\text{hr}}} = 5 \frac{\text{hr}}{\text{falla}}$ El tiempo promedio para que se presente una falla en la máquina es de 5 horas.

$\sigma^2 = \frac{1}{\lambda^2} = \frac{1}{(0.2 \frac{\text{fallas}}{\text{hr}})^2} = 25 \frac{\text{hr}^2}{\text{falla}^2}$ La varianza en el tiempo de que se presente una falla corresponde a $25 \frac{\text{hr}^2}{\text{falla}^2}$

$\sigma = \sqrt{\frac{1}{\lambda^2}} = \sqrt{\frac{1}{(0.2 \frac{fallas}{hr})^2}} = 5 \frac{hr}{falla}$ La desviación estándar en el tiempo de que se presente una falla corresponde a $5 \frac{hr}{falla}$

b) ¿Cuál es la probabilidad de que la máquina falle en 3 horas o menos?

$$P(X \leq t) = 1 - e^{-\lambda t}$$

$$P(X \leq 3) = 1 - e^{-0.2(3)} = 0.4512$$

Existe un 45.12 % de probabilidad de que la máquina falle en 3 horas o menos.

c) ¿Cuál es la probabilidad de que la máquina falle después de 4 horas?

$$P(X \geq t) = e^{-\lambda t}$$

$$P(X \geq 4) = e^{-0.2(4)} = 0.4493$$

Existe un 44.93 % de probabilidad de que la máquina falle después de 4 horas.

d) ¿Cuál es la probabilidad de que la máquina falle entre 1 y 2 horas?

$$P(1 \leq t \leq 2) = P(t \leq 2) - P(t \leq 1)$$

$$P(X \leq 2) = 1 - e^{-0.2(2)} = 0.3297$$

$$P(X \leq 1) = 1 - e^{-0.2(1)} = 0.1813$$

$$P(1 \leq t \leq 2) = P(t \leq 2) - P(t \leq 1) = 0.3217 - 0.1813 = 0.1484$$

Existe un 14.84 % de probabilidad de que la máquina falle en el lapso comprendido de 1 y 2 horas.

Ver video de solución de este ejercicio con Excel dando clic al link <https://youtu.be/TF-2al7y-fA>

Ver video de solución de este ejercicio con GeoGebra dando clic al link <https://youtu.be/iXSu2DJ33FU>

Ejemplo 2: Llegada de camiones de tomate a una planta procesadora

Los camiones que transportan tomates a una planta procesadora llegan a una tasa promedio de $\lambda = 0.5$ camiones por hora. Asumiendo una distribución exponencial para el tiempo entre llegadas:

a) ¿Cuál es el tiempo promedio y la desviación estándar entre la llegada de dos camiones consecutivos?

b) ¿Cuál es la probabilidad de que transcurran más de 2 horas entre la llegada de dos camiones consecutivos?

c) ¿Cuál es la probabilidad de que el próximo camión llegue en menos de 30 minutos?

d) ¿Cuál es la probabilidad de que el siguiente camión llegue dentro del rango de 1.5 a 3 horas?

Solución.

a) ¿Cuál es el tiempo promedio y la desviación estándar entre la llegada de dos camiones consecutivos?

$\mu = \frac{1}{\lambda} = \frac{1}{0.5 \frac{\text{camiones}}{\text{hr}}} = 2 \frac{\text{hrs}}{\text{camión}}$ El tiempo promedio entre la llegada de los camiones es de 2 horas.

$\sigma = \sqrt{\frac{1}{\lambda^2}} = \sqrt{\frac{1}{(0.5 \frac{\text{camiones}}{\text{hr}})^2}} = 2 \frac{\text{hrs}}{\text{camión}}$ La desviación estándar en el tiempo de llegada entre un camión y otro es de 2 horas.

b) ¿Cuál es la probabilidad de que transcurran más de 2 horas entre la llegada de dos camiones consecutivos?

$$P(X \geq t) = e^{-\lambda t}$$

$$P(X \geq 2) = e^{-0.5(2)} = 0.3679$$

Existe un 36.79 % de probabilidad de que el próximo camión llegue después de 2 horas.

c) ¿Cuál es la probabilidad de que el próximo camión llegue en menos de 30 minutos?

30 minutos es igual a 0.5 horas

$$P(X \leq t) = 1 - e^{-\lambda t}$$

$$P(X \leq 0.5) = 1 - e^{-0.5(.5)} = 0.2212$$

Existe un 22.12 % de probabilidad de el siguiente camión llegue en menos de 30 minutos.

d) ¿Cuál es la probabilidad de que el siguiente camión llegue dentro del rango de 1.5 a 3 horas?

$$P(1.5 \leq t \leq 3) = P(t \leq 3) - P(t \leq 1.5)$$

$$P(X \leq 3) = 1 - e^{-0.5(3)} = 0.7769$$

$$P(X \leq 1.5) = 1 - e^{-0.5(1.5)} = 0.5276$$

$$P(1 \leq t \leq 2) = P(t \leq 2) - P(t \leq 1) = 0.3217 - 0.1813 = 0.2492$$

Existe un 24.92 % de probabilidad de que el próximo camión llegue entre 1.5 y 3 horas.

Ejercicios propuestos.

Ejercicio 1: Tiempo de procesamiento de pollos en una línea de producción

El tiempo que tarda una línea de producción en procesar un pollo sigue una distribución exponencial con una tasa (λ) de 4 pollos por minuto, asumiendo esto como cierto.

a) ¿Cuál es el tiempo promedio y la desviación estándar entre el procesamiento de un pollo y otro? Respuesta $\mu=0.25$ minutos por pollo, $\sigma=0.25$ minutos por pollo

b) ¿Cuál es la probabilidad de que el procesamiento del próximo pollo dure menos de 20 segundos? Respuesta 0.7364

c) ¿Cuál es la probabilidad de que el tiempo de procesamiento entre un pollo y otro sea de 30 a 50 segundos? Respuesta 0.0997

d) ¿Cuál es la probabilidad de que el siguiente pollo se procese en más de 45 segundos? Respuesta 0.0497

Ejercicio 2: Fallos en el sistema de envasado de conservas

Una máquina envasadora de conservas presenta fallos de forma aleatoria. El tiempo entre fallos sigue una distribución exponencial con una media de 20 horas, ($\lambda = 0.05$ fallos / hora)

a) ¿Cuál es la probabilidad de que la máquina falle después de 30 horas? Respuesta 0.2231

b) ¿Cuál es la probabilidad de que la máquina falle en el lapso de 18 a 25 horas? Respuesta 0.1201

c) ¿Cuál es la probabilidad de que la máquina falle en 15 horas o menos? Respuesta 0.5276

Ejercicio 3: Tiempo de atención al cliente en una panadería

El tiempo que un panadero tarda en atender a un cliente sigue una distribución exponencial con una media de 2 minutos ($\lambda = 0.5$ clientes / minuto)

a) ¿Cuál es la probabilidad de que un cliente sea atendido en menos de 1 minuto? Respuesta 0.3937

b) ¿Cuál es la probabilidad de que un cliente espere más de 3 minutos para ser atendido?
Respuesta 0.2231

c) ¿Cuál es la probabilidad de que un cliente espere de 1 a 2 minutos para ser atendido?
Respuesta 0.2386

Ejercicio 4: Tiempo de inactividad de una máquina de envasado al vacío

Una máquina utilizada para el envasado al vacío de carne presenta tiempos de inactividad (debido a mantenimientos, reparaciones menores, etc.) que siguen una distribución exponencial con una tasa de $\lambda = 0.07$ fallos por hora.

a) ¿Cuál es la probabilidad de que la máquina falle en las próximas 5 horas?
Respuesta 0.2953

b) ¿Cuál es la probabilidad de que la máquina tenga su primer fallo entre la hora 2 y la hora 4 de funcionamiento? Respuesta 0.1136

c) ¿Cuál es la probabilidad de que la máquina falle después de 8 horas? Respuesta 0.5712

Ejercicio 5: Tiempo de espera en la recepción de materia prima

Los camiones que entregan materia prima a una fábrica de embutidos llegan siguiendo un proceso de Poisson con una tasa media de 3 camiones por hora. El tiempo entre la llegada de un camión y el siguiente sigue una distribución exponencial.

a) ¿Cuál es la probabilidad de que transcurran más de 30 minutos entre la llegada de dos camiones consecutivos? Respuesta 0.1887

b) ¿Cuál es la probabilidad de que llegue entre 1 y 2 horas? Respuesta 0.2031

c) ¿Cuál es la probabilidad de que transcurran menos de 40 minutos entre la llegada de dos camiones consecutivos? Respuesta 0.1993

3.5. Distribución normal.

La distribución normal, también conocida como distribución gaussiana o campana de Gauss, es un concepto fundamental en estadística y probabilidad. Su importancia radica en su frecuente aparición en una amplia gama de fenómenos naturales y procesos industriales, incluyendo el sector alimentario.

La distribución normal se caracteriza por su forma simétrica en forma de campana, donde la mayoría de los datos se concentran alrededor de la media, con una disminución gradual hacia los extremos. Esta distribución se define por dos parámetros: la media (μ), que representa el valor central, y la desviación estándar (σ), que mide la dispersión de los datos alrededor de la media. Una de las propiedades más importantes de la distribución normal es que

aproximadamente el 68% de los datos se encuentra dentro de una desviación estándar de la media, el 95% dentro de dos desviaciones estándar y el 99.7% dentro de tres desviaciones estándar.

En la industria alimentaria, la distribución normal se aplica en diversas áreas. Un ejemplo clave es el control de calidad. Durante la producción de alimentos, se miden constantemente diversas variables, como el peso de los productos, el contenido de humedad, la concentración de un aditivo o la temperatura de cocción. Estas mediciones suelen seguir una distribución normal. Al analizar la distribución de estas variables, los productores pueden identificar si el proceso está bajo control y si los productos cumplen con los estándares de calidad establecidos. Por ejemplo, si el peso de un paquete de galletas se distribuye normalmente con una media de 200 gramos y una desviación estándar de 5 gramos, se puede esperar que la gran mayoría de los paquetes pesen entre 195 y 205 gramos. Si se detectan valores atípicos fuera de este rango, se pueden tomar medidas correctivas para ajustar el proceso de producción.

Otro ámbito de aplicación es la optimización de procesos. Al comprender la distribución normal de las variables del proceso, se pueden realizar ajustes para mejorar la eficiencia y reducir la variabilidad. Por ejemplo, en el envasado de líquidos, la precisión del llenado puede seguir una distribución normal. Al analizar esta distribución, se pueden identificar las fuentes de variabilidad y realizar ajustes en la maquinaria o en los procedimientos operativos para minimizar las pérdidas por sobrellenado o subllenado (Martínez y Martínez, 2011).

Además, la distribución normal es fundamental en la investigación y el desarrollo de nuevos productos alimentarios. Al realizar estudios sensoriales o pruebas de vida útil, se recopilan datos que a menudo siguen una distribución normal. El análisis de estos datos permite obtener conclusiones sobre las preferencias de los consumidores o la estabilidad del producto a lo largo del tiempo. Por ejemplo, al evaluar el sabor de un nuevo refresco, se pueden realizar pruebas con un panel de catadores y analizar la distribución de las puntuaciones otorgadas. Esto permite determinar si el producto tiene una aceptación generalizada o si existen diferencias significativas en las preferencias (Daniel, 1995).

Sin embargo, es importante tener en cuenta que no todas las variables en la industria alimentaria siguen una distribución normal. En algunos casos, las distribuciones pueden ser asimétricas o presentar otras formas. En estos casos, se pueden utilizar otras distribuciones de probabilidad o aplicar transformaciones a los datos para aproximarlos a una distribución normal (Conde, 2015).

En conclusión, la distribución normal es una herramienta estadística de gran utilidad en la industria alimentaria. Su aplicación en el control de calidad, la optimización de procesos y la investigación contribuye a mejorar la eficiencia, garantizar la calidad de los productos y comprender mejor las necesidades de los consumidores. Al comprender los principios de la

distribución normal y su aplicación práctica, los profesionales del sector alimentario pueden tomar decisiones más informadas y mejorar la competitividad de sus empresas.

Para el cálculo de probabilidades correspondientes a una distribución normal, se emplea la distribución normal estándar, la cual es un caso especial de la distribución normal donde la media es igual a 0 ($\mu = 0$) y la desviación estándar es igual a 1 ($\sigma = 1$). Cualquier variable con distribución normal se puede transformar a una distribución normal estándar mediante la siguiente fórmula de estandarización.

$$Z = \frac{X - \mu}{\sigma}$$

Donde:

- Z es el número de desviaciones estándar de un punto a la media.
- X es el valor de la variable.
- μ es la media de la variable.
- σ es la desviación estándar de la variable.

Al igual que en las otras distribuciones de probabilidad de variables aleatorias continuas, el cálculo de probabilidades es para un intervalo o rango comprendido entre dos valores, ya que la probabilidad para un valor en específico se considera prácticamente cero.

A continuación, se presenta un ejemplo de aplicación de la distribución normal.

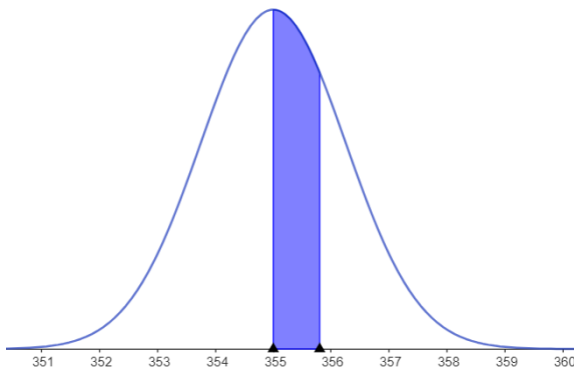
Ejemplo 1.- Considerando que el volumen de las latas de refresco se distribuye de forma normal con una media de 355 mililitros y una desviación estándar de 1.25 mililitros. Cuál es la probabilidad de que un refresco tomado completamente al azar, tenga un volumen:

- a) De 355 a 355.8 mililitros.
- b) De 354.5 a 355 mililitros.
- c) Mayor a 356 mililitros.
- d) Menor a 353.8 mililitros.
- e) Menor a 355.5 mililitros.
- f) Mayor a 354 mililitros.
- g) De 355.4 a 356 mililitros.
- h) De 353.8 a 354.7 mililitros.
- i) De 353.1 a 356.2 mililitros.

Solución.

a) De 355 a 355.8 mililitros.

La representación gráfica de la pregunta corresponde a.



La respuesta de probabilidad se obtiene calculando el valor de Z y luego mediante la tabla de la distribución de probabilidad estándar se obtiene el correspondiente valor de probabilidad.

$$Z = \frac{X - \mu}{\sigma}$$

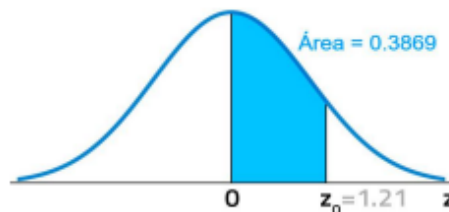
$$Z = \frac{X - \mu}{\sigma} = \frac{355.8 - 355}{1.25} = 0.64$$

Ubicando el valor de 0.64 en la tabla de distribución normal estándar.

$$Z = \frac{x - \mu}{\sigma}$$

$\mu = \text{media}$

$\sigma = \text{desviación estándar}$



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441

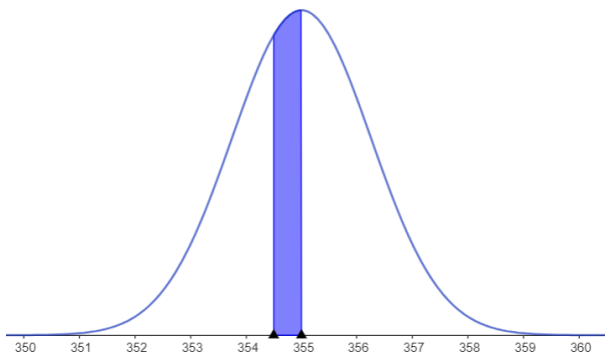
El correspondiente valor de probabilidad es de 0.2389

Existe un 23.89 % de probabilidad de que el refresco tenga un volumen de 355 a 355.8 mililitros.

Ver solución de este inciso con Excel dando clic en el siguiente link <https://youtu.be/ElVPvEOv9J0>

b) De 354.5 a 355 mililitros.

La representación gráfica de la pregunta corresponde a.



La respuesta de probabilidad se obtiene calculando el valor de Z y luego mediante la tabla de la distribución de probabilidad estándar se obtiene el correspondiente valor de probabilidad.

$$Z = \frac{X - \mu}{\sigma}$$

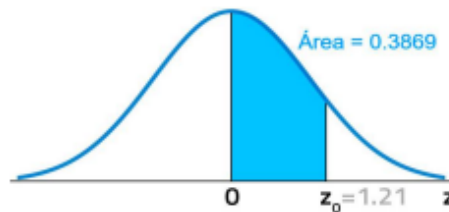
$$Z = \frac{X - \mu}{\sigma} = \frac{354.5 - 355}{1.25} = -0.40$$

El valor negativo de Z indica que el valor de X se encuentra a la izquierda de la media y simplemente en la tabla de la distribución normal estándar consultas el valor positivo, que viene a ser el valor de 0.4.

$$z = \frac{x - \mu}{\sigma}$$

$\mu = \text{media}$

$\sigma = \text{desviación estándar}$



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224

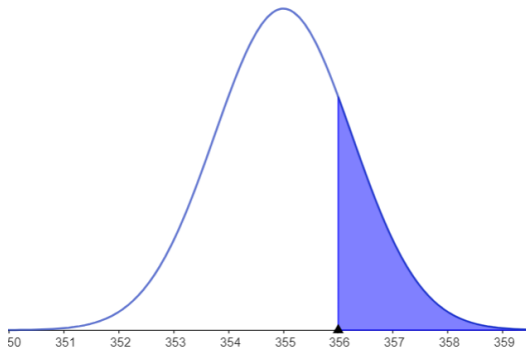
El correspondiente valor de probabilidad es de 0.1554

Existe un 15.54 % de probabilidad de que el refresco tenga un volumen de 354.5 a 355 mililitros.

Ver solución de este inciso con Excel dando clic en el siguiente link <https://youtu.be/l2KoGh1OzuU>

c) Mayor a 356 mililitros.

La representación gráfica de esta presunta es.



Recordar que el valor de Z es la cantidad de desviaciones estándar de un punto a la media, por lo que en este caso se calculará del valor de Z que hay de la media al valor de 356, luego se obtendrá su probabilidad correspondiente, la cual se restará a 0.5, porque lo que se desea es la probabilidad o área bajo la curva correspondiente a un volumen mayor a 356 mililitros.

El valor de Z es

$$Z = \frac{X - \mu}{\sigma}$$

$$Z = \frac{X - \mu}{\sigma} = \frac{356 - 355}{1.25} = 0.80$$

La probabilidad o área bajo la curva corresponde a 0.2881.

$$P(X \geq 356) = 0.5 - P(355 \leq X \leq 356)$$

El valor de 0.5 corresponde a la mitad o 50 % del área bajo la curva de la distribución normal.

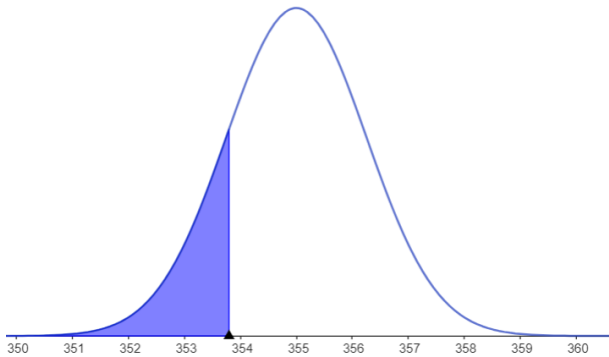
$$P(X \geq 356) = 0.5 - 0.2881 = 0.2119$$

Existe 21.19 % de probabilidad de que el refresco tenga un volumen mayor a 356 mililitros.

Ver solución de este inciso con Excel dando clic en el siguiente link <https://youtu.be/1uKup-GKzBw>

d) Menor a 353.8 mililitros.

La representación gráfica de esta presunta es.



En este caso el valor de Z será negativo porque el punto se encuentra a la izquierda de la media, pero la consulta del valor de Z en la tabla de la distribución normal estándar será con el valor positivo de Z, luego como se desea el área bajo la curva que se encuentra a la izquierda del punto, se procederá a restarle a 0.5 el área bajo la curva del valor de Z consultado.

$$Z = \frac{X - \mu}{\sigma}$$

$$Z = \frac{X - \mu}{\sigma} = \frac{353.8 - 355}{1.25} = -0.96$$

La probabilidad o área bajo la curva corresponde a 0.3315.

$$P(X \leq 353.8) = 0.5 - P(353.8 \leq X \leq 355)$$

El valor de 0.5 corresponde a la mitad o 50 % del área bajo la curva de la distribución normal.

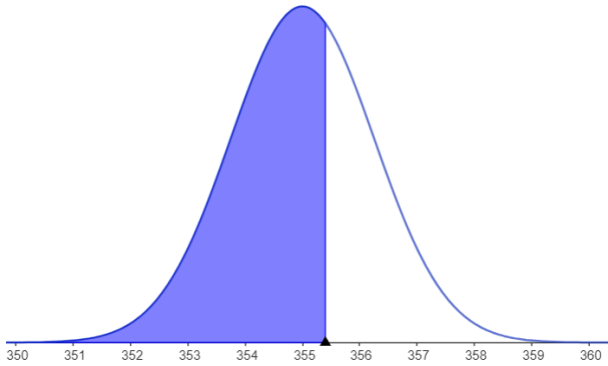
$$P(X \leq 353.8) = 0.5 - 0.3315 = 0.1685$$

Existe 16.85 % de probabilidad de que el refresco tenga un volumen menor a 353.8 mililitros.

Ver solución de este inciso con Excel dando clic en el siguiente link
<https://youtu.be/UhT0XUka4A0>

e) Menor a 355.5 mililitros.

La representación gráfica de esta presunta es.



En esta gráfica se observa que la respuesta a esta pregunta corresponde a la suma del área que se encuentra a la izquierda de la media de 355 mililitros, que es 50 % más el área que se encuentra bajo la curva en el intervalo de 355 a 355.5 mililitros.

$$Z = \frac{X - \mu}{\sigma}$$

$$Z = \frac{X - \mu}{\sigma} = \frac{355.5 - 355}{1.25} = 0.40$$

La probabilidad o área bajo la curva corresponde a 0.1554.

$$P(X \leq 355.5) = 0.5 + P(355 \leq X \leq 355.5)$$

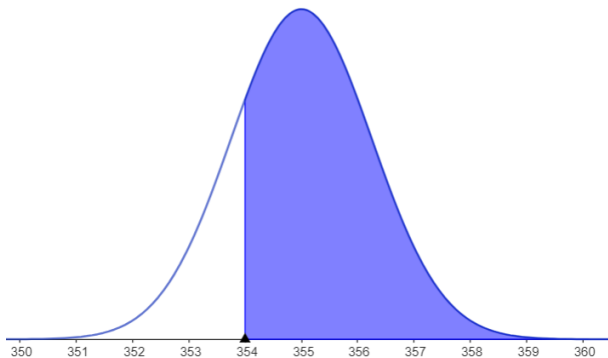
$$P(X \leq 355.5) = 0.5 + 0.1554 = 0.6554$$

Existe 65.54 % de probabilidad de que el refresco tenga un volumen menor a 355.5 mililitros.

Ver solución de este inciso con Excel dando clic en el siguiente link <https://youtu.be/UVpERgbbw-w>

f) Mayor a 354 mililitros.

La representación gráfica de esta pregunta es.



En esta gráfica se observa que la respuesta a esta pregunta corresponde a la suma del área que se encuentra a la derecha de la media de 355 mililitros, que es 50 %, más el área que se encuentra bajo la curva en el intervalo de 354 a 355 mililitros.

$$Z = \frac{X - \mu}{\sigma}$$

$$Z = \frac{X - \mu}{\sigma} = \frac{354 - 355}{1.25} = -0.80$$

La probabilidad o área bajo la curva corresponde a 0.2881.

$$P(X \geq 354) = 0.5 + P(354 \leq X \leq 355)$$

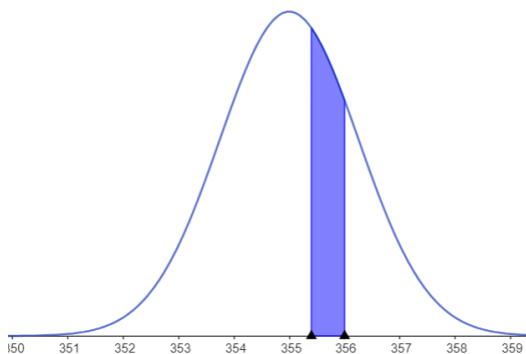
$$P(X \geq 354) = 0.5 + 0.2881 = 0.7881$$

Existe 78.81 % de probabilidad de que el refresco tenga un volumen mayor a 354 mililitros.

Ver solución de este inciso con Excel dando clic en el siguiente link <https://youtu.be/42aIK8jlx-s>

g) De 355.4 a 356 mililitros.

La representación gráfica de esta pregunta es.



Para este caso es necesario calcular la probabilidad de que el refresco tenga un volumen de 355 a 356 mililitros y luego restarle la probabilidad correspondiente a que un refresco tenga un volumen de 355 a 355.4 mililitros.

$$Z = \frac{X - \mu}{\sigma} = \frac{356 - 355}{1.25} = 0.80 \quad \text{por lo tanto, su probabilidad es de 0.2881}$$

$$Z = \frac{X - \mu}{\sigma} = \frac{355.4 - 355}{1.25} = 0.32 \quad \text{por lo tanto, su probabilidad es de 0.1255}$$

$$P(355.4 \leq X \leq 356) = P(355 \leq X \leq 356) - P(355 \leq X \leq 355.4)$$

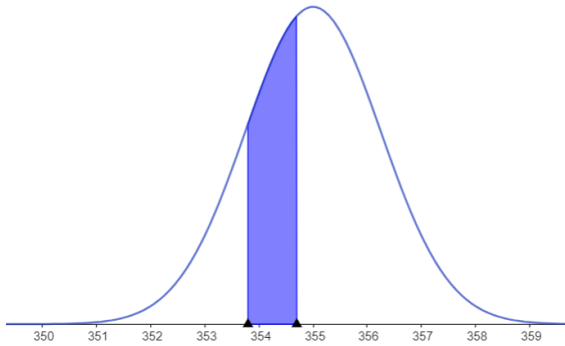
$$P(355.4 \leq X \leq 356) = 0.2881 - 0.1255 = 0.1626$$

Existe 16.26 % de probabilidad de que el refresco tenga un volumen de 355.4 a 356 mililitros.

Ver solución de este inciso con Excel dando clic en el siguiente link
<https://youtu.be/ZXuCsRCoRDo>

h) De 353.8 a 354.7 mililitros.

La representación gráfica de esta pregunta es.



Para este caso es necesario calcular la probabilidad de que el refresco tenga un volumen de 353.8 a 355 mililitros y luego restarle la probabilidad correspondiente a que un refresco tenga un volumen de 354.7 a 355 mililitros.

$$Z = \frac{X - \mu}{\sigma} = \frac{353.8 - 355}{1.25} = -0.96 \quad \text{por lo tanto, su probabilidad es de 0.3315}$$

$$Z = \frac{X - \mu}{\sigma} = \frac{354.7 - 355}{1.25} = -0.24 \quad \text{por lo tanto, su probabilidad es de 0.0948}$$

$$P(353.8 \leq X \leq 354.7) = P(353.8 \leq X \leq 355) - P(354.7 \leq X \leq 355)$$

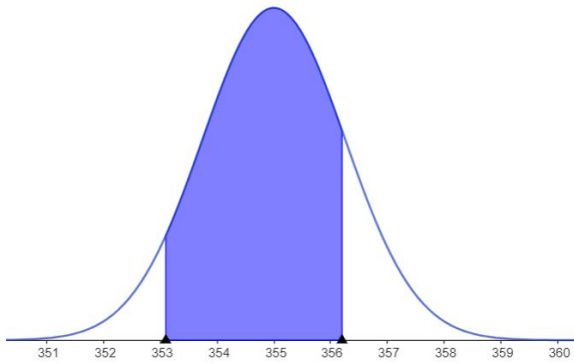
$$P(353.8 \leq X \leq 354.7) = 0.3315 - 0.0948 = 0.2367$$

Existe 23.67 % de probabilidad de que el refresco tenga un volumen de 353.8 a 354.7 mililitros.

Ver solución de este inciso con Excel dando clic en el siguiente link
https://youtu.be/969U3f_knmg

i) De 353.1 a 356.2 mililitros.

La representación gráfica de esta pregunta es.



Para este caso es necesario calcular la probabilidad de que el refresco tenga un volumen de 353.1 a 355 mililitros y luego sumarle la probabilidad correspondiente a que un refresco tenga un volumen de 355 a 356.2 mililitros.

$$Z = \frac{X - \mu}{\sigma} = \frac{353.1 - 355}{1.25} = -1.52 \quad \text{por lo tanto, su probabilidad es de 0.4357}$$

$$Z = \frac{X - \mu}{\sigma} = \frac{356.2 - 355}{1.25} = 0.96 \quad \text{por lo tanto, su probabilidad es de 0.3315}$$

$$P(353.1 \leq X \leq 356.2) = P(353.1 \leq X \leq 355) - P(355 \leq X \leq 356.2)$$

$$P(353.1 \leq X \leq 356.2) = 0.4357 + 0.3315 = 0.7672$$

Existe 76.72 % de probabilidad de que el refresco tenga un volumen de 353.1 a 356.2 mililitros.

Ver solución de este inciso con Excel dando clic en el siguiente link https://youtu.be/9_xIWsxFdQI

Ejercicios propuestos.

Ejercicio 1. Tiempo de horneado de galletas

Considerando que el tiempo de horneado de galletas sigue una distribución normal con una media de 12 minutos y una desviación estándar de 1 minuto.

a) ¿Cuál es la probabilidad de que un lote de galletas se hornee entre 11 y 13 minutos?
Respuesta 0.6827

b) ¿Cuál es la probabilidad de que un lote de galletas se hornee en menos de 12.4 minutos?
Respuesta 0.6554

c) ¿Cuál es la probabilidad de que un lote de galletas se hornee entre 12.5 y 13.8 minutos?
Respuesta 0.2726

d) ¿Cuál es la probabilidad de que un lote de galletas se hornee entre 10 y 11.5 minutos?
Respuesta 0.2858

Ejercicio 2. Peso de los aguacates

Si el peso de los aguacates en un supermercado se distribuye normalmente con una media de 200 gramos y una desviación estándar de 30 gramos.

- a) ¿Cuál es la probabilidad de que un aguacate pese menos de 150 gramos? Respuesta 0.0478
- b) ¿Cuál es la probabilidad de que un aguacate pese más de 180 gramos? Respuesta 0.7475
- c) ¿Cuál es la probabilidad de que un aguacate pese menos de 285 gramos? Respuesta 0.9977
- d) ¿Cuál es la probabilidad de que un aguacate pese de 150 a 265 gramos? Respuesta 0.9371

Ejercicio 3. Volumen de botellas de aceite

Suponiendo que el volumen de aceite contenido en botellas de 1 litro ($\mu=1000\text{mls}$) se distribuye normalmente con una desviación estándar de 5 ml. Cual es la probabilidad de que una botella seleccionada completamente al azar tenga un volumen:

- a) ¿Menor a 993 mililitros? Respuesta 0.0808
- b) ¿Mayor a 1006 mililitros? Respuesta 0.1151
- c) ¿Entre 994 a 1008 mililitros? Respuesta 0.8301
- d) ¿Entre 1002 a 1009 mililitros? Respuesta 0.3086

Ejercicio 4. pH del jugo de naranja

El pH del jugo de naranja recién exprimido se distribuye normalmente con una media de 3.8 y una desviación estándar de 0.2.

- a) ¿Cuál es la probabilidad de que una muestra de jugo tenga un pH inferior a 3.5?
Respuesta 0.0668
- b) ¿Cuál es la probabilidad de que una muestra de jugo tenga un pH superior a 3.7?
Respuesta 0.6915
- c) ¿Cuál es la probabilidad de que una muestra de jugo tenga un pH de 3.7 a 4.2?
Respuesta 0.6687

d) ¿Cuál es la probabilidad de que una muestra de jugo tenga un pH inferior a 4.3?
Respuesta 0.9937

Ejercicio 5. Peso de los filetes de pollo

El peso de los filetes de pollo envasados se distribuye normalmente con una media de 150 gramos y una desviación estándar de 15 gramos. Cuál es la probabilidad de que un filete seleccionado completamente al azar.

a) ¿Pese entre 135 y 165 gramos? Respuesta 0.6827

b) ¿Pese más de 155 gramos? Respuesta 0.3694

c) ¿Pese menos de 158 gramos? Respuesta 0.7031

d) ¿Pese entre 140 y 148 gramos? Respuesta 0.1945

3.5.1. Aproximación de la Distribución binomial a la normal.

La aproximación de la distribución binomial a la distribución normal se realiza en ciertas ocasiones debido a la complejidad de calcular probabilidades binomiales para valores grandes de n (número de ensayos). El cálculo directo de probabilidades binomiales implica el uso de factoriales que se vuelven complejos para n grandes. La distribución normal ofrece una forma más sencilla de estimar estas probabilidades.

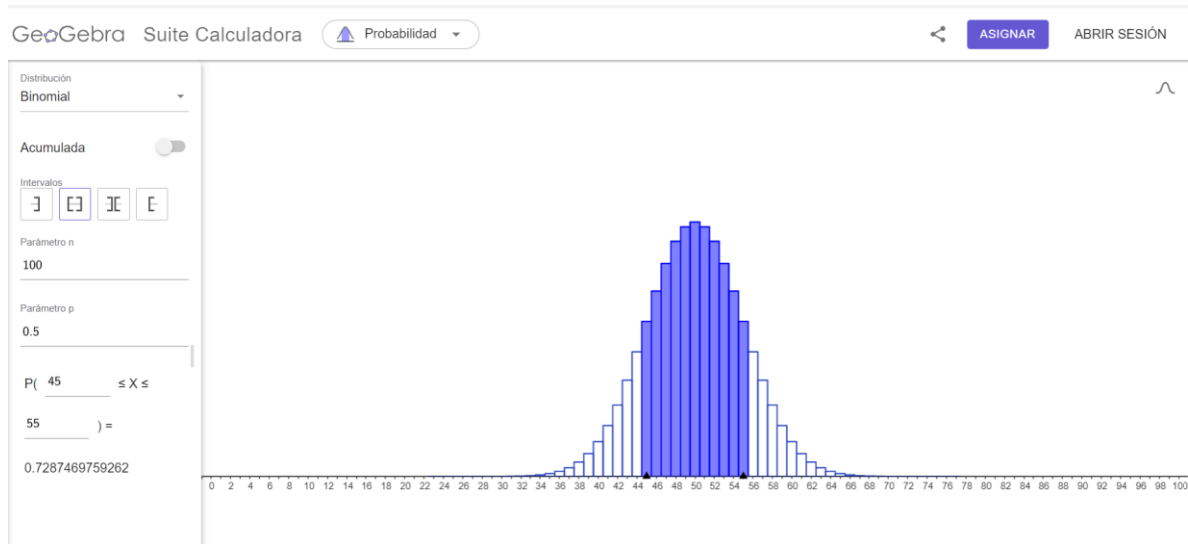
La aproximación de la binomial a la normal es adecuada cuando el número de ensayos es grande, algunos autores sugieren que resulta adecuado cuando $n \geq 30$, siempre y cuando la probabilidad de éxito p no esté demasiado cerca de 0 o 1. Adicional a lo anterior se tiene que si la probabilidad de éxito p es cercano a 0.5, con tamaños de muestra pequeños, se logra una buena aproximación de la binomial a la normal.

Para verificar lo anteriormente expuesto, se presentan los siguientes ejemplos.

Ejemplo 1. Supongamos que se lanza una moneda 100 veces, y queremos calcular la probabilidad de obtener de 45 a 55 caras.

En este caso se cumple con la condición de que el número de ensayos de 100 es mayor a 30 y adicional se tiene que la probabilidad de éxito p es igual a 0.5, de tal forma que es de esperarse una buena aproximación.

La solución de este ejercicio considerando que es una distribución binomial y usando GeoGebra, se tiene que la probabilidad de obtener de 45 a 55 caras es de 0.7287, tal como se aprecia en la siguiente imagen.

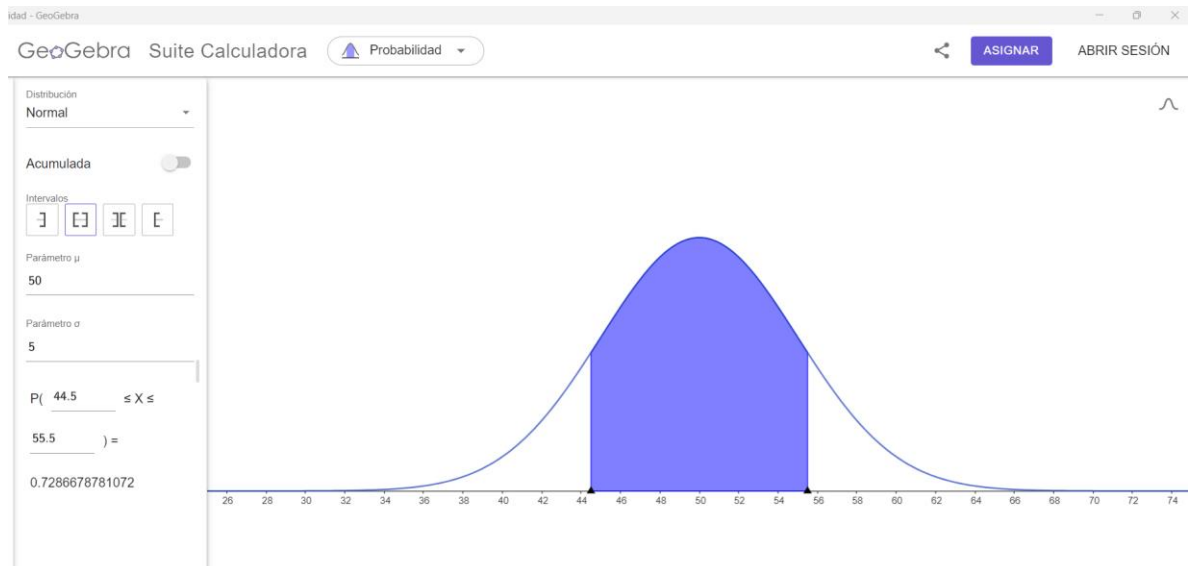


Considerando que la media y la desviación estándar de una distribución de probabilidad binomial son respectivamente:

$$\mu = np \quad \sigma = \sqrt{npq}$$

Resultando para el caso de análisis, en una media de 50 con una desviación estándar de 5.

La solución de este mismo ejercicio usando una distribución normal y el software GeoGebra es de 0.7286, tal como se puede observar en la siguiente imagen.



Observe que el intervalo discreto de 45 a 55 al transformarlo a intervalo continuo es 44.5 a 55.5

De lo anterior se tiene que el resultado obtenido mediante la distribución normal es muy parecido al obtenido mediante la distribución binomial, por lo tanto, se puede decir que la aproximación de la binomial por la normal resulta ser adecuada.

Ejercicios propuestos.

Ejercicio 1: Producción de Galletas

Una máquina produce galletas, y se sabe que el 5% de ellas salen defectuosas. Se toma una muestra aleatoria de 300 galletas. Usando la distribución binomial y la distribución normal, calcule la probabilidad de obtener de 16 a 20 galletas defectuosas. Compare los resultados para establecer en qué medida se aproxima la distribución binomial mediante la distribución normal. Resultado: Distribución Binomial 0.3542, Distribución Normal 0.3747

Ejercicio 2: Preferencias de Consumidores

Una encuesta revela que el 60% de los consumidores prefiere el yogurt natural sin azúcar. Se selecciona una muestra aleatoria de 40 consumidores. Usando la distribución binomial y la distribución normal, calcule la probabilidad de obtener de 26 a 30 consumidores que prefieren el yogurt natural sin azúcar. Compare los resultados para establecer en qué medida se aproxima la distribución binomial mediante la distribución normal. Resultado: Distribución Binomial 0.3019, Distribución Normal 0.2962

Ejercicio 3: Semillas de Tomate

Un paquete de semillas de tomate indica que tiene una tasa de germinación del 85%. Se siembran 180 semillas. Usando la distribución binomial y la distribución normal, calcule la probabilidad de que la cantidad de semillas germinadas sea de 140 a 160. Compare los resultados para establecer en qué medida se aproxima la distribución binomial mediante la distribución normal. Resultado: Distribución Binomial 0.9424, Distribución Normal 0.9389

Ejercicio 4: Conservas de tomate.

Una empresa produce conservas de tomate. Se sabe que, debido a variaciones en el proceso de envasado, un 3% de las latas presentan algún tipo de defecto (abolladuras, cierre defectuoso, etc.). Se toma una muestra aleatoria de 500 latas de la producción diaria. Calcula la probabilidad exacta de encontrar más de 19 latas defectuosas usando la distribución binomial, aproxima la probabilidad del apartado anterior usando la distribución normal y compara los resultados obtenidos. Resultado: Distribución Binomial 0.1214, Distribución Normal 0.1191

Ejercicio 5: Aceptación de un nuevo aditivo alimentario

Una empresa desarrolla un nuevo aditivo para mejorar la textura de un tipo de pan. Se realiza una prueba de degustación con 200 consumidores. Se estima que la probabilidad de que un consumidor apruebe el nuevo aditivo es del 60%. ¿Cuál es la probabilidad de que al menos

100 consumidores aprueben el aditivo? Realiza los cálculos usando la distribución binomial y la distribución normal. Compara los resultados. Resultado: Distribución Binomial 0.9983, Distribución Normal 0.9981

3.6 Distribución t de student y Chi cuadrada

La distribución t de Student, una distribución de probabilidad continua, juega un papel crucial en la inferencia estadística, especialmente cuando se trabaja con muestras pequeñas y se desconoce la desviación estándar de la población (Mendenhall, Beaver, & Beaver, 2010). Desarrollada por William Sealy Gosset bajo el seudónimo de "Student" a principios del siglo XX, esta distribución permite realizar pruebas de hipótesis y construir intervalos de confianza de manera más precisa en situaciones donde la distribución normal estándar podría no ser adecuada.

A principios del siglo XX, Gosset, un químico que trabajaba para la cervecería Guinness, se enfrentaba al problema de analizar datos con muestras pequeñas en el control de calidad de la cerveza. Las muestras pequeñas dificultaban la aplicación directa de la distribución normal, ya que la estimación de la desviación estándar de la población a partir de la muestra era imprecisa. Gosset desarrolló la distribución t para abordar este problema, permitiendo realizar inferencias estadísticas válidas incluso con información limitada. Este contexto histórico es crucial para entender la necesidad y el impacto de la distribución t.

La distribución t de Student se define formalmente como la distribución del cociente:

$$t = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}}$$

Donde:

- $\bar{X}$ es la media muestral.
- μ es la media poblacional.
- s es la desviación estándar muestral.
- n es el tamaño de la muestra.

Un parámetro clave de la distribución t es los *grados de libertad* (gl), que generalmente se calculan como $n - 1$. Los grados de libertad reflejan la cantidad de información independiente disponible para estimar la variabilidad.

La distribución t es simétrica y tiene forma de campana, similar a la distribución normal estándar, pero con colas más pesadas. Esto significa que presenta una mayor probabilidad de observar valores extremos en comparación con la normal (Moore, McCabe, & Craig, 2014).

A medida que aumentan los grados de libertad (es decir, el tamaño de la muestra), la distribución t se aproxima cada vez más a la distribución normal estándar.

La distribución t de Student tiene diversas aplicaciones en estadística:

Pruebas de hipótesis para la media: Se utiliza para realizar pruebas de hipótesis sobre la media de una población cuando se desconoce la desviación estándar poblacional y el tamaño de la muestra es pequeño. La prueba t de Student es un ejemplo clásico (Triola, 2018).

Construcción de intervalos de confianza para la media: Permite construir intervalos de confianza para la media poblacional cuando se cumplen las mismas condiciones que en las pruebas de hipótesis.

Análisis de regresión lineal: Se utiliza para realizar pruebas de significancia de los coeficientes de regresión.

Comparación de medias de dos muestras independientes: La prueba t de Student para muestras independientes se utiliza para determinar si existen diferencias significativas entre las medias de dos grupos.

La distribución chi cuadrada.

La distribución chi-cuadrada (χ^2) es una de las distribuciones de probabilidad continuas más utilizadas en estadística inferencial. Su versatilidad la convierte en una herramienta fundamental en diversas áreas, desde la biología y la medicina hasta la ingeniería y las ciencias sociales.

La distribución chi-cuadrada surge como la suma de los cuadrados de k variables aleatorias independientes con distribución normal estándar (con media 0 y desviación estándar 1). El parámetro k se conoce como grados de libertad y determina la forma de la distribución. A medida que aumentan los grados de libertad, la distribución se asemeja cada vez más a una distribución normal. La distribución chi-cuadrada es asimétrica positiva (sesgada a la derecha), pero a medida que aumentan los grados de libertad, la asimetría disminuye.

Una de las aplicaciones más importantes de la distribución chi-cuadrado es la prueba de bondad de ajuste. Esta prueba se utiliza para determinar si una muestra de datos se ajusta a una distribución teórica específica. Por ejemplo, se podría utilizar para verificar si la distribución de las frecuencias observadas de un determinado evento se ajusta a una distribución de Poisson o binomial. La prueba compara las frecuencias observadas con las frecuencias esperadas bajo la hipótesis de que los datos siguen la distribución teórica. Un valor chi-cuadrado grande sugiere una discrepancia significativa entre los datos observados y la distribución teórica, lo que lleva al rechazo de la hipótesis nula (que establece que los datos se ajustan a la distribución teórica).

Otra aplicación clave es la prueba de independencia, también conocida como prueba de chi-cuadrado de independencia. Esta prueba se utiliza para determinar si existe una asociación

entre dos variables categóricas. Se construye una tabla de contingencia que muestra las frecuencias observadas para cada combinación de categorías de las dos variables. Se calculan las frecuencias esperadas bajo la hipótesis de independencia, y se compara con las frecuencias observadas usando el estadístico chi-cuadrado. Un valor chi-cuadrado grande indica una fuerte evidencia en contra de la hipótesis de independencia, sugiriendo que existe una asociación entre las dos variables.

Además de estas pruebas, la distribución chi-cuadrado también se utiliza en la construcción de intervalos de confianza para la varianza de una población normal. Dado que la distribución muestral de la varianza muestral (multiplicada por los grados de libertad y dividida por la varianza poblacional) sigue una distribución chi-cuadrado, se pueden derivar intervalos de confianza para la varianza poblacional a partir de la varianza muestral observada.

Es importante destacar que la validez de las pruebas chi-cuadrado depende de ciertos supuestos. En la prueba de bondad de ajuste, se requiere que las frecuencias esperadas en cada categoría sean suficientemente grandes (generalmente se recomienda que sean al menos 5). En la prueba de independencia, se asume que las observaciones son independientes entre sí. El incumplimiento de estos supuestos puede afectar la validez de los resultados.

En resumen, la distribución chi-cuadrado es una herramienta estadística esencial con una amplia gama de aplicaciones. Su uso en pruebas de bondad de ajuste, pruebas de independencia y la construcción de intervalos de confianza para la varianza la convierte en una parte fundamental del análisis estadístico. Su comprensión y correcta aplicación son cruciales para la interpretación adecuada de los datos y la toma de decisiones informadas. Como señalan Moore, McCabe y Craig (2014), la distribución chi-cuadrado "es una herramienta invaluable para analizar datos categóricos y evaluar la bondad de ajuste de modelos estadísticos". De manera similar, Walpole, Myers, Myers y Ye (2012) destacan su importancia en la inferencia estadística relacionada con varianzas y pruebas de hipótesis que involucran datos categóricos.

Una sencilla evaluación del contenido de esta unidad puede ser realizada escaneando el siguiente código QR.



4. Estadística Descriptiva e Inferencial.

4.1. Datos no agrupados.

La industria alimentaria, un sector crucial para la subsistencia humana, se enfrenta a desafíos constantes en cuanto a la calidad, seguridad, eficiencia y sostenibilidad de sus procesos. En este contexto, la recopilación, el análisis y la interpretación de datos se han convertido en herramientas indispensables para la toma de decisiones informadas.

Los datos en la industria alimentaria abarcan una amplia gama de información, desde las características físicas y químicas de los alimentos hasta los datos de producción, distribución y consumo. Se pueden clasificar en diferentes categorías:

- **Datos de calidad:** Incluyen mediciones de pH, acidez, contenido de humedad, composición nutricional, presencia de contaminantes, características organolépticas (sabor, olor, textura), entre otros.
- **Datos de producción:** Comprenden información sobre la cantidad de materia prima utilizada, el tiempo de procesamiento, la temperatura, la presión, el rendimiento de las líneas de producción, el consumo de energía y agua, etc.
- **Datos de seguridad alimentaria:** Abarcan el registro de la temperatura de almacenamiento, la trazabilidad de los productos, los resultados de análisis microbiológicos, las fechas de caducidad, etc.
- **Datos de consumo:** Incluyen información sobre las preferencias de los consumidores, las tendencias del mercado, las ventas, las encuestas de satisfacción, etc.
- Dentro de estas categorías, podemos distinguir entre datos no agrupados y datos agrupados. Los datos no agrupados representan mediciones individuales sin clasificar. En la industria alimentaria, ejemplos de datos no agrupados serían:
 - El peso individual de cada pollo en una muestra de 100 pollos.
 - La temperatura individual de cada lata de conserva medida al salir de la autoclave.
 - El valor de pH de cada muestra individual de leche en un lote de producción.
 - Los resultados individuales de pruebas sensoriales (ej. puntuaciones de sabor en una escala del 1 al 10) dadas por un panel de catadores.
 - El análisis de datos no agrupados, aunque a veces desafiante para grandes conjuntos de datos, ofrece ventajas significativas en la industria alimentaria:
 - **Control de calidad preciso:** Permite identificar variaciones sutiles en los procesos de producción y detectar problemas específicos en lotes pequeños, como diferencias en el

peso de los productos o fluctuaciones en la temperatura de procesamiento. Esto es crucial para garantizar la calidad y la homogeneidad de los productos.

- Seguridad alimentaria mejorada: Facilita el seguimiento individual de productos y la identificación rápida de posibles contaminaciones o desviaciones en los parámetros de seguridad. Por ejemplo, el monitoreo continuo de la temperatura de almacenamiento de alimentos refrigerados a nivel individual permite detectar fallos en el sistema de refrigeración antes de que se produzca un deterioro generalizado.
- Optimización de procesos: El análisis de datos no agrupados puede revelar información valiosa sobre la eficiencia de los procesos de producción. Por ejemplo, el registro del tiempo de cocción individual de cada producto puede ayudar a identificar variaciones y optimizar los parámetros del proceso para reducir el consumo de energía o mejorar la textura del producto.
- Investigación y desarrollo: En la fase de desarrollo de nuevos productos, el análisis de datos no agrupados provenientes de pruebas sensoriales o análisis de laboratorio permite comprender mejor las características del producto y realizar ajustes precisos en la formulación.

Como señalan Montgomery y Runger (2010), el análisis estadístico, incluyendo el manejo de datos no agrupados, es esencial para la mejora continua de la calidad en la industria. Permite identificar las fuentes de variabilidad en los procesos y tomar medidas para reducirla, lo que se traduce en productos más consistentes y de mayor calidad.

Tipos de Análisis para datos no agrupados en la Industria Alimentaria:

Las técnicas estadísticas para datos no agrupados incluyen:

- Estadística descriptiva: Cálculo de medidas de tendencia central (media, mediana, moda) y dispersión (rango, varianza, desviación estándar, desviación media absoluta) para resumir las características de los datos.
- Gráficos: Diagramas de puntos, diagramas de tallo y hojas y, en ciertos casos, histogramas con intervalos muy pequeños para visualizar la distribución de los datos.
- Análisis de valores atípicos: Identificación y estudio de valores que se desvían significativamente del resto de los datos. Estos valores pueden indicar errores de medición, problemas en el proceso de producción o características particulares de algunos productos.

Como destacan Walpole, Myers, Myers y Ye (2012), la correcta aplicación de estas técnicas requiere un entendimiento profundo de los datos y del contexto en el que se han recopilado.

En conclusión, se tiene que los datos, y en particular los datos no agrupados, juegan un papel crucial en la industria alimentaria. Su análisis adecuado permite mejorar el control de calidad,

garantizar la seguridad alimentaria, optimizar los procesos de producción y apoyar la investigación y el desarrollo de nuevos productos. La capacidad de interpretar correctamente la información contenida en los datos es fundamental para la toma de decisiones estratégicas en este sector. Como concluyen Anderson, Sweeney y Williams (2012), el análisis de datos se ha convertido en una ventaja competitiva clave para las empresas de la industria alimentaria que buscan la excelencia en sus operaciones.

4.2. Datos agrupados.

La industria alimentaria, un sector vital para la sociedad, requiere un riguroso control y análisis de datos para asegurar la calidad, seguridad, eficiencia y sostenibilidad de sus procesos. El análisis de datos agrupados, en particular, juega un papel crucial en la gestión de grandes volúmenes de información generados en esta industria. Este ensayo se centra en la importancia y utilidad de los datos agrupados dentro del contexto de la industria alimentaria, abordando su definición, tipos, aplicaciones y la comparación con datos no agrupados.

Los datos agrupados son aquellos que se han organizado en categorías o intervalos de clase, perdiendo la información individual de cada dato, pero facilitando el análisis y la visualización de grandes conjuntos de datos. En la industria alimentaria, ejemplos de datos agrupados serían:

- Distribución de pesos de pollos en rangos: En lugar de registrar el peso individual de cada pollo, se registran cuántos pollos caen dentro de ciertos rangos de peso (ej., 1.5-2 kg, 2.1-2.5 kg, 2.6-3 kg).
- Frecuencia de temperaturas de latas de conserva en intervalos: Se registra cuántas latas alcanzan temperaturas dentro de rangos específicos durante el proceso de esterilización (ej., 110-115°C, 116-120°C, 121-125°C).
- Distribución de valores de pH de lotes de leche: Se registra cuántos lotes de leche tienen un pH dentro de ciertos rangos (ej., 6.5-6.6, 6.7-6.8, 6.9-7.0).
- Resultados de pruebas sensoriales agrupados por categorías: En lugar de las puntuaciones individuales, se registra cuántos catadores calificaron un producto como "excelente," "bueno," "regular" o "malo."

El análisis de datos agrupados ofrece varias ventajas en la industria alimentaria:

- Gestión de grandes volúmenes de datos: Facilita el manejo y la comprensión de grandes conjuntos de datos generados en la producción a gran escala.
- Identificación de tendencias generales: Permite identificar patrones y tendencias en los datos, como la distribución general de pesos de los productos, la variación de las

temperaturas de procesamiento a lo largo del tiempo o la preferencia general de los consumidores por ciertas características de los productos.

- Visualización efectiva: Facilita la creación de histogramas, polígonos de frecuencia y otras representaciones gráficas que permiten una mejor comprensión de la distribución de los datos.
- Comparación entre grupos: Permite comparar diferentes lotes de producción, diferentes proveedores de materia prima o diferentes formulaciones de productos.
- Como señalan Montgomery y Runger (2010), el uso de datos agrupados es fundamental en el control estadístico de procesos, permitiendo monitorear la variabilidad de los procesos de producción y detectar desviaciones que puedan afectar la calidad del producto.
- Las técnicas estadísticas para datos agrupados incluyen:
- Construcción de tablas de frecuencia: Organizar los datos en intervalos de clase y contar la frecuencia de observaciones en cada intervalo.
- Cálculo de medidas de tendencia central y dispersión para datos agrupados: Se utilizan fórmulas específicas que involucran las frecuencias y los límites de los intervalos para estimar la media, la mediana, la moda, la varianza y la desviación estándar.
- Representaciones gráficas: Histogramas, polígonos de frecuencia, ojivas y diagramas de barras para visualizar la distribución de los datos.
- La interpretación de datos agrupados en la industria alimentaria debe considerar la pérdida de información individual y el uso de estimaciones. Es importante analizar las tendencias generales, comparar distribuciones entre diferentes grupos y utilizar las representaciones gráficas para comunicar eficazmente los resultados.

En conclusión, se tiene que los datos agrupados son una herramienta esencial en la industria alimentaria para el manejo y análisis de grandes volúmenes de información. Su correcta aplicación permite optimizar procesos, controlar la calidad, monitorear la seguridad alimentaria y apoyar la toma de decisiones estratégicas. La elección entre utilizar datos agrupados o no agrupados depende del contexto específico del análisis y del equilibrio entre la necesidad de simplificar la información y la necesidad de mantener el detalle relevante.

4.2.1. Tablas de distribución de frecuencias y gráficas.

La industria alimentaria, caracterizada por la producción a gran escala y la necesidad de un estricto control de calidad, se beneficia enormemente del análisis estadístico y la visualización de datos. Los histogramas y polígonos de frecuencias son herramientas gráficas fundamentales para comprender la distribución de variables continuas, proporcionando información valiosa para la toma de decisiones en este sector.

Definición de Histogramas y Polígonos de Frecuencias:

Histograma: Un histograma es una representación gráfica de la distribución de frecuencias de una variable continua. Se compone de barras adyacentes, donde el eje horizontal representa los intervalos de clase (rangos de valores) y el eje vertical representa la frecuencia (o frecuencia relativa) de las observaciones que caen dentro de cada intervalo. El área de cada barra es proporcional a la frecuencia del intervalo.

Polígono de Frecuencias: Un polígono de frecuencias es una gráfica lineal que se construye a partir de un histograma. Se obtienen los puntos medios de la parte superior de cada barra del histograma y se unen con líneas rectas. El polígono de frecuencias representa la misma información que el histograma, pero de una forma más suavizada y continua.

En la industria alimentaria, los histogramas y polígonos de frecuencias son útiles para analizar una variedad de datos continuos, tales como:

- **Peso de productos:** Distribución de pesos de paquetes de alimentos, latas, etc.
- **Temperaturas de procesamiento:** Distribución de temperaturas durante la pasteurización, esterilización, refrigeración, etc.
- **Contenido de humedad:** Distribución del contenido de humedad en alimentos deshidratados, harinas, etc.
- **Concentraciones de componentes:** Distribución de la concentración de azúcares, grasas, proteínas, aditivos, etc.
- **Tiempos de procesamiento:** Distribución de tiempos de cocción, fermentación, enfriamiento, etc.
- **El análisis de estas distribuciones permite:**
- **Control de calidad:** Identificar variaciones en los procesos de producción y detectar desviaciones de los estándares de calidad.
- **Optimización de procesos:** Ajustar parámetros de producción para minimizar la variabilidad y mejorar la eficiencia.
- **Cumplimiento normativo:** Verificar el cumplimiento de los límites establecidos por las regulaciones sanitarias.
- **Investigación y desarrollo:** Estudiar el efecto de diferentes ingredientes o procesos en las características de los productos.

No existe una regla fija sobre la cantidad exacta de datos a partir de la cual es conveniente elaborar histogramas, polígonos de frecuencia u otras gráficas. Sin embargo, hay algunas pautas generales que ayudan a decidir cuándo es adecuado utilizarlas, dentro de las consideraciones clave se tienen las siguientes:

1.- Tamaño de la muestra (n): Esta es la consideración más importante, teniendo con ello lo siguiente:

- Muestras pequeñas (menos de 20-30 datos): Con pocos datos, las gráficas como diagramas de puntos o diagramas de tallo y hojas suelen ser más informativas que los histogramas o polígonos de frecuencia. Estos últimos pueden verse "truncados" o poco representativos con tan pocos datos.
- Muestras medianas (entre 30 y 100 datos): En este rango, los histogramas y polígonos de frecuencia comienzan a ser útiles para visualizar la distribución de los datos.
- Muestras grandes (más de 100 datos): Con grandes conjuntos de datos, los histogramas y polígonos de frecuencia son esenciales para resumir y comprender la distribución. Se hacen necesarios para visualizar patrones que serían imposibles de detectar examinando cada dato individualmente.

2.- Naturaleza de los datos:

- Datos continuos: Los histogramas y polígonos de frecuencia son especialmente adecuados para datos continuos (ej., altura, peso, temperatura, tiempo).
- Datos discretos: Para datos discretos con pocos valores distintos, los diagramas de barras suelen ser más apropiados. Sin embargo, si los datos discretos tienen muchos valores posibles, se pueden agrupar en intervalos y representarlos con un histograma.

3.- Objetivo del análisis:

- Visualizar la distribución: Si el objetivo principal es mostrar la forma de la distribución (ej., simétrica, asimétrica, multimodal), un histograma o un polígono de frecuencia son ideales.
- Comparar distribuciones: Los histogramas y polígonos de frecuencia también son útiles para comparar las distribuciones de diferentes grupos o conjuntos de datos.
- Mostrar tendencias a lo largo del tiempo: Para mostrar cambios en los datos a lo largo del tiempo, los gráficos de líneas (series de tiempo) son más adecuados.

4.- Recomendaciones específicas:

- Histogramas: Generalmente se recomienda usar al menos 30 datos para construir un histograma que represente adecuadamente la distribución. Con menos datos, la forma del histograma puede verse influenciada por fluctuaciones aleatorias.
- Polígonos de frecuencia: Se utilizan principalmente para comparar dos o más distribuciones o para mostrar la forma de una distribución con más detalle que un histograma. Requieren una cantidad similar de datos que los histogramas.
- Gráficos de dispersión: Se utilizan para mostrar la relación entre dos variables. Se necesita al menos un par de datos por cada observación (un valor para cada variable). No

hay un límite superior estricto, pero con demasiados puntos el gráfico puede volverse confuso.

- Gráficos de barras: Son adecuados para variables categóricas o discretas con pocos valores. No hay un límite inferior estricto en la cantidad de datos, pero con muy pocos datos la gráfica puede no ser necesaria.
- Gráficos circulares (de pastel): Se utilizan para mostrar la proporción de cada categoría en relación con el total. No se recomiendan para comparar muchas categorías, ya que pueden volverse difíciles de interpretar.

La mejor manera de decidir qué tipo de gráfica utilizar es considerar la naturaleza de los datos y el mensaje que se desea transmitir. A veces, probar diferentes tipos de gráficas ayuda a encontrar la que mejor se adapta a las necesidades del caso.

Un ejemplo de aplicación en la industria alimentaria puede ser que una empresa produce leche pasteurizada y registra la temperatura de pasteurización de cada lote. Después de recolectar datos de 100 lotes, se construye un histograma y un polígono de frecuencias. El análisis de estas gráficas permite observar la distribución de las temperaturas, identificar si la mayoría de los lotes se pasteurizaron dentro del rango de temperatura recomendado y detectar posibles desviaciones que podrían afectar la calidad o la seguridad del producto. Si se observa una alta variabilidad en las temperaturas, la empresa puede investigar las causas y tomar medidas correctivas para optimizar el proceso de pasteurización.

En conclusión, se tiene que los histogramas y polígonos de frecuencias son herramientas valiosas para el análisis y la visualización de datos continuos en la industria alimentaria. Su correcta aplicación, junto con otras gráficas complementarias, permite a las empresas tomar decisiones informadas en cuanto al control de calidad, la optimización de procesos, el cumplimiento normativo y la investigación y desarrollo de nuevos productos. La elección de la gráfica adecuada dependerá del tipo de datos y del objetivo del análisis, pero en general, para visualizar la distribución de una variable continua con una cantidad considerable de datos, los histogramas y polígonos de frecuencia son una excelente opción.

El procedimiento para la construcción de histogramas y polígonos de frecuencias, consiste en los siguientes pasos:

1. Determinar el rango de los datos: Encontrar el valor máximo, el valor mínimo y la diferencia entre ellos corresponde al rango.
2. Establecer la cantidad de clases a utilizar, lo cual se puede realizar mediante:
 - La Regla de Sturges, la cual consiste en aplicar la formula $K = 1 + 3.322 * \log(n)$ Euna de las reglas más comunes y se calcula como:

Donde k es el número de clases y n es el número de datos.

- Regla de la raíz cuadrada. Para lo cual se calcula la raíz cuadrada del número de datos. $K = \sqrt{n}$.
 - Eligiendo un valor de K de 5 a 15 y evaluando si los histograma y polígonos son satisfactorios o no.
3. Establecer la amplitud de clase, para lo cual se divide el rango entre la cantidad K de clases.
 4. Contar la frecuencia de observaciones en cada intervalo: Determinar cuántos datos caen dentro de cada intervalo.
 5. Dibujar el histograma: Representar los intervalos en el eje horizontal y las frecuencias en el eje vertical. Dibujar barras adyacentes cuya altura corresponda a la frecuencia de cada intervalo.
 6. Construir el polígono de frecuencias: Marcar el punto medio de la parte superior de cada barra del histograma y unir estos puntos con líneas rectas.

A continuación, se presenta un ejemplo de elaboración de histogramas y polígonos de frecuencia.

Ejemplo 1.- Control de peso en la producción de bolsas de papas fritas

Una empresa productora de papas fritas envasa sus productos en bolsas cuyo peso nominal es de 150 gramos. Para asegurar la calidad y el cumplimiento con las regulaciones de peso, se realiza un control de calidad tomando una muestra aleatoria de 100 bolsas de papas fritas de la línea de producción. Los pesos (en gramos) de las bolsas se registran a continuación:

153.52	146.74	154.89	151.76	150.40	150.73	154.55	150.25
149.05	153.57	150.41	144.73	154.54	151.51	150.52	144.56
153.17	150.02	148.44	150.47	149.81	144.59	145.08	151.75
148.88	149.83	151.57	150.42	145.37	149.71	154.95	149.98
150.18	152.39	153.01	150.91	148.73	146.90	147.98	146.61
152.47	145.91	153.98	147.71	153.04	149.19	147.95	147.69
148.72	152.22	145.90	150.33	153.10	152.38	147.06	144.63
147.77	154.24	148.38	145.46	149.02	145.64	144.11	150.28
145.06	152.79	144.51	151.07	148.64	148.80	149.58	146.74
151.11	147.27	145.40	144.28	151.05	152.16	145.16	152.66

Realice lo siguiente:

1. Construir una tabla de distribución de frecuencias.

2. Elaborar un histograma y polígono de: frecuencia, frecuencia relativa, frecuencia acumulada y frecuencia relativa acumulada.

3. Analizar e interpretar los resultados:

a)¿Cuál es el peso más frecuente de las bolsas?

b)¿Cómo se distribuyen los pesos alrededor del peso nominal de 150 gramos?

c)¿Hay bolsas con pesos significativamente diferentes al peso nominal (valores atípicos)?

d)¿El proceso de envasado es consistente?

Solución.

1. Construir una tabla de distribución de frecuencias.

Para construir la tabla de frecuencias es necesario establecer la cantidad de clases, la amplitud de clase, definir los límites de cada clase y realizar el conteo.

Minino	144.11
Máximo	154.95
Rango	10.84
n	80
K	7.32
Amplitud	1.355

La amplitud de clase se obtuvo de dividir el rango entre la cantidad de clase K, el rango corresponde a la diferencia de máximo menos el mínimo y la cantidad de clases K se calculó mediante la Regla de Sturges ($K = 1 + 3.322 * \log(n)$), redondeando el valor obtenido a 8.

En la tabla de frecuencias se puede observar las 8 clases, cada una con su límite inferior y superior, la primera clase inicia con el valor mínimo y la octava clase termina con el valor máximo, cada clase tiene la amplitud de clase establecida en el paso anterior (1.355).

Adicional a lo anterior se presenta cuantos datos quedaron en cada una de las clases (frecuencia), pudiendo apreciarse el correcto conteo al momento de que el total de las frecuencias (80) coincide con la cantidad total de datos.

La clase media corresponde a promedio de los límites inferior y superior.

La frecuencia relativa indica que porcentaje de los datos corresponde a cada una de las clases, el cual se calcula dividiendo la cantidad de datos de la clase entre el total de datos y multiplicando por 100, por ejemplo, para la clase, se procede a dividir 13 entre 80 y multiplicar por 100, obteniendo con ello 16.25 %.

La frecuencia acumulada corresponde al total que se tiene hasta cada una de las clases, por ejemplo, la frecuencia acumulada de la clase 4 correspondiente a 37, se obtuvo de la suma de los valores de las clases 1, 2, 3 y 4, los cuales respectivamente son: 13, 6, 8 y 10. De forma análoga sucede con la frecuencia relativa acumulada.

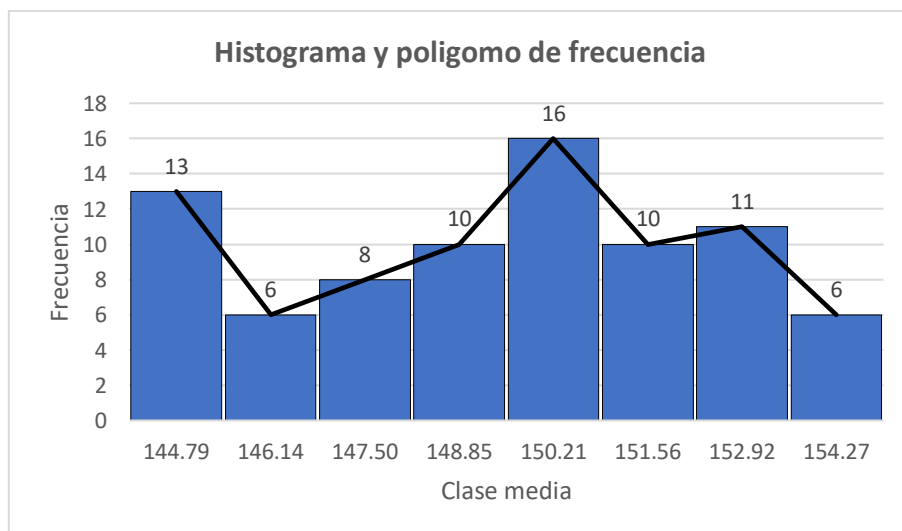
Clase	límite inferior	límite superior	Clase media	Frecuencia	Frecuencia Relativa (%)	Frecuencia Acumulada	Frecuencia Relativa Acumulada (%)
1	144.11	145.47	144.79	13	16.25	13	16.25
2	145.47	146.82	146.14	6	7.5	19	23.75
3	146.82	148.18	147.50	8	10	27	33.75
4	148.18	149.53	148.85	10	12.5	37	46.25
5	149.53	150.89	150.21	16	20	53	66.25
6	150.89	152.24	151.56	10	12.5	63	78.75
7	152.24	153.60	152.92	11	13.75	74	92.5
8	153.60	154.95	154.27	6	7.5	80	100

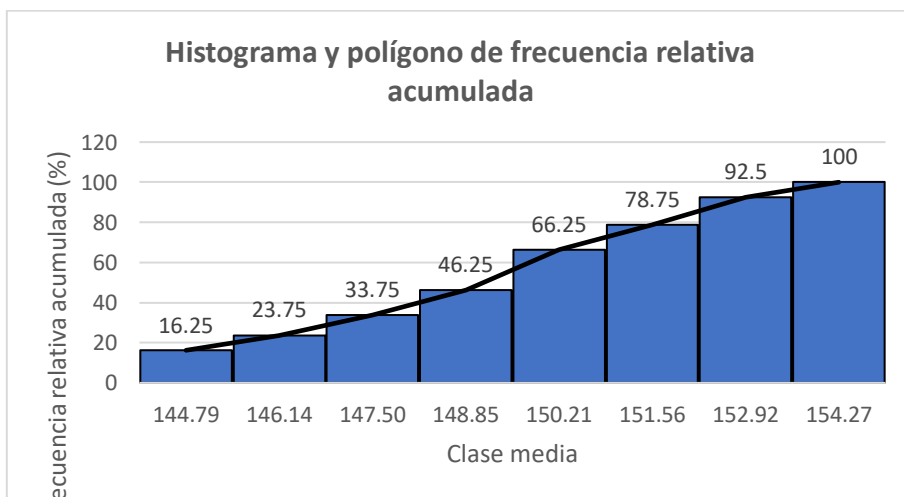
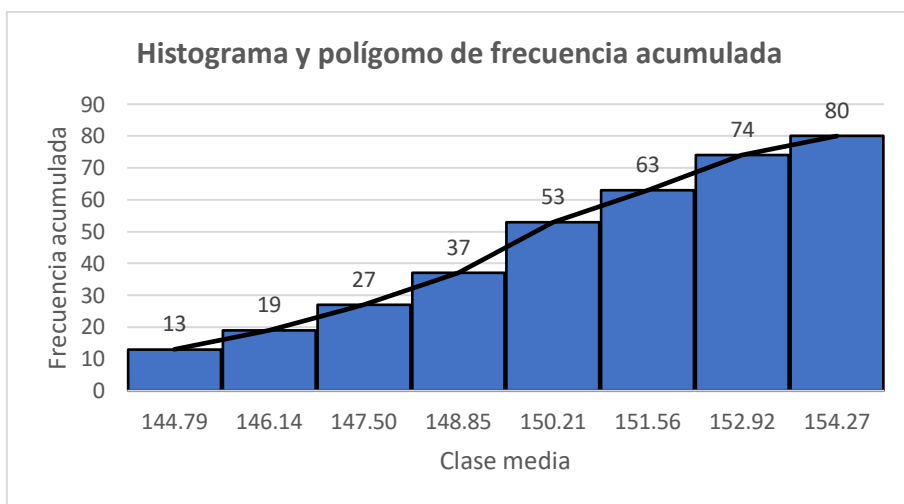
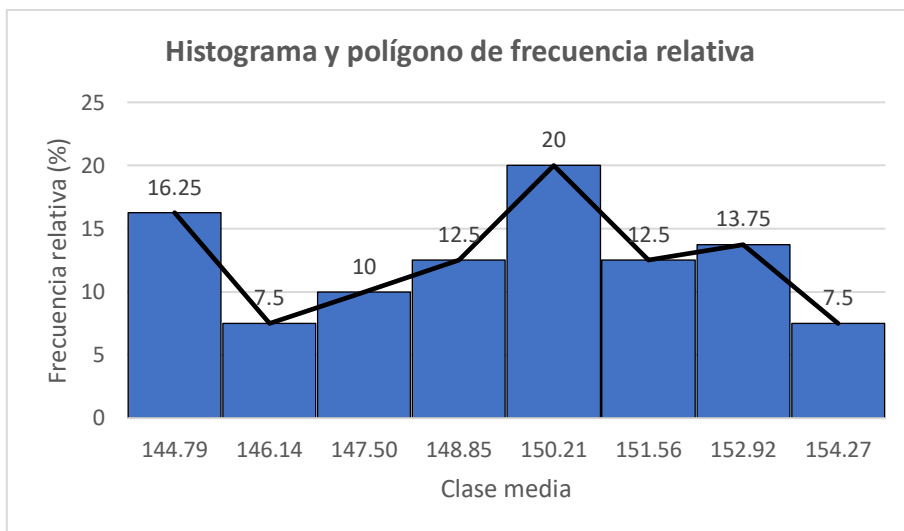
80

100

2. Elaborar un histograma y polígono de: frecuencia, frecuencia relativa, frecuencia acumulada y frecuencia relativa acumulada.

Los histogramas y polígonos de frecuencia, frecuencia relativa, frecuencia acumulada y frecuencia relativa acumulada, se presentan a continuación.





Ver como elaborar estos histogramas en Excel dando clic en el siguiente enlace
<https://youtu.be/ZoTKeWrW4Fs>

3. Analizar e interpretar los resultados:

a) ¿Cuál es el peso más frecuente de las bolsas?

El peso más frecuente de las bolsas de papas oscila de 149.53 a 150.58 gramos (ver histograma de frecuencia)

b) ¿Cómo se distribuyen los pesos alrededor del peso nominal de 150 gramos?

Los pesos de las bolsas se distribuyen de forma irregular alrededor de la media con cierta inclinación hacia los pesos bajos, ya que el 66.25 % de las bolsas tienen un peso de 150 gramos hacia abajo. (ver histograma de frecuencia relativa acumulada)

c) ¿Hay bolsas con pesos significativamente diferentes al peso nominal (valores atípicos)?

Es destacable que hay una cantidad significativa de bolsas con bajo peso (144.11 a 145.47 gramos), correspondiente a 13 bolsas, que equivalen al 16.25 % del total.

d) ¿El proceso de envasado es consistente?

En base a observar los histogramas y polígonos de frecuencia, se puede considerar que el proceso de envasado no es consistente, que es conveniente realizar ajustes.

El ejemplo que se presenta a continuación corresponde a un conjunto de datos discretos, para lo cual es conveniente la elaboración de una gráfica de barras y/o un diagrama circular.

El administrador de la paletería arcoíris registro durante un día las ventas de las nieves de yogurt en sus diferentes sabores, teniendo con ello lo siguiente:

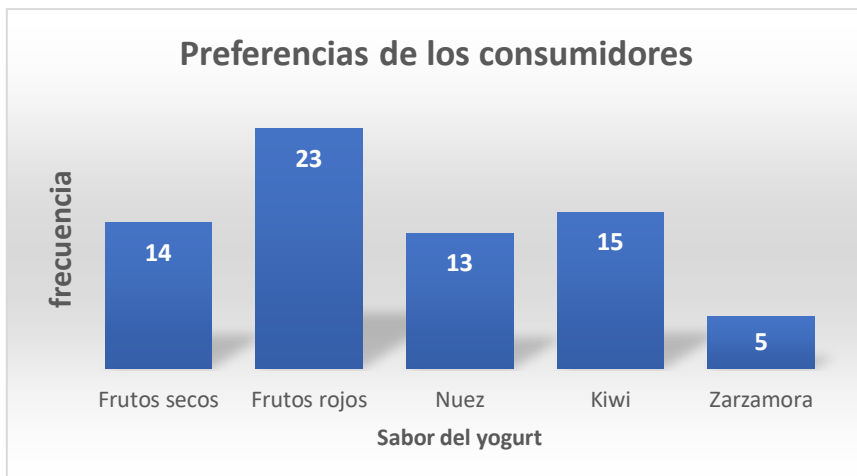
Frutos secos	Zarzamora	Frutos secos	Frutos rojos	Frutos rojos	Nuez	Zarzamora
Frutos rojos	Zarzamora	Frutos rojos	Frutos secos	Frutos rojos	Kiwi	Nuez
Nuez	Kiwi	Frutos secos	Kiwi	Nuez	Kiwi	Frutos rojos
Nuez	Kiwi	Frutos rojos	Nuez	Frutos rojos	Frutos rojos	Kiwi
Frutos rojos	Frutos secos	Kiwi	Frutos secos	Frutos secos	Kiwi	Nuez
Frutos rojos	Kiwi	Frutos secos	Frutos secos	Frutos rojos	Zarzamora	Frutos rojos
Frutos rojos	Nuez	Kiwi	Frutos rojos	Kiwi	Nuez	Kiwi
Frutos secos	Frutos secos	Frutos rojos	Frutos secos	Frutos rojos	Kiwi	Nuez
Nuez	Frutos secos	Nuez	Zarzamora	Kiwi	Nuez	Frutos rojos
Frutos secos	Frutos rojos	Frutos rojos	Frutos rojos	Frutos rojos	Frutos rojos	Kiwi

Para elaborar las gráficas solicitadas es necesario obtener la tabla de frecuencias, para lo cual se hace un listado de los diferentes sabores de nieve de yogurt y posteriormente se realiza el conteo. El resultado de ello es.

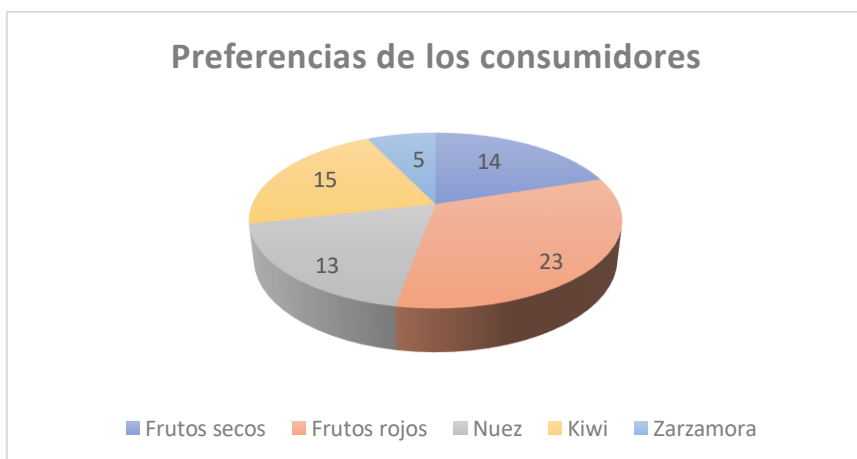
Sabor	Frecuencia
Frutos secos	14
Frutos rojos	23
Nuez	13
Kiwi	15
Zarzamora	5

70

La gráfica de barras correspondiente a esta tabla de frecuencias es.



Su gráfica circular o de pastel es.



Ver como elaborar este tipo de gráficas en Excel dando clic en el siguiente link
<https://youtu.be/24VgQfwJH6k>

Ejercicios propuestos.

Ejercicio 1.

El jefe del departamento de control de calidad de una embotelladora de refrescos solicito que se midiera el volumen de los últimos 150 refrescos que se envasaron en la línea de producción de refrescos en lata. Los volúmenes obtenidos en mililitros, se presentan a continuación:

353.21	353.34	355.8	355.35	354.62	356.23	355.84	355.92	353.21	354.62
355.61	354.63	354.47	355.36	354.59	355.42	356.22	353.64	354.53	354.97
355.29	353.88	355.31	356.06	353.63	355.72	354.92	354.63	355.18	353.39
354.79	353.85	355.75	355.91	354.61	353.46	356.24	353.34	355.18	353.9
354.3	353.8	354.48	355.89	355.26	355.36	355.32	355.72	353.37	354.08
353.24	355.95	354.71	356.15	354.33	355.65	355.14	356.34	354.28	355.47
354.12	355.93	354.46	355.33	354.04	355.53	355.81	355.4	354.21	354.54
354.06	355.46	353.64	355.26	353.45	356.21	355.84	354.45	353.84	355.23
354.94	354.85	356.3	355.68	355.69	353.22	354.16	355.93	356.07	353.41
356.35	354.53	356.38	354.56	354.05	353.28	353.25	353.25	354.56	355.13
354.22	355.78	353.89	354.75	353.64	354.06	354.14	355.91	355.79	354.37
355.62	353.9	356.34	354.54	353.52	353.9	355.22	355.8	355.82	355.6
354.62	356.22	353.43	353.44	355.86	354.8	353.64	354.65	356	355.1
353.88	353.53	355.33	355.17	353.87	355.97	356.02	353.61	354.19	353.76
355.88	353.81	353.28	354.34	355.17	353.3	353.33	354.21	356.34	355.96

Realice lo siguiente:

- Elabore una tabla de distribución de frecuencias.
- Elaborar un histograma y polígono de: frecuencia, frecuencia relativa, frecuencia acumulada y frecuencia relativa acumulada.

Ejercicio 2.

Del libro de registros de la empresa Tierra Mojada se obtuvieron los datos del peso de las truchas Arcoíris que fueron capturadas en días pasados, siendo estos los siguientes:

288.95	432.47	381.34	308.54	321.05	413.74	370.91	434.77	393.38
309.36	353.43	384.92	314.12	290.02	368.05	310.80	371.37	360.97
449.58	435.71	301.03	297.28	452.02	350.86	415.56	372.00	323.91
402.60	417.14	371.38	285.71	374.90	440.80	324.89	307.11	361.92
438.01	434.45	375.54	451.16	341.95	439.08	400.79	347.18	314.21
452.34	383.54	314.96	385.54	462.33	440.34	324.85	290.23	311.45
314.87	378.03	413.03	451.03	359.12	399.94	367.92	300.02	310.49
349.88	335.62	421.96	336.63	447.17	374.11	452.66	334.56	327.59
379.51	373.83	355.67	434.87	377.55	317.52	372.69	464.66	431.00
457.47	335.44	293.84	389.42	351.33	331.97	415.87	428.58	306.27
372.70	390.65	431.54	347.84	467.62	378.97	381.02	343.19	408.38

430.68	382.95	368.12	404.64	294.18	467.31	460.44	332.21	341.62
344.56	315.79	333.28	405.34	329.50	377.98	326.49	297.64	331.94

Realice lo siguiente:

- Elabore una tabla de distribución de frecuencias.
- Elaborar un histograma y polígono de: frecuencia, frecuencia relativa, frecuencia acumulada y frecuencia relativa acumulada.

Ejercicio 3.

A continuación, se presentan los datos de la concentración en grados Brix de los últimos lotes producidos de concentrado de jugo de manzana de una empresa del municipio de Cuauhtémoc Chihuahua.

70.21	71.78	71.67	69.12	69.82	71.71	71.93
70.18	70.01	69.19	69.81	69.2	70.71	71.9
71.56	71.57	69.58	71.83	71.32	69.4	69.77
70.77	71.69	70.88	69.02	69.84	71.42	71.7
71.64	69.49	70.67	71.29	69.21	70.75	71.32
69.48	69.79	70.1	69.19	71.79	70.06	70.49
69.42	69.1	70.92	71.94	71.05	69.31	69.11
69.24	69.36	69.56	70.46	70.14	71.73	69.4
70.49	70.78	69.26	71.51	71.45	69.06	70.15
70.26	69.87	71.39	71.08	69.95	70.27	71.09
69.36	70.77	71.32	70.38	71.71	71	69.09
70.53	69.34	71.77	70.18	69.72	70.21	69.15

Realice lo siguiente:

- Elabore una tabla de distribución de frecuencias.
- Elaborar un histograma y polígono de: frecuencia, frecuencia relativa, frecuencia acumulada y frecuencia relativa acumulada.

Ejercicio 4.

En una investigación de mercado con pruebas de degustación efectuada a 135 personas, en relación a una salsa de chile habanero con mango, algunas de las preguntas que se les hicieron fueron las siguientes:

- Le gusto: Si ____ No ____
- La compraría: Si ____ No ____

3.- Cuanto estaría dispuesto a pagar por un frasco con un contenido de 200 mililitros:

50 pesos _____

60 pesos _____

70 pesos _____

80 pesos _____

90 pesos _____

Después de aplicar las encuestas y realizar el conteo para cada pregunta, las respuestas fueron las siguientes.

1.- Le gusto: Si 95 personas No 40 personas

2.- La compraría: Si 80 personas No 15 personas

3.- Cuanto estaría dispuesto a pagar por un frasco con un contenido de 200 mililitros:

50 pesos 10 personas

60 pesos 15 personas

70 pesos 35 personas

80 pesos 15 personas

90 pesos 5 personas

Para las respuestas de cada pregunta, elabore una gráfica de barras y una gráfica circular o de pastel.

4.2.2. Medidas de tendencia central: Medidas de dispersión Cuartiles, Deciles y Percentiles.

Las medidas de resumen son herramientas estadísticas esenciales para sintetizar y comprender conjuntos de datos. En la industria alimentaria, donde la calidad, la seguridad y la eficiencia son primordiales, estas medidas juegan un papel crucial en diversas áreas, desde el control de calidad hasta la investigación de mercado. Las principales medidas de resumen, se clasifican en dos categorías: medidas de tendencia central y medidas de dispersión.

Medidas de Tendencia Central:

Estas medidas buscan identificar un valor central o típico dentro de un conjunto de datos. Las más comunes son:

1.- Media Aritmética (Promedio): Se calcula sumando todos los valores y dividiendo el resultado entre el número total de valores, su fórmula es.

$$Media = \frac{\sum_{i=1}^n x_i}{n}$$

Donde $\sum x_i$ es la suma de los valores de x y n es el número de valores.

Ejemplos de interpretación en la industria alimentaria:

- Peso promedio de productos: Permite verificar el cumplimiento de normativas de peso y detectar posibles desajustes en las líneas de producción. Por ejemplo, el peso promedio de un lote de galletas debe acercarse al peso declarado en el empaque.
- Contenido promedio de nutrientes: Esencial para el etiquetado nutricional. La media del contenido de grasa, azúcar o sodio permite a los consumidores tomar decisiones informadas.
- Tiempo promedio de vida útil: Crucial para la gestión de inventarios y la fijación de fechas de caducidad.
- Calificación promedio en pruebas sensoriales: Facilita la comparación de productos y la evaluación de la aceptación por parte de los consumidores.

Limitaciones: Es sensible a valores atípicos. Un único valor extremadamente alto o bajo puede distorsionar la media y no representar adecuadamente el conjunto de datos.

2.- Mediana: Es el valor central cuando los datos se ordenan de menor a mayor. Si hay un número par de datos, la mediana es el promedio de los dos valores centrales.

Ejemplos de interpretación en la industria alimentaria:

- Tamaño de partícula: En alimentos en polvo, la mediana del tamaño de las partículas es más robusta que la media ante la presencia de algunas partículas muy grandes o pequeñas.
- Preferencias de consumidores: La mediana de las calificaciones en encuestas de sabor ayuda a identificar el punto medio de las preferencias, minimizando el impacto de valoraciones extremas.

Ventajas: No se ve afectada por valores atípicos.

3.- Moda: Es el valor que aparece con mayor frecuencia. Puede haber una moda (unimodal), dos (bimodal) o más (multimodal).

Ejemplos de interpretación en la industria alimentaria:

- Sabor más popular: La moda en encuestas de sabor indica la preferencia mayoritaria.
- Tamaño de envase más vendido: Facilita la gestión de inventarios y la planificación de la producción.
- Defecto más frecuente: En control de calidad, identificar el tipo de defecto más común permite enfocar las acciones correctivas.

Un ejemplo en específico corresponde a que en una encuesta sobre el tipo de leche preferida, se obtuvieron los siguientes resultados: Entera (80 votos), Semidesnatada (120 votos), Desnatada (50 votos). La moda es la leche Semidesnatada.

Medidas de Dispersión:

Estas medidas de resumen cuantifican la variabilidad o dispersión de los datos alrededor de la tendencia central. Las más importantes son:

1.- Rango: Es la diferencia entre el valor máximo y el valor mínimo.

Interpretación en la Industria Alimentaria: Ofrece una idea rápida de la amplitud de variación en una característica. Por ejemplo, el rango de temperaturas durante el almacenamiento de un producto refrigerado.

Limitaciones: Solo considera los valores extremos, sin tener en cuenta la distribución de los datos intermedios.

2.- Varianza: Mide la dispersión promedio de los datos con respecto a la media. Se calcula como el promedio de las diferencias al cuadrado entre cada valor y la media. Su formula de cálculo es la siguiente.

$$\text{Varianza poblacional} = \frac{\sum_{i=1}^n (x_i - \text{media})^2}{n}$$

$$\text{Varianza muestral} = \frac{\sum_{i=1}^n (x_i - \text{media})^2}{n-1}$$

Interpretación en la Industria Alimentaria: Indica la consistencia de un proceso. Una baja varianza en el peso de los productos indica un proceso de envasado preciso.

Limitaciones: Se expresa en unidades al cuadrado, lo que dificulta su interpretación directa.

3.- Desviación Estándar: Es la raíz cuadrada de la varianza.

Interpretación en la Industria Alimentaria: Se interpreta en las mismas unidades que los datos originales, lo que facilita su comprensión. Una baja desviación estándar indica que los datos están concentrados cerca de la media, mientras que una alta desviación estándar indica una mayor dispersión. Por ejemplo, una baja desviación estándar en el contenido de sal en un lote de salsa indica una producción consistente.

Un ejemplo en específico corresponde a que, si la varianza del peso de las manzanas es 25 gramos cuadrados, la desviación estándar es $\sqrt{25} = 5 \text{ gramos}$.

4.- Coeficiente de Variación: Es la relación entre la desviación estándar y la media, expresada en porcentaje. Su fórmula es.

$$CV = \frac{\text{Desviación estándar}}{\text{media}} \times 100$$

Interpretación en la Industria Alimentaria: Permite comparar la variabilidad entre conjuntos de datos con diferentes unidades o escalas. Por ejemplo, se puede comparar la variabilidad del peso de las manzanas con la variabilidad del tamaño de las mismas, aunque estén medidas en unidades diferentes.

En conclusión, se tiene que las medidas de resumen son herramientas fundamentales para el análisis de datos en la industria alimentaria. Las medidas de tendencia central (media, mediana y moda) proporcionan información sobre el valor típico de un conjunto de datos, mientras que las medidas de dispersión (rango, varianza, desviación estándar y coeficiente de variación) cuantifican la variabilidad de los datos. La elección de la medida más adecuada depende del tipo de datos y del objetivo del análisis. Es crucial interpretar estas medidas en el contexto específico de la industria alimentaria, considerando las implicaciones para la calidad, la seguridad y la eficiencia de los procesos. El uso conjunto de medidas de tendencia central y dispersión proporciona una visión completa y robusta de los datos, permitiendo tomar decisiones informadas y optimizar las operaciones en la industria alimentaria.

Ejemplos de aplicación del cálculo de medidas de resumen para un conjunto de datos se presentan a continuación.

Ejemplo 1.

Considere que el volumen de 10 envases de aceite de cártamo es: 998, 995, 1002, 1002, 998, 996, 997, 1004, 997, 998 mililitros. Calcule: la media, la mediana, la moda, la varianza, la desviación estándar y el coeficiente de variación.

Media.

$$\begin{aligned} \text{Media} &= \frac{\sum_{i=1}^n x_i}{n} = \frac{998 + 995 + 1002 + 998 + 996 + 997 + 1004 + 997 + 998}{10} \\ &= \frac{9987}{10} = 998.7 \text{ mililitros} \end{aligned}$$

Mediana.

Los datos ordenados de menor a mayor se presentan a continuación

995 996 997 997 **998** **998** 998 1002 1002 1004

En este caso por ser par la cantidad de datos, la mediana se obtiene sumando el 5° y 6° dato y dividiéndolo entre 2.

$$\text{Mediana} = \frac{998 + 998}{2} = 998 \text{ mililitros.}$$

Moda.

El valor que más se repite en este conjunto de datos es 998 mililitros.

Varianza.

Considerando que los datos corresponden a una muestra, la formula para la varianza es.

$$\text{Varianza muestral} = \frac{\sum_{i=1}^n (x_i - \text{media})^2}{n-1} = \frac{78.1}{10-1} = 8.68 \text{ mililitros al cuadrado}$$

x_i	$(x_i - \text{media})^2$		
998	0.49		
995	13.69		
1002	10.89		
1002	10.89		
998	0.49		
996	7.29		
997	2.89		
1004	28.09		
997	2.89		
998	0.49		
	78.1	$\sum (x_i - \text{media})^2$	

Desviación estándar.

Recordando que la desviación estándar es la raíz cuadrada de la varianza, se tiene lo siguiente.

$$\text{Desviación estándar} = \sqrt{\text{varianza}} = \sqrt{8.68} = 2.95 \text{ mililitros.}$$

Coefficiente de variación.

$$CV = \frac{\text{Desviación estandar}}{\text{media}} \times 100 = \frac{2.95}{998.7} \times 100 = 0.295$$

Una empresa productora de papas fritas envasa sus productos en bolsas cuyo peso nominal es de 150 gramos. Para asegurar la calidad y el cumplimiento con las regulaciones de peso, se realiza un control de calidad tomando una muestra aleatoria de 100 bolsas de papas fritas de la línea de producción. Los pesos (en gramos) de las bolsas se registran a continuación:

153.52	146.74	154.89	151.76	150.40	150.73	154.55	150.25
149.05	153.57	150.41	144.73	154.54	151.51	150.52	144.56
153.17	150.02	148.44	150.47	149.81	144.59	145.08	151.75
148.88	149.83	151.57	150.42	145.37	149.71	154.95	149.98
150.18	152.39	153.01	150.91	148.73	146.90	147.98	146.61
152.47	145.91	153.98	147.71	153.04	149.19	147.95	147.69
148.72	152.22	145.90	150.33	153.10	152.38	147.06	144.63

147.77	154.24	148.38	145.46	149.02	145.64	144.11	150.28
145.06	152.79	144.51	151.07	148.64	148.80	149.58	146.74
151.11	147.27	145.40	144.28	151.05	152.16	145.16	152.66

Atendiendo los datos proporcionados, calcule: la media, la mediana, la moda, la varianza, la desviación estándar y el coeficiente de variación.

Solución.

$$Media = \frac{\sum_{i=1}^n x_i}{n} = \frac{153.52 + 149.05 + 153.17 + \dots + 146.74 + 152.06}{80} = \frac{11955.44}{80} = 149.45$$

La media o promedio de este conjunto de datos es 149.45.

Para calcular la mediana es necesario ordenarlos de menor a mayor, localizar los datos que se encuentran en la posición 40 y 41 y calcular la media de dichos datos, para este caso se tiene que los valores ubicados en las posiciones 40 y 41 son 149.81 y 149.83, de tal forma que la mediana es 149.82

La moda, correspondiente al valor que más se repite en este conjunto de datos es 146.74.

Varianza.

Considerando que el conjunto de datos proporcionados corresponde a una muestra, la formula a utilizar es.

$$Varianza\ muestral = \frac{\sum_{i=1}^n (x_i - media)^2}{n-1} = \frac{720.13}{80-1} = 9.12 \text{ gramos al cuadrado}$$

La varianza para este conjunto de datos resulto en un valor de 9.12 gramos al cuadrado.

Desviación estándar.

Teniendo en cuenta que la desviación estándar es la raíz cuadrada de la varianza, se tiene que la desviación estándar para este conjunto de datos es de 3.02 gramos.

Coeficiente de variación.

$$CV = \frac{Desviación\ estandar}{media} \times 100 = \frac{3.02}{149.45} \times 100 = 2.02$$

El coeficiente de variación CV para este conjunto de datos resulto en 2.02

Ver video de solución usando Excel para un conjunto grande de datos.

Ejercicios propuestos.

Ejercicio 1.

Teniendo como referencia que la densidad se mide con un lactodensímetro y se expresa en g/ml a una temperatura específica (generalmente 15 °C) y que la densidad normal de la leche de vaca oscila entre 1.028 y 1.033 g/ml, por lo que valores fuera de este rango pueden indicar adulteración con agua o descremado. Si en una de las queserías establecidas en el municipio de Cuauhtémoc se hicieron las pruebas de densidad a la leche proporcionada por 8 de sus proveedores obteniéndose los siguientes datos en g/ml: 1.032, 1.032, 1.029, 1.028, 1.033, 1.033, 1.028, 1.033

Calcule: la media, la mediana, la moda, la varianza, la desviación estándar y el coeficiente de variación. Respuestas.

Media	1.031
Mediana	1.032
Moda	1.033
Varianza	0.0000051
Desviación estándar	0.0023
Coeficiente de variación	0.2200

Ver video de solución usando Excel para un conjunto pequeño de datos, dando clic en el siguiente enlace <https://youtu.be/HVGB5huHw-0>

Ejercicio 2.

Se tiene el conocimiento de que un control adecuado de la actividad acuosa (A_w) ayuda a mantener las características organolépticas de los productos deshidratados, como su textura, color, sabor y aroma, durante un mayor tiempo. Si en una empresa dedicada a la producción de deshidratados de manzana se hizo la medición de los últimos 7 lotes, en los cuales los valores de actividad acuosa fueron: 0.52, 0.55, 0.54, 0.53, 0.52, 0.52, 0.55.

Calcule: la media, la mediana, la moda, la varianza, la desviación estándar y el coeficiente de variación. Respuestas.

Media	0.5329
Mediana	0.53
Moda	0.52
Varianza	0.00019
Desviación estándar	0.0138
Coeficiente de variación	2.5901

Ejercicio 3.

En una empresa dedicada a la venta de pollo es común que estos sean pesados, datos en relación a ello son los siguientes:

2.67	2.20	2.04	2.02	2.10	1.69	2.69	2.74
2.15	1.70	1.66	2.25	2.50	1.85	2.43	1.94
1.85	2.10	2.25	1.65	1.96	1.76	1.92	1.99
2.66	2.77	2.00	2.30	2.46	2.57	2.54	2.44
2.49	1.72	2.51	1.72	2.14	1.72	2.15	1.93
1.99	2.73	1.70	2.30	2.68	2.08	2.59	2.01
1.88	2.16	2.42	2.73	2.60	2.72	2.08	2.35
2.33	1.91	2.50	1.70	2.70	2.66	1.69	2.55
2.20	1.70	2.74	2.47	1.71	2.13	2.34	2.39
2.25	2.15	1.76	1.68	1.88	2.74	2.02	2.44
2.58	2.27	2.30	2.42	1.84	2.45	1.79	2.70
1.95	1.77	2.22	1.68	2.03	1.66	2.11	2.01
2.43	2.33	1.68	2.19	2.25	1.65	2.37	2.31
2.02	2.73	2.07	2.30	2.11	1.90	2.06	2.04
1.83	2.13	2.18	2.45	2.42	2.55	2.09	2.04
2.11	2.22	2.66	2.44	2.48	2.28	2.47	2.62
1.65	2.45	1.86	2.65	2.69	1.66	2.30	2.06
1.82	2.73	2.59	2.35	2.39	2.13	2.78	2.38
2.75	2.57	2.21	2.31	2.62	2.41	2.13	2.33
1.70	2.40	1.88	2.68	1.78	2.56	2.65	1.75
1.88	2.21	1.85	2.77	1.79	2.39	2.00	2.01
2.50	1.89	1.88	1.93	2.48	2.07	1.98	2.03
2.42	1.88	2.37	2.19	2.13	1.66	2.18	2.42
1.67	2.32	1.81	2.13	2.44	1.98	2.10	2.11
2.17	2.07	1.74	2.47	2.32	2.25	2.06	2.57

Calcule: la media, la mediana, la moda, la varianza, la desviación estándar y el coeficiente de variación. Respuestas.

Media	2.20
Mediana	2.19
Moda	1.88
Varianza	0.10
Desviación estándar	0.32
Coeficiente de variación	14.74

Ver video de solución usando Excel para este conjunto de datos, dando clic en el siguiente enlace <https://youtu.be/g3DlyvZpYi0>

Ejercicio 4.

A continuación, se presentan los datos de la concentración en grados Brix de los últimos lotes producidos de concentrado de jugo de manzana de una empresa del municipio de Cuauhtémoc Chihuahua.

70.21	71.78	71.67	69.12	69.82	71.71	71.93
70.18	70.01	69.19	69.81	69.2	70.71	71.9
71.56	71.57	69.58	71.83	71.32	69.4	69.77
70.77	71.69	70.88	69.02	69.84	71.42	71.7
71.64	69.49	70.67	71.29	69.21	70.75	71.32
69.48	69.79	70.1	69.19	71.79	70.06	70.49
69.42	69.1	70.92	71.94	71.05	69.31	69.11
69.24	69.36	69.56	70.46	70.14	71.73	69.4
70.49	70.78	69.26	71.51	71.45	69.06	70.15
70.26	69.87	71.39	71.08	69.95	70.27	71.09
69.36	70.77	71.32	70.38	71.71	71	69.09
70.53	69.34	71.77	70.18	69.72	70.21	69.15

Calcule: la media, la mediana, la moda, la varianza, la desviación estándar y el coeficiente de variación. Respuestas.

Media	70.41
Mediana	70.27
Moda	71.32
Varianza	0.89
Desviación estándar	0.94
Coeficiente de variación	1.34

4.3. Técnicas de Muestreo.

El muestreo es una herramienta fundamental en la industria alimentaria, ya que permite obtener información representativa sobre la calidad, seguridad e inocuidad de los alimentos sin necesidad de analizar la totalidad de la producción.

El muestreo se define como el proceso de selección de una porción representativa de un conjunto mayor, llamado población o lote, con el objetivo de estudiar sus características y extraer conclusiones que puedan generalizarse a toda la población. En el contexto

alimentario, la población puede ser un lote de producción de un determinado alimento, un cargamento de materia prima, o incluso la producción total de una fábrica en un período de tiempo (Montgomery, 2004).

La importancia del muestreo radica en su capacidad para proporcionar información valiosa con un menor costo y tiempo en comparación con el análisis del 100% de la producción. Esto es especialmente relevante en la industria alimentaria debido a:

- Grandes volúmenes de producción: Las empresas procesan grandes cantidades de alimentos, lo que hace inviable el análisis exhaustivo de cada unidad.
- Naturaleza destructiva de algunas pruebas: Algunas pruebas de calidad, como los análisis microbiológicos o la determinación de la vida útil, implican la destrucción de la muestra.
- Limitaciones de tiempo y recursos: El análisis completo de toda la producción requeriría una gran cantidad de tiempo, personal y recursos económicos.

El muestreo permite superar estas limitaciones al proporcionar una imagen confiable de la calidad del lote a partir del análisis de una muestra representativa.

El muestreo se utiliza en diversas etapas de la cadena alimentaria, con diferentes objetivos:

- Control de calidad de materias primas: Se evalúa la calidad de los ingredientes que ingresan a la planta de producción, verificando que cumplan con los estándares establecidos (calidad fisicoquímica, microbiológica, organoléptica).
- Control de calidad en el proceso de producción: Se monitorean las diferentes etapas del proceso para asegurar que se cumplan los parámetros de calidad y se detecten posibles desviaciones.
- Control de calidad del producto terminado: Se verifica que el producto final cumpla con las especificaciones de calidad y seguridad establecidas por las normativas y la empresa.
- Estudios de vida útil: Se toman muestras a lo largo del tiempo para evaluar la estabilidad del producto y determinar su fecha de caducidad.
- Investigación y desarrollo de nuevos productos: Se utilizan técnicas de muestreo para evaluar las preferencias de los consumidores y optimizar las formulaciones.
- Vigilancia sanitaria: Las autoridades sanitarias realizan muestreos para verificar el cumplimiento de las normativas y proteger la salud pública.

A continuación, se presentan algunos ejemplos de muestreo en la Industria Alimentaria:

- Muestreo de leche cruda: Se toman muestras de leche de diferentes tanques de almacenamiento en una granja lechera para analizar su contenido de grasa, proteínas, sólidos no grasos y recuento bacteriano.

- **Muestreo de latas de conservas:** Se seleccionan aleatoriamente algunas latas de un lote de producción para verificar su hermeticidad, vacío interno, peso neto y calidad del contenido.
- **Muestreo de carne molida:** Se toman muestras de diferentes puntos de un lote de carne molida para analizar su contenido de grasa, humedad y presencia de microorganismos patógenos como *Escherichia coli* O157:H7.
- **Muestreo de frutas y verduras frescas:** Se seleccionan muestras de diferentes cajas o lotes para evaluar su madurez, tamaño, color, presencia de defectos y residuos de pesticidas.
- **Muestreo de agua utilizada en la producción:** Se toman muestras de agua en diferentes puntos del sistema de abastecimiento para verificar su calidad microbiológica y fisicoquímica.

Tipos de Muestreo

La elección del tipo de muestreo adecuado es crucial para obtener resultados válidos y generalizables. A continuación, se presentan los principales tipos de muestreo, clasificándolos en probabilísticos y no probabilísticos, mostrando sus características, ventajas y desventajas, con ejemplos para una mejor comprensión.

Muestreo Probabilístico:

En el muestreo probabilístico, cada elemento de la población tiene una probabilidad conocida y no nula de ser seleccionado en la muestra. Esto permite realizar inferencias estadísticas válidas y generalizar los resultados a la población. Los principales tipos de muestreo probabilístico son:

1.- **Muestreo Aleatorio Simple (MAS):** Es el método más básico. Cada elemento de la población tiene la misma probabilidad de ser seleccionado. Se utiliza una tabla de números aleatorios o un generador de números aleatorios para seleccionar la muestra.

Ventajas: Fácil de entender y aplicar en poblaciones pequeñas y homogéneas.

Desventajas: Requiere un listado completo de la población (marco muestral), lo cual puede ser difícil de obtener en poblaciones grandes. Puede ser ineficiente si la población es heterogénea.

Ejemplo: Seleccionar al azar 50 estudiantes de una lista alfabética de todos los estudiantes de una escuela.

2.- **Muestreo Estratificado:** La población se divide en subgrupos homogéneos llamados estratos (ej., por edad, género, nivel socioeconómico). Luego, se realiza un muestreo aleatorio simple dentro de cada estrato.

Ventajas: Asegura la representación de todos los estratos en la muestra, lo que aumenta la precisión de las estimaciones, especialmente en poblaciones heterogéneas.

Desventajas: Requiere conocer la composición de la población en términos de los estratos.

Ejemplo: Realizar una encuesta sobre preferencias alimentarias dividiendo a la población por grupos de edad (niños, adolescentes, adultos, adultos mayores) y luego seleccionando una muestra aleatoria dentro de cada grupo.

3.- Muestreo Sistemático: Se selecciona un punto de partida aleatorio y luego se seleccionan elementos a intervalos regulares (ej., cada 10 personas).

Ventajas: Fácil de aplicar y eficiente cuando la población está ordenada.

Desventajas: Puede ser problemático si existe un patrón periódico en la población que coincida con el intervalo de selección.

Ejemplo: Seleccionar cada 50 botellas de un lote de producción en una línea de embotellado para controlar su volumen.

4.- Muestreo por Conglomerados: La población se divide en grupos o conglomerados (ej., escuelas, barrios, empresas). Se seleccionan algunos conglomerados al azar y luego se incluyen en la muestra todos los elementos dentro de los conglomerados seleccionados o una muestra de ellos.

Ventajas: Útil cuando la población es grande y dispersa geográficamente. Reduce los costos de muestreo.

Desventajas: Puede ser menos preciso que otros métodos si los conglomerados son heterogéneos entre sí.

Ejemplo: Realizar un estudio sobre el consumo de lácteos seleccionando al azar algunas ciudades de un país y luego encuestando a todas las familias en algunos barrios seleccionados dentro de esas ciudades.

Muestreo No Probabilístico:

En el muestreo no probabilístico, la selección de los elementos de la muestra no se basa en la probabilidad, sino en criterios subjetivos del investigador. Esto limita la posibilidad de generalizar los resultados a la población. Los principales tipos son:

1.- Muestreo por Conveniencia: Se seleccionan los elementos que están más fácilmente disponibles para el investigador.

Ventajas: Fácil y rápido de aplicar. Bajo costo.

Desventajas: No es representativo de la población. Los resultados no se pueden generalizar.

Ejemplo: Encuestar a las personas que pasan por un centro comercial.

2.- Muestreo por Juicio o Criterio: El investigador selecciona los elementos que considera más representativos de la población según su conocimiento o experiencia.

Ventajas: Útil cuando se necesita información rápida y no se requiere alta precisión.

Desventajas: Subjetivo y propenso a sesgos del investigador. No se pueden generalizar los resultados.

Ejemplo: Seleccionar expertos en un tema para una entrevista.

3.- Muestreo por Cuotas: Se establecen cuotas para diferentes categorías (ej., género, edad) y se seleccionan elementos hasta completar cada cuota.

Ventajas: Asegura la representación de diferentes grupos en la muestra.

Desventajas: La selección dentro de cada cuota no es aleatoria, lo que limita la generalización de los resultados.

Ejemplo: Realizar una encuesta de mercado estableciendo cuotas de género y edad y luego seleccionando personas hasta completar cada cuota.

4.- Muestreo de Bola de Nieve (Snowball): Se identifican algunos participantes iniciales que luego recomiendan a otros participantes.

Ventajas: Útil para estudiar poblaciones de difícil acceso.

Desventajas: No es representativo de la población. Puede haber sesgos en la selección.

Ejemplo: Estudiar una comunidad de personas con una enfermedad rara.

En conclusión, se tiene que la elección del tipo de muestreo depende del objetivo de la investigación, los recursos disponibles y las características de la población. El muestreo probabilístico es preferible cuando se requiere generalizar los resultados a la población, mientras que el muestreo no probabilístico puede ser útil en estudios exploratorios o cuando los recursos son limitados. Es crucial comprender las ventajas y desventajas de cada tipo de muestreo para seleccionar el método más adecuado y obtener resultados válidos y confiables.

Con la finalidad de comprender con mayor facilidad el proceso de muestreo y determinación del tamaño de la muestra para la estimación de la media y la proporción, se plantea el escenario de un productor avícola que tiene la necesidad de establecer cuál es el peso promedio de los 10,000 pollos que tiene en su granja, así como definir qué proporción de ellos tiene un peso mayor a 2 kilos.

Las fórmulas para el objetivo planteado en el párrafo anterior son:

		Estimador	
		Media	Proporción
Población	Infinita	$n = \left(\frac{z \sigma}{E} \right)^2$	$n = \frac{z^2 p q}{E^2}$
	Finita	$n = \frac{z^2 N \sigma^2}{E^2 (N - 1) + z^2 \sigma^2}$	$n = \frac{N z^2 p q}{E^2 (N - 1) + z^2 p q}$

En esta tabla de fórmulas se puede observar que se tienen poblaciones finitas o pequeñas y poblaciones infinitas o grandes.

Para definir a que tipo de población corresponde el caso de análisis, se calcula el tamaño de muestra aplicando la fórmula para población infinita y si el tamaño de muestra calculado es igual o mayor al tamaño de la población ($n \geq 0.05 N$), implica que la población es pequeña o finita, por lo cual es necesario recalcular el tamaño de muestra usando la fórmula para población finita, la cual contiene un factor de corrección.

Procedimiento para estimar la media o promedio.

El primer paso es aplicar la formula.

$$n = \left(\frac{z \sigma}{E} \right)^2 \text{ donde:}$$

n es el tamaño de la muestra.

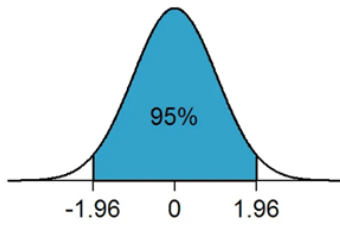
Z es el número de desviaciones estándar de un punto a la media y está en función del nivel de confianza, el cual es definido por quien lleva a cabo el estudio o la investigación, teniéndose que a mayor nivel de confianza el tamaño de muestra se incrementa.

σ es la desviación estándar de la población, la cual se puede conocer por datos históricos o mediante un muestreo previo o piloto.

E es el error de estimación, el cual es definido por quien lleva a cabo el estudio o la investigación, teniéndose que a menor error de estimación mayor tamaño de muestra.

Aplicando esta fórmula para el caso del productor avícola, se tendría lo siguiente.

Si el nivel de confianza deseado es del 95 %, el valor de z es igual a 1.96, el cual se obtiene consultando la tabla de la distribución normal estándar.

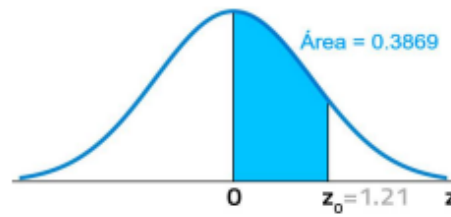


En la imagen anterior se muestra como la distribución normal es simétrica respecto a la media, teniendo que la mitad del 95 % se encuentra a la derecha de la media, o sea 47.5 %, que equivale a 0.475, que al buscarlo en el cuerpo de la tabla de la distribución normal estándar se tiene que el valor de z es igual a 1.96, tal como se muestra a continuación.

$$z = \frac{x - \mu}{\sigma}$$

$\mu = \text{media}$

$\sigma = \text{desviación estándar}$



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990
3.1	0.4990	0.4991	0.4991	0.4991	0.4992	0.4992	0.4992	0.4992	0.4993	0.4993
3.2	0.4993	0.4993	0.4994	0.4994	0.4994	0.4994	0.4994	0.4995	0.4995	0.4995
3.3	0.4995	0.4995	0.4995	0.4996	0.4996	0.4996	0.4996	0.4996	0.4996	0.4997
3.4	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4998
3.5	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998
3.6	0.4998	0.4998	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.7	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.8	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.9	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000
4.0	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000

Al no tener datos históricos para calcular la desviación estándar, considerar que se tomó una muestra completamente al azar de 20 pollos, se pesaron y los datos en gramos fueron los siguientes.

2031, 1811, 2374, 2177, 2394, 2236, 1593, 2141, 2165, 1968, 1842, 2488, 2274, 1861, 1749, 1558, 1992, 2288, 1994, 1525

La desviación estándar del peso de los pollos resulto en 285.3 gramos.

Se considero que un error de estimación (E) de 50 gramos era aceptable.

Una vez que se obtuvieron todos los datos requeridos por la fórmula de tamaño de muestra se tuvo lo siguiente.

$$n = \left(\frac{z \sigma}{E} \right)^2$$

$$n = \left(\frac{1.96 \ 285.3}{50} \right)^2 = 125.07$$

Ver video de determinación del tamaño de muestra para la media en Excel dando clic en el siguiente enlace <https://youtu.be/NpwZyTummaQ>

Como el tamaño de muestra calculado es menor al 5 % de la población (500) la población se considera grande y con aplicar esta fórmula es suficiente, en caso contrario, hubiera sido requerido aplicar la fórmula para población pequeña o finita.

La cantidad de pollos que es necesario pesar es de 125, siendo conveniente aplicar un muestreo aleatorio simple, para lo cual simplemente se pueden ir tomando sin ver un pollo y luego otro, pesarlos, anotar los datos y luego calcular la media o promedio.

Como en la muestra previa ya se habían tomado 20 pollos, solo faltan 105 para completar el tamaño de muestra de 125.

Considerar que la actividad se realizó y se obtuvieron los 125 pesos en gramos que se presentan a continuación.

2031	1842	2453	2327	2482	2148	2353	1948
1811	2488	2326	1706	1543	1662	1955	1808
2374	2274	1996	1510	1578	2441	1802	1961
2177	1861	1635	2300	1968	2100	1895	2507
2394	1749	1582	1536	1970	2255	1512	2035
2236	1558	1814	1743	1656	1698	2197	1974
1593	1992	1892	2141	1866	1866	2384	2092
2141	2288	2198	1534	2412	1929	1625	1692
2165	1994	2282	1770	1779	2182	1620	1800
1968	1525	1615	1875	2112	1896	1605	2274
1804	2368	1728	1604	2215	2351	1534	1860
2394	2187	2201	2129	2458	2062	2292	2068
2417	2086	1880	1829	2038	2280	2520	2237
1497	2245	1526	2287	2262	1641	2150	2283
2407	2556	1706	1499	1999	2110	2501	2543

1565	2315	2530	2032	1604			
------	------	------	------	------	--	--	--

De este conjunto de datos resulta que la media o promedio del peso de los pollos es de 2008.6 gramos.

Procedimiento para estimar proporción.

Recordar que el dueño de la granja avícola desea estimar que proporción de los pollos tiene un peso superior a 2000 gramos.

El primer paso es aplicar la formula.

$$n = \frac{z^2 p q}{E^2} \text{ Donde:}$$

Z es el número de desviaciones estándar de un punto a la media y está en función del nivel de confianza, el cual es definido por quien lleva a cabo el estudio o la investigación, teniéndose que a mayor nivel de confianza el tamaño de muestra se incrementa.

p es la proporción de la población, la cual se puede conocer por datos históricos o mediante un muestreo previo o piloto.

q corresponde al valor de 1 - p

E es el error de estimación, el cual es definido por quien lleva a cabo el estudio o la investigación, teniéndose que a menor error de estimación mayor tamaño de muestra.

Aplicando esta fórmula para el caso del productor avícola, se tendría lo siguiente.

Tal como se había visto anteriormente, para un nivel de confianza del 95 %, el valor de e es igual a 1.96.

Al no tener datos históricos para calcular la proporción, considerar que se tomó una muestra completamente al azar de 20 pollos, se pesaron y los datos en gramos fueron los siguientes.

2031, 1811, 2374, 2177, 2394, 2236, 1593, 2141, 2165, 1968, 1842, 2488, 2274, 1861, 1749, 1558, 1992, 2288, 1994, 1525

De estos 20 pollos, 10 de ellos pesan más de 2000 gramos, por lo tanto, la proporción p es igual a 50 % que equivale a 0.5 y adicional a ello se tiene que q es igual a 0.5 (q = 1 - p).

Se considero que un error de estimación (E) de 5% era aceptable.

Una vez que se obtuvieron todos los datos requeridos por la fórmula de tamaño de muestra se tuvo lo siguiente.

$$n = \frac{z^2 p q}{E^2}$$

$$n = \frac{z^2 p q}{E^2} = \frac{1.96^2 * 0.5 * 0.5}{0.05^2} = 384.15$$

La cantidad de pollos que deben ser pesados es de 385 (se redondea hacia arriba).

Ver video de determinación del tamaño de muestra para la proporción en Excel dando clic en el siguiente enlace <https://youtu.be/-rBtLuAcj7c>

Como el tamaño de muestra calculado es menor al 5 % de la población (500) la población se considera grande y con aplicar esta fórmula es suficiente, en caso contrario, hubiera sido requerido aplicar la fórmula para población pequeña o finita.

La cantidad de pollos que es necesario pesar es de 385, siendo conveniente aplicar un muestreo aleatorio simple, para lo cual simplemente se pueden ir tomando sin ver un pollo y luego otro, pesarlos, anotar los datos y luego calcular la proporción.

Como en la muestra previa ya se habían tomado 20 pollos, solo faltan 365 para completar el tamaño de muestra de 385.

Considerar que la actividad se realizó y se obtuvieron los 385 pesos en gramos que se presentan a continuación.

2031	1842	2453	2327	2482	2148	2353	1948	1766
1811	2488	2326	1706	1543	1662	1955	1808	1671
2374	2274	1996	1510	1578	2441	1802	1961	2078
2177	1861	1635	2300	1968	2100	1895	2507	2156
2394	1749	1582	1536	1970	2255	1512	2035	2505
2236	1558	1814	1743	1656	1698	2197	1974	1639
1593	1992	1892	2141	1866	1866	2384	2092	1620
2141	2288	2198	1534	2412	1929	1625	1692	2762
2165	1994	2282	1770	1779	2182	1620	1800	2307
1968	1525	1615	1875	2112	1896	1605	2274	2108
1804	2368	1728	1604	2215	2351	1534	1860	2760
2394	2187	2201	2129	2458	2062	2292	2068	2390
2417	2086	1880	1829	2038	2280	2520	2237	2269
1497	2245	1526	2287	2262	1641	2150	2283	2624
2407	2556	1706	1499	1999	2110	2501	2543	2071
1565	2315	2530	2032	1604	1964	2295	2363	2552
2655	1591	2378	1992	2088	2375	2317	2317	2108
1808	2497	1853	1915	2304	2512	2036	2466	2698
1701	1785	2438	1986	1874	1648	2741	2583	2423
2124	1597	2567	2806	2679	1819	2043	1770	2065
2285	2630	2574	2526	2601	1818	2324	2550	2096
2141	2836	1626	2251	1971	2377	2693	2183	2403

2844	1661	2746	2850	2251	2423	1931	2767	2833
2558	1667	2743	2255	1707	1720	2272	2022	2751
2171	2619	2710	2053	2235	2626	2100	1994	2702
2391	1737	2299	2845	2670	2827	1600	2658	1903
2822	1708	2715	2679	2406	1920	2297	2621	2815
2220	1843	2541	2111	1659	1830	1846	2293	2844
1711	2055	2370	2073	2133	1930	2669	1999	1586
1853	2647	1611	1583	1611	2274	2345	2280	2340
1660	2323	2538	2524	1877	2706	2191	1675	2086
2558	2592	2431	2253	2747	2109	2817	2795	2369
2124	1892	2408	1922	2362	2262	1881	1869	2689
1636	1576	2031	1964	2283	2033	1794	2597	1959
2095	1700	2038	2389	2430	1631	2117	2031	1606
2580	1889	2778	2836	1612	1770	2606	2787	1742
2089	2354	2538	2520	2329	1910	2092	2529	2614
2028	2816	2842	2550	2748	2035	1685	2506	2657
2711	1933	1633	2415	1624	2161	2193	2538	1846
2482	2232	1573	2142	2068	1664	2577	1871	2807
1973	2272	1921	2708	1891	1997	2209	1937	2444
1970	2830	1571	2532	2817	2147	1697	2181	1662
1897	1728	2065	1932	2369	1589	2217		

De estos datos, resulta que 237 son superiores a 2000 gramos, de forma que la proporción de pollos con un peso superior a 2000 gramos es de.

$$\hat{p} = \frac{237}{385} = 0.6156$$

El 61.56 % de los pollos tienen un peso superior a 2000 gramos.

Como el tamaño de muestra para la proporción resulto ser mayor que el tamaño de muestra requerido para la media, resulta conveniente utilizar esos datos con que ya se cuenta para recalcular la media, esto con la finalidad de tener una mejor estimación de la misma.

Por lo anterior, se tiene que la media o peso promedio de los pollos es de 2150.24 gramos.

Ejemplo 2 de cálculo del tamaño de la muestra para estimar la media y la proporción.

Considere que desea vender donas de chocolate en el Instituto Tecnológico de Cd. Cuauhtémoc que cuenta con una población entre alumnos, maestros y personal administrativo de 2650 personas.

Antes de emprender usted decide llevar a cabo una investigación de mercado con pruebas de degustación, con la finalidad de establecer de que magnitud pueden ser las ventas por semana, por lo que elabora la encuesta siguiente.

- 1.- Le gusto la dona de chocolate. Si_____ No_____
- 2.- La compraría. Si_____ No_____
- 3.- Cuantas veces a la semana. _____
- 4.- Cuanto estaría dispuesto a pagar (pesos) por una dona. _____

Usted sabe que necesita estimar ciertos datos para aplicar las fórmulas de tamaño de muestra y verificar que las preguntas de la encuesta son entendibles, por lo que decide aplicar una encuesta piloto con pruebas de degustación a 30 personas.

Orientado a realizar un muestreo aleatorio simple decide que se va a colocar en la entrada del edificio C y les va a aplicar a las primeras 30 personas que pasen por el lugar.

Una vez que se aplicó el plan, se obtuvieron los siguientes datos.

- 1.- Le gusto la dona de chocolate. Si 25 No 5
- 2.- La compraría. Si 20 No 5
- 3.- Cuantas veces a la semana. 1, 3, 3, 3, 3, 2, 3, 2, 1, 1, 2, 3, 3, 2, 1, 1, 3, 2, 3, 1
- 4.- Cuanto estaría dispuesto a pagar (pesos) por una dona. 13, 12, 13, 10, 16, 16, 12, 11, 13, 16, 12, 12, 13, 15, 12, 16, 12, 13, 15, 11

Ahora se desea establecer el tamaño de la muestra para estimar el precio promedio y la proporción de personas que si comprarían las donas.

Las fórmulas para ello son.

		Estimador	
		Media	Proporción
Población	Infinita	$n = \left(\frac{z \sigma}{E} \right)^2$	$n = \frac{z^2 p q}{E^2}$
	Finita	$n = \frac{z^2 N \sigma^2}{E^2 (N - 1) + z^2 \sigma^2}$	$n = \frac{N z^2 p q}{E^2 (N - 1) + z^2 p q}$

Estimación de la media o promedio.

Aplicar la formula $n = \left(\frac{z \sigma}{E} \right)^2$.

Para un nivel de confianza del 95 % ($z = 1.96$), un error de estimación (E) de 25 centavos (0.25 pesos).

El valor de la desviación estándar se estima en base a los datos obtenidos del precio que están dispuestos a pagar, los cuales son: 13, 12, 13, 10, 16, 16, 12, 11, 13, 16, 12, 12, 13, 15, 12, 16, 12, 13, 15, 11. La desviación estándar estimada resultó en un valor de 1.87 pesos.

$$n = \left(\frac{z \sigma}{E} \right)^2$$

$$n = \left(\frac{z \sigma}{E} \right)^2 = \left(\frac{1.96 * 1.87}{0.25} \right)^2 = 215.28 \text{ Es necesario aplicar 216 encuestas.}$$

Como 216 es mayor al 5% de la población ($0.05 \times 2650 = 132.5$), es necesario aplicar la fórmula para población finita.

$$n = \frac{z^2 N \sigma^2}{E^2 (N - 1) + z^2 \sigma^2}$$

$$n = \frac{1.96^2 * 2650 * 1.87^2}{0.25^2 (2650 - 1) + 1.96^2 1.87^2} = 199.18$$

Es necesario aplicar 200 encuestas

Ver video de calculo de tamaño de encuesta para determinar el precio promedio de las donas dando clic en el siguiente enlace <https://youtu.be/xXHdy29xjM8>

Estimación de la proporción.

Aplicar la formula $n = \frac{z^2 p q}{E^2}$

Para un nivel de confianza del 95 % ($z = 1.96$), un error de estimación (E) del 5 % (0.05).

Considerando que, de 30 personas encuestadas, 20 si las comprarían, se tiene que la proporción estimada es de 0.6666 (20/30).

$$q = 1 - p = 1 - 0.6666 = 0.3333$$

$$n = \frac{z^2 p q}{E^2}$$

$$n = \frac{z^2 p q}{E^2} = \frac{1.96^2 * 0.6666 * 0.3333}{.05^2} = 341.5 \text{ Es necesario aplicar 342 encuestas.}$$

Como 342 es mayor al 5% de la población ($0.05 \times 2650 = 132.5$), es necesario aplicar la fórmula para población finita.

$$n = \frac{N z^2 p q}{E^2 (N - 1) + z^2 p q}$$

$$n = \frac{2650 * 1.96^2 * 0.6666 * 0.3333}{.05^2 (2650 - 1) + 1.96^2 * 0.6666 * 0.3333} = 302.59 \text{ Es necesario aplicar 303 encuestas.}$$

Ver video de cálculo de tamaño de encuesta para determinar la proporción de personas que comprarían donas dando clic en el siguiente enlace <https://youtu.be/lPzw5WyiVFo>

De los cálculos del tamaño de muestra para la media y para la proporción, se toma el de mayor magnitud, siendo el de 303, correspondiente a la proporción.

Considerar que se aplicó un muestreo aleatorio simple para obtener las 273 encuestas faltantes, por el hecho de que ya se habían aplicado 30 y se obtuvieron los siguientes resultados.

De las 303 personas encuestadas 233 dijeron que, si las comprarían, dando con ello una proporción de 0.7690 (76.9 %).

En relación a la pregunta 3.

3.- Cuantas veces a la semana.

Las respuestas fueron.

1	3	3	3	3	2	3	2	1	1	2	3	3	2	1
2	2	2	3	2	2	3	1	1	3	3	3	1	3	1
1	3	2	3	1	3	1	3	1	1	2	2	1	1	1
1	3	1	2	3	3	3	3	1	1	2	3	3	1	1
1	2	3	2	1	1	2	3	1	3	2	2	3	3	2
1	3	1	2	1	2	1	2	1	2	1	2	1	3	2
3	3	2	1	2	2	2	2	2	1	2	1	3	1	3
1	2	3	2	2	3	1	2	2	3	1	2	1	2	1
1	3	2	3	1	1	3	1	3	2	2	3	2	3	3
2	3	3	2	1	2	1	1	1	2	1	2	1	3	1
3	2	1	3	1	2	1	3	2	2	3	2	1	2	1
1	3	1	1	1	1	1	3	1	2	2	1	2	1	2
3	2	1	1	3	1	3	1	2	1	1	3	1	2	1
1	1	3	3	3	2	2	3	3	2	3	3	2	1	2
3	2	3	2	2	3	2	2	1	2	1	1	3	3	2
3	3	1	2	3	3	3	1							

La media o promedio de frecuencia de compra a la semana resulto en 1.97, prácticamente 2 donas por semana.

En relación a la pregunta 4.

4.- Cuanto estaría dispuesto a pagar (pesos) por una dona.

Las respuestas fueron.

15	15	13	13	13	12	12	13	15	14	14	15	13	14	12
12	12	14	15	14	14	12	15	14	12	12	15	15	12	14
14	13	13	12	15	14	13	15	13	13	14	14	15	13	13
14	13	14	15	15	12	15	12	14	13	13	13	15	12	15
13	14	12	15	14	15	13	12	12	14	14	14	15	15	15
12	14	12	13	13	13	13	12	13	15	14	12	13	13	12
15	12	12	15	15	15	15	12	12	15	14	12	13	14	12
14	14	14	15	13	13	14	14	15	12	14	15	14	12	12
14	15	12	14	15	12	15	12	12	14	13	12	12	14	12
12	15	12	14	13	12	14	14	13	14	13	12	15	13	14
12	15	12	12	14	15	15	15	14	12	14	14	13	15	15
14	12	14	14	12	13	14	14	13	15	15	12	12	14	14
15	13	12	12	12	13	13	15	14	12	13	15	13	15	13
15	12	12	14	15	15	12	15	15	13	13	14	14	13	13
12	14	15	12	15	12	13	15	13	12	12	15	14	15	13
15	15	14	12	14	13	13	14							

La media o promedio de cuanto estarían dispuestos a pagar por dona resulto en 13.50 pesos.

Muy probablemente se preguntaría que tanto potencial tiene la venta de donas de chocolate en el Instituto Tecnológico de Cd. Cuauhtémoc. La respuesta a ello, sería.

Cantidad de personas que si comprarían donas $2650 * .7690 = 2037.9$, redondeado sería 2040.

Considerando que la frecuencia estimada de compra a la semana es de 2 donas, se tiene que la cantidad de donas compradas a la semana sería de $2040 * 2 = 4080$ donas por semana.

Asumiendo que el precio de venta se fijaría en 13 pesos (por cuestiones de facilitar el cobro), la magnitud de ventas por semana sería de $4080 * 13 = 53040$ pesos.

Recuerde que lo anterior se refiere a la potencialidad del mercado. Probablemente alguien diría, con que venda 200 donas por día, es suficiente.

Ejercicios propuestos.

Ejercicio 1. Considere que desea establecer que proporción de latas de chiles en escabeche se encuentran dañadas de un lote de 1500 piezas. Establezca el tamaño de muestra necesario para ello, considerando un nivel de confianza del 90 % y un error de estimación del 4 %.

Invente los datos que sean necesarios para cumplir el objetivo de establecer la proporción de latas defectuosas

Ejercicio 2. Invente los datos que sean necesarios para definir el tamaño de muestra y realice los cálculos para establecer el peso promedio de las manzanas de una de las huertas de la Norteña de Cd. Cuauhtémoc.

Ejercicio 3. Usted desea vender conos de jamoncillo en las primarias de Cd. Cuauhtémoc. Considere que la cantidad de niños en dichas primarias es del orden de 8750 e invente lo necesario para definir la magnitud de las ventas potenciales por semana.

Ejercicio 4. Realice el procedimiento requerido para establecer que proporción de los alumnos del Instituto Tecnológico de Cd. Cuauhtémoc tiene al menos un tatuaje.

4.4. Distribuciones muestrales.

La industria alimentaria, un sector de vital importancia para la sociedad, se sustenta en la calidad y seguridad de sus productos. Para garantizar estos estándares, las empresas emplean una amplia gama de herramientas estadísticas, siendo las distribuciones muestrales una de las más relevantes. Estas distribuciones, que describen la probabilidad de obtener ciertos resultados al seleccionar muestras aleatorias de una población, desempeñan un papel crucial en diversos aspectos del control de calidad y la toma de decisiones en este sector.

Control de Calidad: Más allá de los límites

Una de las aplicaciones más comunes de las distribuciones muestrales en la industria alimentaria es el control de calidad. Por ejemplo, en una fábrica de galletas, se establecen límites de control estadísticos para el peso de cada galleta. Al tomar muestras aleatorias de la línea de producción y calcular el peso promedio y la desviación estándar, se puede construir un intervalo de confianza. Si el peso promedio de la muestra cae fuera de este intervalo, se activa una alarma, indicando que el proceso de producción podría estar fuera de control y que es necesario realizar ajustes.

Diseño de Experimentos: Optimizando recetas

Imagine a empresa que desea desarrollar una nueva salsa de tomate. Los ingenieros de alimentos diseñan un experimento variando el tipo de tomate, la cantidad de azúcar y el tiempo de cocción. Al analizar los resultados utilizando un diseño experimental y técnicas de análisis de varianza, pueden determinar qué combinación de factores produce la salsa con el sabor más agradable para los consumidores. Las distribuciones muestrales permiten evaluar la significancia estadística de las diferencias entre los tratamientos y construir modelos predictivos para optimizar la receta.

Evaluación de la Vida Útil: Más allá de la fecha de caducidad

Las distribuciones muestrales también son fundamentales para estimar la vida útil de los alimentos. Por ejemplo, una empresa de yogur desea determinar cuánto tiempo puede almacenarse su producto a una temperatura de refrigeración específica antes de que se produzcan cambios significativos en su textura o sabor. Al someter muestras de yogur a diferentes tiempos de almacenamiento y medir parámetros como la acidez y la viscosidad, se pueden ajustar modelos de regresión que permitan predecir la vida útil con un cierto nivel de confianza.

Otros Ejemplos:

- **Análisis sensorial:** Las distribuciones muestrales se utilizan para evaluar las preferencias de los consumidores y comparar diferentes productos. Por ejemplo, al realizar una prueba de cata a ciegas, se puede utilizar una distribución normal para modelar las puntuaciones otorgadas por los consumidores a diferentes marcas de un mismo producto.
- **Control de contaminantes:** Se emplean para determinar si la concentración de un contaminante en un alimento supera los límites establecidos por las regulaciones sanitarias. Por ejemplo, en el análisis de muestras de agua para detectar la presencia de E. coli, se utiliza una distribución binomial para calcular la probabilidad de encontrar un número determinado de bacterias en una muestra.
- **Optimización de procesos:** Se utilizan para identificar las variables que más influyen en la calidad de un producto y mejorar la eficiencia de los procesos de producción. Por ejemplo, en la producción de cerveza, se puede utilizar un diseño factorial para evaluar el efecto de diferentes variables, como la temperatura de fermentación y el tipo de levadura, en el sabor final del producto.

En conclusión, se tiene que las distribuciones muestrales son una herramienta estadística indispensable en la industria alimentaria. Su aplicación en diversas áreas, desde el control de calidad hasta el desarrollo de nuevos productos, permite garantizar la seguridad y calidad de los alimentos, optimizar los procesos de producción y tomar decisiones basadas en evidencia. Al comprender los principios básicos de las distribuciones muestrales, los profesionales de la industria alimentaria pueden tomar decisiones más informadas y mejorar la competitividad de sus empresas.

Principios Básicos de las Distribuciones Muestrales

Las distribuciones muestrales son un concepto fundamental en estadística inferencial que permite hacer inferencias sobre una población a partir de una muestra. En esencia, una distribución muestral describe la probabilidad de obtener diferentes valores de un estadístico (como la media o la proporción) al tomar múltiples muestras de una misma población.

¿Por qué son importantes las distribuciones muestrales?

- Inferencia estadística: Nos permiten hacer generalizaciones sobre una población completa a partir de una pequeña muestra.
- Construcción de intervalos de confianza: Ayudan a determinar un rango de valores en el que probablemente se encuentre el parámetro poblacional.
- Pruebas de hipótesis: Nos permiten tomar decisiones sobre afirmaciones acerca de una población.

Conceptos Clave

- Población: El conjunto completo de todos los elementos de interés en un estudio.
- Muestra: Un subconjunto de la población seleccionado para el estudio.
- Estadístico: Una medida calculada a partir de los datos de una muestra (por ejemplo, media, desviación estándar, proporción).
- Distribución muestral: La distribución de probabilidad de un estadístico cuando se toman múltiples muestras de la misma población.

Teorema de limite central (TLC).

El teorema de limite central establece que una distribución muestral de medias sigue una distribución normal en los siguientes dos casos.

- 1.- La distribución de la media de múltiples muestras aleatorias tomadas de una población se aproxima a una distribución normal, independientemente de la distribución de la población original, siempre y cuando el tamaño de las muestras sea suficientemente grande. Generalmente, se considera que una muestra de 30 observaciones es suficiente para que esto suceda.
- 2.- Si una población tiene una distribución normal, la distribución muestral de las medias también será normal, independientemente del tamaño de la muestra.

4.5. Estimación.

Estimación en la Industria Alimentaria: Un Pilar para la Toma de Decisiones

La industria alimentaria, en su constante búsqueda de eficiencia y optimización, depende en gran medida de la estimación. Desde la predicción de la demanda hasta la evaluación de la calidad de los productos, las técnicas de estimación juegan un papel crucial en la toma de decisiones estratégicas y operativas. La importancia de la estimación en la industria alimentaria, destacando su aplicación en diversas áreas y su impacto en la rentabilidad y la sostenibilidad, se presenta a continuación.

Estimación de la Demanda: Anticipando las Necesidades del Mercado

La estimación de la demanda es una de las aplicaciones más fundamentales de la estimación en la industria alimentaria. Al predecir la cantidad de producto que los consumidores demandarán en un período determinado, las empresas pueden ajustar su producción, gestionar el inventario y planificar estrategias de marketing efectivas. Una estimación precisa de la demanda permite evitar el exceso de inventario, que puede resultar en pérdidas económicas, y la escasez de producto, que puede dañar la reputación de la marca y la lealtad del cliente (Hyndman & Athanasopoulos, 2018).

Existen diversos métodos para estimar la demanda, desde modelos estadísticos basados en datos históricos hasta técnicas de investigación de mercado que involucran encuestas y análisis de tendencias. La elección del método adecuado dependerá de la disponibilidad de datos, la complejidad del mercado y los recursos disponibles.

Estimación de la Calidad del Producto: Garantizando la Satisfacción del Cliente

Además de la demanda, la estimación también se utiliza para evaluar y garantizar la calidad de los productos alimentarios. Las empresas pueden utilizar técnicas de muestreo y análisis estadístico para estimar la calidad de un lote de producción, identificando posibles defectos o desviaciones de los estándares de calidad. Esta información es esencial para tomar decisiones sobre la aceptación o el rechazo de un lote, así como para implementar medidas correctivas en el proceso de producción (Montgomery, Runger, & Hubele, 2010).

La estimación de la calidad del producto no solo es importante para garantizar la satisfacción del cliente, sino también para cumplir con las regulaciones y normativas de seguridad alimentaria. Al implementar sistemas de control de calidad basados en técnicas de estimación, las empresas pueden reducir el riesgo de brotes de enfermedades transmitidas por los alimentos y proteger la salud de los consumidores.

Estimación de Costos y Rentabilidad: Optimizando la Producción

La estimación también juega un papel fundamental en la gestión de costos y la evaluación de la rentabilidad en la industria alimentaria. Al estimar los costos de producción, las empresas pueden fijar precios competitivos, evaluar la viabilidad de nuevos productos y tomar decisiones sobre la inversión en equipos y tecnologías. La estimación de la rentabilidad, por su parte, permite a las empresas identificar los productos o líneas de producción más rentables, así como los factores que influyen en la rentabilidad, como los costos de producción, los precios de venta y la demanda del mercado.

En resumen, la estimación es una herramienta esencial en la industria alimentaria, que permite a las empresas tomar decisiones informadas en diversas áreas, desde la planificación de la producción hasta la gestión de la calidad y la evaluación de la rentabilidad. Al utilizar técnicas de estimación adecuadas, las empresas pueden optimizar sus operaciones, reducir costos, mejorar la calidad de sus productos y garantizar la satisfacción del cliente. En un

mercado cada vez más competitivo y exigente, la capacidad de estimar con precisión es un factor clave para el éxito y la sostenibilidad en la industria alimentaria.

Estimación Puntual vs. Estimación de Intervalos: Dos Enfoques para la Inferencia Estadística

En el campo de la inferencia estadística, el objetivo principal es obtener información sobre una población a partir de una muestra de datos. Dos enfoques comunes para lograr esto son la estimación puntual y la estimación de intervalos. Si bien ambos métodos buscan proporcionar información sobre un parámetro poblacional desconocido, difieren en su enfoque y en la cantidad de información que proporcionan. Este ensayo explorará las características de cada enfoque, destacando sus ventajas y desventajas, y proporcionando ejemplos para ilustrar sus diferencias.

Estimación Puntual: Un Valor Único como Estimación

La estimación puntual implica calcular un único valor que se considera la mejor estimación del parámetro poblacional desconocido. Este valor se obtiene a partir de los datos de la muestra y se denomina estimador puntual. Por ejemplo, si queremos estimar la altura promedio de todos los estudiantes de una universidad, podríamos calcular la altura promedio de una muestra de estudiantes y utilizar este valor como nuestra estimación puntual.

La principal ventaja de la estimación puntual es su simplicidad. Proporciona un valor único y fácil de interpretar que puede utilizarse para tomar decisiones o hacer predicciones. Sin embargo, la estimación puntual también tiene una limitación importante: no proporciona información sobre la precisión o la incertidumbre asociada a la estimación. En otras palabras, no nos dice qué tan cerca está nuestra estimación del verdadero valor del parámetro poblacional.

Estimación de Intervalos

A diferencia de la estimación puntual, la estimación de intervalos implica calcular un rango de valores dentro del cual se espera que se encuentre el parámetro poblacional desconocido con un cierto nivel de confianza. Este rango se denomina intervalo de confianza y se construye utilizando los datos de la muestra y un nivel de confianza predefinido. Por ejemplo, se puede estimar que la altura promedio de los estudiantes de la universidad se encuentra entre 165 cm y 175 cm, con un nivel de confianza del 95%.

La principal ventaja de la estimación de intervalos es que proporciona información sobre la precisión de la estimación. El ancho del intervalo de confianza indica la incertidumbre asociada a la estimación: cuanto más ancho es el intervalo, más incierta es la estimación. Además, el nivel de confianza dice con qué frecuencia se espera que el intervalo de confianza contenga el verdadero valor del parámetro poblacional si se repitiera el proceso de muestreo muchas veces.

Para ilustrar las diferencias entre la estimación puntual y la estimación de intervalos, considere el siguiente ejemplo. Suponga que se quiere estimar el peso promedio de los paquetes de cereal que se producen en una fábrica. Si se toma una muestra de 100 paquetes y se calcula el peso promedio de la muestra, que resulta ser 500 gramos. Esta sería la estimación puntual.

Si se utiliza la estimación de intervalos, se puede calcular un intervalo de confianza del 95% y obtener un resultado como este: el peso promedio de los paquetes de cereal se encuentra entre 495 gramos y 505 gramos. Este intervalo de confianza dice que se está un 95% seguros de que el verdadero peso promedio de los paquetes de cereal se encuentra dentro de este rango.

En resumen, la estimación puntual y la estimación de intervalos son dos enfoques complementarios para la inferencia estadística. La estimación puntual proporciona un valor único como estimación del parámetro poblacional, pero no proporciona información sobre la precisión de la estimación. La estimación de intervalos, por otro lado, proporciona un rango de valores posibles y un nivel de confianza asociado, lo que permite evaluar la precisión de la estimación. La elección entre un enfoque u otro dependerá del contexto específico y de la información que se requiera. En general, la estimación de intervalos es preferible cuando se desea tener una idea de la incertidumbre asociada a la estimación.

Caso 1. Estimación de intervalo para la media con desviación estándar poblacional conocida.

La fórmula a utilizar para este caso es.

$$\mu = \bar{x} \pm z \frac{\sigma}{\sqrt{n}}$$

Donde:

$\bar{x}$ es la media muestral.

z es la cantidad de desviaciones estándar de un punto a la media y esta en función del nivel de confianza.

σ es la desviación estándar poblacional.

n es el tamaño de muestra.

Ejemplo 1. Considere que el peso de las manzanas se distribuye de forma normal con desviación estándar de 12.5 gramos. Se tomo una muestra de 10 manzanas, las cuales pesaron en gramos: 213, 206, 191, 200, 193, 196, 203, 201, 197, 198.

Calcule un intervalo para la media con un nivel de confianza del:

a) 85 %.

b)90 %.

c)95 %.

Solución.

a)

Media muestral	$\bar{x}$	199.8
Desviación estándar	σ	12.5
Nivel de confianza	NC	0.85
Valor tipificado	z	1.44
Tamaño de muestra	n	10
Lado derecho	LD	5.69

$$\mu = \bar{x} \pm z \frac{\sigma}{\sqrt{n}} \text{ Lado derecho (LD)}$$

$$\mu = 199.8 \pm 1.44 * \frac{12.5}{\sqrt{10}}$$

$$\mu = 199.8 \pm 5.69$$

$$194.11 \leq \mu \leq 205.49$$

A un nivel de confianza del 85 % se tiene que el peso promedio de las manzanas oscila de 194.11 a 205.49 gramos

b)

Media muestral	$\bar{x}$	199.8
Desviación estándar	σ	12.5
Nivel de confianza	NC	0.9
Valor tipificado	z	1.64
Tamaño de muestra	n	10
Lado derecho	LD	6.50

$$\mu = \bar{x} \pm z \frac{\sigma}{\sqrt{n}} \text{ Lado derecho (LD)}$$

$$\mu = 199.8 \pm 1.64 * \frac{12.5}{\sqrt{10}}$$

$$\mu = 199.8 \pm 6.5$$

$$193.30 \leq \mu \leq 206.30$$

A un nivel de confianza del 90 % se tiene que el peso promedio de las manzanas oscila de 193.30 a 206.30 gramos

c)

Media muestral	$\bar{x}$	199.8
Desviación estándar	σ	12.5
Nivel de confianza	NC	0.95
Valor tipificado	z	1.96
Tamaño de muestra	n	10
Lado derecho	LD	7.75

$$\mu = \bar{x} \pm z \frac{\sigma}{\sqrt{n}} \text{ Lado derecho (LD)}$$

$$\mu = 199.8 \pm 1.96 * \frac{12.5}{\sqrt{10}}$$

$$\mu = 199.8 \pm 7.75$$

$$192.05 \leq \mu \leq 207.55$$

A un nivel de confianza del 95 % se tiene que el peso promedio de las manzanas oscila de 192.05 a 207.55 gramos

Ver video de solución dando clic en el siguiente enlace <https://youtu.be/0h4sABdV1Gk>

Ejemplo 2. En el laboratorio de control de calidad de miel Norteña se realizaron las pruebas de densidad de la miel de abeja que se cosecho de diferentes localidades, obteniéndose los

siguientes resultados en kilogramos por litro: 1.4, 1.44, 1.39, 1.4, 1.45, 1.39, 1.43, 1.42, 1.44, 1.44, 1.4, 1.38, 1.42, 1.41, 1.43, 1.41, 1.4, 1.41, 1.41, 1.45, 1.45, 1.44, 1.41, 1.41, 1.38, 1.45, 1.43, 1.42, 1.38, 1.41, 1.43, 1.43, 1.4

Calcule un intervalo para la media con un nivel de confianza de:

- a) 90 %
- b) 94 %
- c) 96 %

Solución.

En este ejercicio aun cuando no se proporciona la desviación estándar poblacional y no se dice que la densidad de la miel de abeja se distribuye de forma normal, si tiene una distribución normal dado que el tamaño de muestra es mayor o igual a 30, tal como lo establece el teorema de limite central, por lo tanto, si corresponde a este caso 1.

Adicional a lo anterior, se tiene que al no contar con la desviación estándar poblacional, se usa la desviación estándar muestral como estimador de ella.

a)

Media muestral	$\bar{x}$	1.417
Desviación estándar	σ	0.0214
Nivel de confianza	NC	0.9
Valor tipificado	z	1.64
Tamaño de muestra	n	33
Lado derecho	LD	0.006

$$\mu = \bar{x} \pm z \frac{\sigma}{\sqrt{n}} \text{ Lado derecho (LD)}$$

$$\mu = 1.417 \pm 1.64 * \frac{.0214}{\sqrt{33}}$$

$$\mu = 1.417 \pm 0.006$$

1.411	$\leq \mu \leq$	1.423
--------------	-----------------	--------------

A un nivel de confianza de 90 % se tiene que la densidad promedio de la miel de abeja oscila de 1.411 a 1.423 kilogramos por litro.

b)

Media muestral	$\bar{x}$	1.417
Desviación estándar	σ	0.0214
Nivel de confianza	NC	0.94
Valor tipificado	z	1.88
Tamaño de muestra	n	33
Lado derecho	LD	0.007

$$\mu = \bar{x} \pm z \frac{\sigma}{\sqrt{n}} \text{ Lado derecho (LD)}$$

$$\mu = 1.417 \pm 1.88 * \frac{.0214}{\sqrt{33}}$$

$$\mu = 1.417 \pm 0.007$$

1.410	$\leq \mu \leq$	1.424
--------------	-----------------	--------------

A un nivel de confianza de 94 % se tiene que la densidad promedio de la miel de abeja oscila de 1.410 a 1.424 kilogramos por litro

c)

Media muestral	$\bar{x}$	1.417
Desviación estándar	σ	0.0214
Nivel de confianza	NC	0.96
Valor tipificado	z	2.05
Tamaño de muestra	n	33
Lado derecho	LD	0.008

$$\mu = \bar{x} \pm z \frac{\sigma}{\sqrt{n}} \text{ Lado derecho (LD)}$$

$$\mu = 1.417 \pm 2.05 * \frac{.0214}{\sqrt{33}}$$

$$\mu = 1.417 \pm 0.008$$

$$1.409 \leq \mu \leq 1.425$$

A un nivel de confianza de 96 % se tiene que la densidad promedio de la miel de abeja oscila de 1.409 a 1.425 kilogramos por litro

Ejercicios propuestos.

Ejercicio 1: Control de calidad en la producción de leche

Una empresa láctea produce leche envasada en botellas de un litro. Se sabe por experiencia que la desviación estándar de la cantidad de leche en cada botella es de 5 ml y que el volumen se distribuye de forma normal. Se toma una muestra aleatoria de 25 botellas y se mide la cantidad de leche en cada una, obteniendo una media muestral de 998 ml.

a) Estima el intervalo de confianza del 90% para la cantidad promedio de leche en todas las botellas producidas por la empresa. Respuesta $996.36 \leq \mu \leq 999.65$

b) Estima el intervalo de confianza del 95% para la cantidad promedio de leche en todas las botellas producidas por la empresa. Respuesta $996.04 \leq \mu \leq 999.96$

Ejercicio 2: Peso promedio de galletas

Una fábrica de galletas produce paquetes de galletas con un peso declarado de 200 gramos. Se sabe que la desviación estándar del peso de los paquetes es de 3 gramos y que el peso de las galletas se distribuye de forma normal. Se toma una muestra aleatoria de 12 paquetes y se pesa cada uno, obteniendo los siguientes datos en gramos: 199.1, 200.4, 201.2, 200.5, 200.3, 198.2, 198.8, 198.1, 201.7, 199.5, 198.2, 198.6

a) Estima el intervalo de confianza del 85% para el peso promedio de todos los paquetes de galletas producidos por la fábrica. Respuesta $198.30 \leq \mu \leq 200.79$

b) Estima el intervalo de confianza del 94% para el peso promedio de todos los paquetes de galletas producidos por la fábrica. Respuesta $197.92 \leq \mu \leq 201.18$

Ejercicio 3: Contenido de azúcar en refrescos

Una empresa de bebidas produce refrescos de cola. Se sabe que la desviación estándar del contenido de azúcar en cada botella es de 2 gramos. Se toma una muestra aleatoria de 36

botellas y se mide el contenido de azúcar en cada una, obteniendo una media muestral de 38 gramos.

a) Estima el intervalo de confianza del 90% para el contenido promedio de azúcar en todas las botellas de refresco producidas por la empresa. Respuesta $37.45 \leq \mu \leq 38.55$

b) Estima el intervalo de confianza del 94% para el contenido promedio de azúcar en todas las botellas de refresco producidas por la empresa. Respuesta $37.37 \leq \mu \leq 38.63$

Ejercicio 4: Tiempo de cocción de alimentos preparados

Una empresa de alimentos preparados produce platos listos para consumir. Se sabe que la desviación estándar del tiempo de cocción de estos platos es de 5 minutos. Se toma una muestra aleatoria de 49 platos y se mide el tiempo de cocción de cada uno, obteniendo una media muestral de 25 minutos.

a) Estima el intervalo de confianza del 88% para el tiempo de cocción promedio de todos los platos producidos por la empresa. Respuesta $23.89 \leq \mu \leq 26.11$

b) Estima el intervalo de confianza del 92% para el tiempo de cocción promedio de todos los platos producidos por la empresa. Respuesta $23.75 \leq \mu \leq 26.25$

Ejercicio 5: Longitud de salchichas

Una empresa de embutidos produce salchichas. Se sabe que la desviación estándar de la longitud de las salchichas es de 1 cm. Se toma una muestra aleatoria de 100 salchichas y se mide la longitud de cada una, obteniendo una media muestral de 15 cm.

a) Estima el intervalo de confianza del 94% para la longitud promedio de todas las salchichas producidas por la empresa. Respuesta $14.81 \leq \mu \leq 15.19$

b) Estima el intervalo de confianza del 99% para la longitud promedio de todas las salchichas producidas por la empresa. Respuesta $14.74 \leq \mu \leq 15.26$

Caso 2. Estimación de intervalo para la media con desviación estándar poblacional desconocida.

Este caso corresponde a las ocasiones en que el tamaño de muestra es menor a 30, no se conoce la desviación estándar poblacional y la variable se distribuye de forma aproximadamente normal.

La fórmula a utilizar para es.

$$\mu = \bar{x} \pm t \frac{s}{\sqrt{n}}$$

Donde:

$\bar{x}$ es la media muestral.

t es valor t de student.

s es la desviación estándar muestral.

n es el tamaño de muestra.

Para este ejercicio es necesario consultar la tabla de t de student, la cual se presenta a continuación.

Tabla A.4 Valores críticos de la distribución t

737

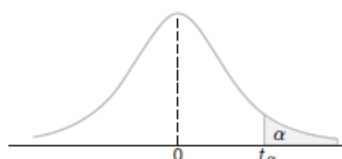


Tabla A.4 Valores críticos de la distribución t

ν	α						
	0.40	0.30	0.20	0.15	0.10	0.05	0.025
1	0.325	0.727	1.376	1.963	3.078	6.314	12.706
2	0.289	0.617	1.061	1.386	1.886	2.920	4.303
3	0.277	0.584	0.978	1.250	1.638	2.353	3.182
4	0.271	0.569	0.941	1.190	1.533	2.132	2.776
5	0.267	0.559	0.920	1.156	1.476	2.015	2.571
6	0.265	0.553	0.906	1.134	1.440	1.943	2.447
7	0.263	0.549	0.896	1.119	1.415	1.895	2.365
8	0.262	0.546	0.889	1.108	1.397	1.860	2.306
9	0.261	0.543	0.883	1.100	1.383	1.833	2.262
10	0.260	0.542	0.879	1.093	1.372	1.812	2.228
11	0.260	0.540	0.876	1.088	1.363	1.796	2.201
12	0.259	0.539	0.873	1.083	1.356	1.782	2.179
13	0.259	0.538	0.870	1.079	1.350	1.771	2.160
14	0.258	0.537	0.868	1.076	1.345	1.761	2.145
15	0.258	0.536	0.866	1.074	1.341	1.753	2.131
16	0.258	0.535	0.865	1.071	1.337	1.746	2.120
17	0.257	0.534	0.863	1.069	1.333	1.740	2.110
18	0.257	0.534	0.862	1.067	1.330	1.734	2.101
19	0.257	0.533	0.861	1.066	1.328	1.729	2.093
20	0.257	0.533	0.860	1.064	1.325	1.725	2.086
21	0.257	0.532	0.859	1.063	1.323	1.721	2.080
22	0.256	0.532	0.858	1.061	1.321	1.717	2.074
23	0.256	0.532	0.858	1.060	1.319	1.714	2.069
24	0.256	0.531	0.857	1.059	1.318	1.711	2.064
25	0.256	0.531	0.856	1.058	1.316	1.708	2.060
26	0.256	0.531	0.856	1.058	1.315	1.706	2.056
27	0.256	0.531	0.855	1.057	1.314	1.703	2.052
28	0.256	0.530	0.855	1.056	1.313	1.701	2.048
29	0.256	0.530	0.854	1.055	1.311	1.699	2.045
30	0.256	0.530	0.854	1.055	1.310	1.697	2.042
40	0.255	0.529	0.851	1.050	1.303	1.684	2.021
60	0.254	0.527	0.848	1.045	1.296	1.671	2.000
120	0.254	0.526	0.845	1.041	1.289	1.658	1.980
∞	0.253	0.524	0.842	1.036	1.282	1.645	1.960

Para consultarla se requiere el nivel de significancia α y los grados de libertad.

El nivel de significancia es igual a 1 menos en nivel de confianza.

Los grados de libertad son igual a número de datos menos 1.

Cuál es el valor de t de student para un nivel de confianza del 95% para una muestra de tamaño 12.

Como el nivel de confianza es 95%, el nivel de significancia es 5%. Adicional a lo anterior se tiene que el nivel de significancia se distribuye por igual entre los dos extremos de la distribución de probabilidad, por lo que el valor a consultar en la tabla será $.05 / 2 = 0.025$.

Los grados de libertad (gl) son 11 (12 – 1).

Consultando estos datos en la tabla t de student, se tiene que el valor correspondiente es 2.201, tal como se aprecia a continuación.

Tabla A.4 Valores críticos de la distribución t

737

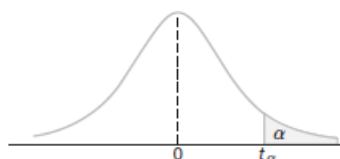
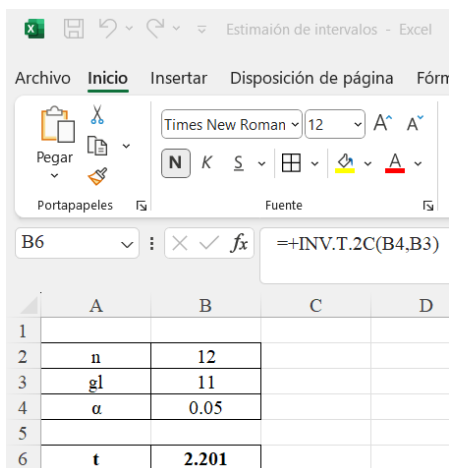


Tabla A.4 Valores críticos de la distribución t

v	α						
	0.40	0.30	0.20	0.15	0.10	0.05	0.025
1	0.325	0.727	1.376	1.963	3.078	6.314	12.706
2	0.289	0.617	1.061	1.386	1.886	2.920	4.303
3	0.277	0.584	0.978	1.250	1.638	2.353	3.182
4	0.271	0.569	0.941	1.190	1.533	2.132	2.776
5	0.267	0.559	0.920	1.156	1.476	2.015	2.571
6	0.265	0.553	0.906	1.134	1.440	1.943	2.447
7	0.263	0.549	0.896	1.119	1.415	1.895	2.365
8	0.262	0.546	0.889	1.108	1.397	1.860	2.306
9	0.261	0.543	0.883	1.100	1.383	1.833	2.262
10	0.260	0.542	0.879	1.093	1.372	1.812	2.228
11	0.260	0.540	0.876	1.088	1.363	1.796	2.201
12	0.259	0.539	0.873	1.083	1.356	1.782	2.179
13	0.259	0.538	0.870	1.079	1.350	1.771	2.160
14	0.258	0.537	0.868	1.076	1.345	1.761	2.145
15	0.258	0.536	0.866	1.074	1.341	1.753	2.131
16	0.258	0.535	0.865	1.071	1.337	1.746	2.120
17	0.257	0.534	0.863	1.069	1.333	1.740	2.110
18	0.257	0.534	0.862	1.067	1.330	1.734	2.101
19	0.257	0.533	0.861	1.066	1.328	1.729	2.093
20	0.257	0.533	0.860	1.064	1.325	1.725	2.086
21	0.257	0.532	0.859	1.063	1.323	1.721	2.080
22	0.256	0.532	0.858	1.061	1.321	1.717	2.074
23	0.256	0.532	0.858	1.060	1.319	1.714	2.069
24	0.256	0.531	0.857	1.059	1.318	1.711	2.064
25	0.256	0.531	0.856	1.058	1.316	1.708	2.060
26	0.256	0.531	0.856	1.058	1.315	1.706	2.056
27	0.256	0.531	0.855	1.057	1.314	1.703	2.052
28	0.256	0.530	0.855	1.056	1.313	1.701	2.048
29	0.256	0.530	0.854	1.055	1.311	1.699	2.045
30	0.256	0.530	0.854	1.055	1.310	1.697	2.042
40	0.255	0.529	0.851	1.050	1.303	1.684	2.021
60	0.254	0.527	0.848	1.045	1.296	1.671	2.000
120	0.254	0.526	0.845	1.041	1.289	1.658	1.980
∞	0.253	0.524	0.842	1.036	1.282	1.645	1.960

Este valor se puede obtener en Excel mediante la función INV.T.2C, tal como se observa en la siguiente imagen.



Ejemplo 1. Una fábrica de conservas produce latas de atún en aceite. Se desea estimar el peso promedio de las latas, por lo que se toma una muestra aleatoria de 15 latas y se pesa cada una, obteniendo una media muestral de 120 gramos y una desviación estándar muestral de 8 gramos.

- Estima el intervalo de confianza del 90% para el peso promedio de todas las latas de atún producidas por la fábrica.
- Estima el intervalo de confianza del 95% para el peso promedio de todas las latas de atún producidas por la fábrica

Solución

a)

Media muestral	$\bar{x}$	120
Desviación estándar	s	8
Nivel de confianza	NC	0.9
Nivel de significancia	α	0.1
Tamaño de muestra	n	15
Grados de libertad	gl	14
Valor t de student	t	1.761
Lado derecho	LD	3.64

$$\mu = \bar{x} \pm t \frac{s}{\sqrt{n}}$$

$$\mu = 120 \pm 1.761 * \frac{8}{\sqrt{15}}$$

$$\mu = 120 \pm 3.64$$

Lado derecho (LD)

116.36

$\leq \mu \leq$

123.64

A un nivel de confianza del 90 % se tiene que el peso promedio de las latas de atún oscila de 116.36 a 123.64 gramos

[illegible]

Ver video de solución dando clic en el siguiente link <https://youtu.be/QxaAP4F10aw>

a) Estima el intervalo de confianza del 90% para el contenido promedio de azúcar en todo el zumo de naranja producido por la empresa.

Solución.

A un nivel de confianza del 90 % se tiene que el contenido promedio de azúcar en el zumo de naranja producido por la empresa oscila de 9.49 a 10.51 gramos por cada 100 mililitros.

b)

Media muestral	$\bar{x}$	10
Desviación estándar	s	1.5
Nivel de confianza	NC	0.99
Nivel de significancia	α	0.01
Tamaño de muestra	n	25
Grados de libertad	gl	24
Valor t de student	t	2.797
Lado derecho	LD	0.84

$$\mu = \bar{x} \pm t \frac{s}{\sqrt{n}} \quad \text{Lado derecho (LD)}$$

$$\mu = 10 \pm 2.797 * \frac{1.5}{\sqrt{25}}$$

$$\mu = 120 \pm 0.84$$

9.16	$\leq \mu \leq$	10.84
-------------	-----------------	--------------

A un nivel de confianza del 90 % se tiene que el contenido promedio de azúcar en el zumo de naranja producido por la empresa oscila de 9.16 a 10.84 gramos por cada 100 mililitros

Ejercicios propuestos.

Ejercicio 1. Tiempo de fermentación de cerveza

Una cervecería artesanal desea estimar el tiempo promedio de fermentación de su cerveza tipo Ale. Se toma una muestra aleatoria de 12 lotes de cerveza y se mide el tiempo de fermentación de cada uno, obteniendo una media muestral de 14 días y una desviación estándar muestral de 2 días.

a) Estima el intervalo de confianza del 95% para el tiempo promedio de fermentación de toda la cerveza Ale producida por la cervecería. Respuesta $12.73 \leq \mu \leq 15.27$

b) Estima el intervalo de confianza del 98% para el tiempo promedio de fermentación de toda la cerveza Ale producida por la cervecería. Respuesta $12.43 \leq \mu \leq 15.57$

Ejercicio 2. Nivel de pH en salsa de tomate

Una empresa de alimentos produce salsa de tomate. Se desea estimar el nivel promedio de pH en su producción. Se toma una muestra aleatoria de 18 lotes de salsa de tomate y se mide el nivel de pH en cada uno, obteniendo una media muestral de 4.2 y una desviación estándar muestral de 0.2.

a) Estima el intervalo de confianza del 90% para el nivel promedio de pH en toda la salsa de tomate producida por la empresa. Respuesta $4.118 \leq \mu \leq 4.282$

b) Estima el intervalo de confianza del 96% para el nivel promedio de pH en toda la salsa de tomate producida por la empresa. Respuesta $4.0952 \leq \mu \leq 4.3048$

Ejercicio 3. Contenido de grasa en carne picada

Un productor de carne desea estimar el porcentaje promedio de grasa en su carne picada. Se toma una muestra aleatoria de 16 paquetes de carne picada y se mide el contenido de grasa

en cada uno, obteniendo los siguientes datos en porcentaje: 15.2, 14.5, 15.4, 14.9, 16.1, 13.9, 14.8, 15.4, 15.9, 16.3, 14.8, 13.9, 14.8, 13.8, 14.8, 15.6.

a) Estima el intervalo de confianza del 89% para el porcentaje promedio de grasa en toda la carne picada producida por el productor. Respuesta $14.6824 \leq \mu \leq 15.3296$

b) Estima el intervalo de confianza del 95% para el porcentaje promedio de grasa en toda la carne picada producida por el productor. Respuesta $14.600 \leq \mu \leq 15.412$

Ejercicio 4. Peso de barras de chocolate

Una fábrica de chocolate produce barras de chocolate de 100 gramos. Se desea estimar el peso promedio real de las barras. Se toma una muestra aleatoria de 20 barras y se pesan, obteniendo una media muestral de 98 gramos y una desviación estándar muestral de 3 gramos.

a) Estima el intervalo de confianza del 90% para el peso promedio de todas las barras de chocolate producidas por la fábrica. Respuesta $96.8491 \leq \mu \leq 99.1599$

b) Estima el intervalo de confianza del 94% para el peso promedio de todas las barras de chocolate producidas por la fábrica. Respuesta $96.6583 \leq \mu \leq 99.3417$

Ejercicio 85. Acidez en salsa de tomate

Una empresa de alimentos desea estimar el nivel promedio de acidez en su salsa de tomate. Se toma una muestra aleatoria de 10 lotes de salsa de tomate y se mide el nivel de acidez en cada uno, obteniendo los siguientes datos: 4.45, 4.39, 4.58, 4.51, 4.36, 4.85, 4.56, 4.66, 4.54, 4.49.

a) Estima el intervalo de confianza del 89% para el nivel promedio de acidez en toda la salsa de tomate producida por la empresa. Respuesta $4.4599 \leq \mu \leq 4.6181$

b) Estima el intervalo de confianza del 97% para el nivel promedio de acidez en toda la salsa de tomate producida por la empresa. Respuesta $4.4242 \leq \mu \leq 4.6538$

Caso 3. Estimación de intervalo para la diferencia de medias con desviaciones estándar poblacional conocidas.

La fórmula correspondiente a este caso es.

$$\mu_1 - \mu_2 = (\bar{x}_1 - \bar{x}_2) \pm z \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

En esta fórmula se puede apreciar que son dos poblaciones y que de cada una de ellas se tiene información en cuanto a su media muestral, su desviación estándar poblacional y su tamaño de muestra. Adicional a lo anterior se observa el valor de z el cual está en función del nivel de confianza e indica que la diferencia de medias se distribuye de forma normal.

Ejemplo 1. Una empresa de lácteos quiere estimar la diferencia en la cantidad promedio de grasa entre dos marcas de leche entera. Se toman muestras aleatorias de 60 botellas de la marca A y 70 botellas de la marca B. Las desviaciones estándar poblacionales son de 0.6 gramos para la marca A y 0.8 gramos para la marca B.

a) Si la media muestral de grasa para la marca A es de 3.5 gramos y para la marca B es de 3.2 gramos, calcule un intervalo de confianza del 90% para la diferencia de medias poblacionales.

b) Si la media muestral de grasa para la marca A es de 3.5 gramos y para la marca B es de 3.2 gramos, calcule un intervalo de confianza del 95% para la diferencia de medias poblacionales.

Solución.

a)

		Marca	
		A	B
Media muestral	$\bar{x}$	3.5	3.2
Desviación estándar	σ	0.6	0.8
Varianza	σ^2	0.36	0.64
Tamaño de muestra	n	60	70

$$\mu_1 - \mu_2 = (\bar{x}_1 - \bar{x}_2) \pm z \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Lado derecho (LD)

Diferencia de medias	$\bar{x}_1 - \bar{x}_2$	0.3
Nivel de confianza	NC	0.9
Valor tipificado	z	1.64
Lado derecho	LD	0.2024

$$\bar{\mu}_1 - \mu_2 = 0.3 \pm 1.64 \sqrt{\frac{0.36}{60} + \frac{0.64}{70}}$$

$$\bar{\mu}_1 - \mu_2 = 0.3 \pm 0.2024$$

0.0976	$\leq \mu_1 - \mu_2 \leq$	0.5024
--------	---------------------------	--------

A un nivel de confianza del 90 % se tiene que la diferencia del promedio de grasa de la marca A menos la marca B oscila de 0.0976 a 0.5024 gramos

b)

		Marca	
		A	B
Media muestral	$\bar{x}$	3.5	3.2
Desviación estándar	σ	0.6	0.8
Varianza	σ^2	0.36	0.64
Tamaño de muestra	n	60	70

Diferencia de medias	$\bar{x}_1 - \bar{x}_2$	0.3
Nivel de confianza	NC	0.95
Valor tipificado	z	1.96
Lado derecho	LD	0.2412

$$\mu_1 - \mu_2 = (\bar{x}_1 - \bar{x}_2) \pm z \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Lado derecho (LD)

$$\bar{\mu}_1 - \mu_2 = 0.3 \pm 1.96 \sqrt{\frac{0.36}{60} + \frac{0.64}{70}}$$

$$\bar{\mu}_1 - \mu_2 = 0.3 \pm 0.2412$$

0.0588	$\leq \mu_1 - \mu_2 \leq$	0.5412
--------	---------------------------	--------

A un nivel de confianza del 95 % se tiene que la diferencia del promedio de grasa de la marca A menos la marca B oscila de 0.0588 a 0.5412 gramos

Ver video de solución dando clic en el siguiente link <https://youtu.be/2uvgEvG7vUs>

Ejemplo 2. Un fabricante de galletas está interesado en estimar la diferencia en la preferencia de los consumidores por dos nuevas recetas de galletas. Se realizan encuestas a 150 consumidores para la receta A y 120 consumidores para la receta B. Las desviaciones estándar poblacionales para la preferencia (en una escala del 1 al 10) son de 1.5 para la receta A y 1.2 para la receta B.

a) Si la media de preferencia muestral para la receta A es 7.2 y para la receta B es 6.8, calcule un intervalo de confianza del 85% para la diferencia de medias poblacionales.

b) Si la media de preferencia muestral para la receta A es 7.2 y para la receta B es 6.8, calcule un intervalo de confianza del 94% para la diferencia de medias poblacionales.

a)

		Marca	
		A	B
Media muestral	$\bar{x}$	7.2	6.8
Desviación estándar	σ	1.5	1.2
Varianza	σ^2	2.25	1.44
Tamaño de muestra	n	150	120

$$\mu_1 - \mu_2 = (\bar{x}_1 - \bar{x}_2) \pm z \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Lado derecho (LD)

Diferencia de medias	$\bar{x}_1 - \bar{x}_2$	0.4
Nivel de confianza	NC	0.85
Valor tipificado	z	1.44
Lado derecho	LD	0.2365

$$\bar{\mu}_1 - \mu_2 = 0.4 \pm 1.44 \sqrt{\frac{2.25}{150} + \frac{1.44}{120}}$$

$$\bar{\mu}_1 - \mu_2 = 0.4 \pm 0.2365$$

$$0.1635 \leq \mu_1 - \mu_2 \leq 0.6365$$

A un nivel de confianza del 85 % se tiene que la diferencia del promedio de aceptación de la receta A menos la receta B oscila de 0.1635 a 0.6365

b)

		Marca	
		A	B
Media muestral	$\bar{x}$	7.2	6.8
Desviación estándar	σ	1.5	1.2
Varianza	σ^2	2.25	1.44
Tamaño de muestra	n	150	120

$$\mu_1 - \mu_2 = (\bar{x}_1 - \bar{x}_2) \pm z \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Lado derecho (LD)

Diferencia de medias	$\bar{x}_1 - \bar{x}_2$	0.4
Nivel de confianza	NC	0.94
Valor tipificado	z	1.88
Lado derecho	LD	0.3090

$$\bar{\mu}_1 - \mu_2 = 0.4 \pm 1.88 \sqrt{\frac{2.25}{150} + \frac{1.44}{120}}$$

$$\bar{\mu}_1 - \mu_2 = 0.4 \pm 0.3090$$

$$0.0910 \leq \mu_1 - \mu_2 \leq 0.7090$$

A un nivel de confianza del 94 % se tiene que la diferencia del promedio de aceptación de la receta A menos la receta B oscila de 0.0910 a 0.7090

Ejercicios propuestos.

Ejercicio 1. Una empresa de alimentos para mascotas desea estimar la diferencia en el aumento de peso promedio de perros alimentados con dos nuevas fórmulas. Se alimentan muestras de 90 perros con la fórmula X y 100 perros con la fórmula Y durante dos meses. Las desviaciones estándar poblacionales para el aumento de peso son de 2.5 kg para la fórmula X y 2 kg para la fórmula Y.

a) Si la media del aumento de peso muestral para la fórmula X es 4 kg y para la fórmula Y es 3.5 kg, construye un intervalo de confianza del 99% para la diferencia de medias poblacionales. Respuesta $-.03521 \leq \mu_X - \mu_Y \leq 1.3521$

Ejercicio 2. Un productor de jugos quiere estimar la diferencia en el contenido de vitamina C entre dos métodos de extracción de jugo de naranja. Se miden los niveles de vitamina C en muestras de 80 naranjas procesadas con el método 1 y 90 naranjas procesadas con el método 2. Las desviaciones estándar poblacionales son de 0.7 mg/100 ml para el método 1 y 0.5 mg/100 ml para el método 2.

a) Si la media del contenido de vitamina C muestral para el método 1 es 28 mg/100 ml y para el método 2 es 26 mg/100 ml, construye un intervalo de confianza del 90% para la diferencia de medias poblacionales. Respuesta $1.8448 \leq \mu_1 - \mu_2 \leq 2.1552$

Ejercicio 3. Una empresa de snacks está comparando dos nuevos sabores de papas fritas. Se realizan encuestas a 120 consumidores para el sabor A y 150 consumidores para el sabor B. Las desviaciones estándar poblacionales para la preferencia (en una escala del 1 al 5) son de 0.8 para el sabor A y 0.6 para el sabor B.

a) Si la media de preferencia muestral para el sabor A es 4.1 y para el sabor B es 3.9, construye un intervalo de confianza del 95% para la diferencia de medias poblacionales. Respuesta $0.0276 \leq \mu_A - \mu_B \leq 0.3724$

Ejercicio 4. Un fabricante de cereales está interesado en estimar la diferencia en el peso promedio de los paquetes llenados por dos máquinas envasadoras. Se pesan muestras de 100 paquetes llenados por cada máquina. Las desviaciones estándar poblacionales para el peso de los paquetes son de 4 gramos para la máquina 1 y 3 gramos para la máquina 2.

a) Si la media del peso de los paquetes muestral para la máquina 1 es 505 gramos y para la máquina 2 es 502 gramos, construye un intervalo de confianza del 99% para la diferencia de medias poblacionales. Respuesta $1.7121 \leq \mu_1 - \mu_2 \leq 4.2879$

Ejercicio 5. Una empresa de bebidas quiere estimar la diferencia en la pureza de la azúcar suministrada por dos proveedores. Se miden los niveles de pureza en muestras de 80 lotes de azúcar del proveedor A y 90 lotes de azúcar del proveedor B. Las desviaciones estándar poblacionales para la pureza son de 0.3% para el proveedor A y 0.2% para el proveedor B.

a) Si la media de pureza muestral para el proveedor A es 99.2% y para el proveedor B es 99.5%, construye un intervalo de confianza del 90% para la diferencia de medias poblacionales. Respuesta $0.198 \leq \mu_B - \mu_A \leq 0.402$

Caso 4. Estimación de intervalo para la diferencia de medias con desviaciones estándar poblacional desconocidas e iguales

Las fórmulas correspondientes a este caso son.

$$\mu_1 - \mu_2 = (\bar{x}_1 - \bar{x}_2) \pm t s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$gl = n_1 + n_2 - 2$$

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

En estas fórmulas se puede apreciar que son dos poblaciones y que de cada una de ellas se tiene información en cuanto a su media muestral, su desviación estándar muestral y su tamaño de muestra. Adicional a lo anterior se observa que se emplea el estadístico t de student.

Ejemplo 1. Una empresa de procesamiento de carne está comparando dos métodos de conservación para la carne de res. Se toman muestras aleatorias de 10 piezas de carne conservadas con el método A y 12 piezas con el método B. Si la media de vida útil muestral para el método A es de 25 días y para el método B es de 28 días, y las desviaciones estándar muestrales son de 3 días y 2.5 días respectivamente. Calcule un intervalo de confianza del 95% para la diferencia de medias poblacionales considerando que las desviaciones estándar poblacionales son desconocidas e iguales.

		Método	
		A	B
Media muestral	$\bar{x}$	25	28
Desviación estándar	s	3	2.5
Varianza	s^2	9	6.25
Tamaño de muestra	n	10	12

Diferencia de medias	$\bar{x}_B - \bar{x}_A$	3
Nivel de confianza	NC	0.95
Nivel de significancia	α	0.05
Grados de libertad	gl	20
Valor t de student	t	2.086
Desviación std conjunta	s_p	2.736
Lado derecho	LD	2.44

$$gl = n_1 + n_2 - 2$$

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

$$\mu_1 - \mu_2 = (\bar{x}_1 - \bar{x}_2) \pm t s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Lado derecho (LD)

$$gl = 10 + 12 - 2 = 20$$

$$s_p = \sqrt{\frac{(10-1)*9 + (12-1)*6.25}{10+12-2}} = 2.736$$

$$\mu_1 - \mu_2 = 3 \pm 2.086 * 2.736 * \sqrt{\frac{1}{10} + \frac{1}{12}}$$

$$\bar{\mu}_1 - \mu_2 = 0.3 \pm 2.44$$

0.556	$\leq \mu_1 - \mu_2 \leq$	5.444
-------	---------------------------	-------

A un nivel de confianza del 95 % se tiene que la diferencia del promedio de conservación del método B menos el método A oscila de 0.556 a 5.444 días.

Ver video de solución dando clic en el siguiente link <https://youtu.be/dRk1y4TJIss>

Ejemplo 2. Un fabricante de aderezos para ensaladas está probando dos nuevas fórmulas. Se realizan pruebas de sabor con consumidores y se mide su preferencia en una escala del 1 al 10. Si la media de preferencia proporcionada por 8 personas de la fórmula X es 7.8 y de 12 personas para la fórmula Y es 6.5. Si las desviaciones estándar muestrales son de 1.5 y 1.2 respectivamente, construye un intervalo de confianza del 90% para la diferencia de medias poblacionales considerando que las desviaciones estándar poblacionales son desconocidas e iguales.

		Fórmula	
		X	Y
Media muestral	$\bar{x}$	7.8	6.5
Desviación estándar	s	1.5	1.2
Varianza	s^2	2.25	1.44
Tamaño de muestra	n	8	9

Diferencia de medias	$\bar{x}_X - \bar{x}_Y$	1.3
Nivel de confianza	NC	0.9
Nivel de significancia	α	0.1
Grados de libertad	gl	15
Valor t de student	t	1.753
Desviación std conjunta	s_p	1.348
Lado derecho	LD	1.15

$$gl = n_1 + n_2 - 2$$

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

$$\mu_1 - \mu_2 = (\bar{x}_1 - \bar{x}_2) \pm t s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Lado derecho (LD)

$$gl = 8 + 9 - 2 = 15$$

$$s_p = \sqrt{\frac{(8-1)*2.25 + (9-1)*1.44}{8+9-2}} = 1.348$$

$$\mu_1 - \mu_2 = 1.3 \pm 1.753 * 1.348 * \sqrt{\frac{1}{8} + \frac{1}{9}}$$

$$\bar{\mu}_1 - \mu_2 = 3.3 \pm 1.15$$

0.151	$\leq \mu_1 - \mu_2 \leq$	2.449
-------	---------------------------	-------

A un nivel de confianza del 90 % se tiene que la diferencia del promedio de preferencia de la fórmula X menos la fórmula Y oscila de 0.151 a 2.449 días.

Ejercicios propuestos

Ejercicio 1. Una empresa de producción de aceite de oliva está comparando el rendimiento de dos variedades de aceitunas. Se cosechan muestras de 15 árboles de cada variedad y se mide la cantidad de aceite producido por árbol. Si la media de producción de aceite muestral para la variedad A es de 12 kg y para la variedad B es de 10 kg, y las desviaciones estándar muestrales son de 2 kg y 1.5 kg respectivamente, construye un intervalo de confianza del 99% para la diferencia de medias poblacionales. Considerando que las desviaciones estándar poblacionales son desconocidas e iguales. Respuesta $0.2163 \leq \mu_A - \mu_B \leq 3.7837$

Ejercicio 2. Un productor de vino está evaluando el efecto de dos tipos de levadura en el proceso de fermentación. Se miden los niveles de alcohol en 13 muestras de vino elaboradas con cada tipo de levadura. Si la media del contenido de alcohol muestral para la levadura 1 es 13.5% y para la levadura 2 es 14%, y las desviaciones estándar muestrales son de 0.5% y 0.4% respectivamente, construye un intervalo de confianza del 90% para la diferencia de medias poblacionales. Considerando que las desviaciones estándar poblacionales son desconocidas e iguales. *Respuesta* $0.1962 \leq \mu_2 - \mu_1 \leq 0.8038$

Ejercicio 3. Una empresa de alimentos congelados está comparando dos métodos de congelación. Se mide la pérdida de peso de los alimentos después de un cierto período de tiempo. Si la media de pérdida de peso de 14 muestras del método A es de 5% y para 9 muestras del método B es de 3%, y las desviaciones estándar muestrales son de 1% y 0.8% respectivamente, construye un intervalo de confianza del 95% para la diferencia de medias poblacionales. Considerando que las desviaciones estándar poblacionales son desconocidas e iguales. *Respuesta* $0.198 \leq \mu_B - \mu_A \leq 0.402$

Ejercicio 4. Una empresa que produce alimentos congelados quiere comparar la cantidad de tiempo que tardan en descongelarse dos tipos de pescado: el salmón y el bacalao. Se toman muestras aleatorias de 14 filetes de salmón y 15 filetes de bacalao, y se mide el tiempo de descongelación en minutos. Si el tiempo promedio de descongelación del salmón es de 28 minutos con una desviación estándar muestral de 5 minutos, y el tiempo promedio de descongelación del bacalao es de 32 minutos con una desviación estándar muestral de 6 minutos, construye un intervalo de confianza del 95% para la diferencia de medias poblacionales. Considerando que las desviaciones estándar poblacionales son desconocidas e iguales. *Respuesta* $1.1747 \leq \mu_A - \mu_B \leq 2.8253$

Ejercicio 5. Una compañía que fabrica jugos está investigando dos métodos de extracción de jugo de naranja. Se toman muestras aleatorias de 20 naranjas para el método A y 18 naranjas para el método B, y se mide la cantidad de jugo extraído en mililitros. Si la cantidad promedio de jugo extraído con el método A es de 120 ml con una desviación estándar muestral de 10 ml, y la cantidad promedio de jugo extraído con el método B es de 130 ml con una desviación estándar muestral de 12 ml, construye un intervalo de confianza del 90% para la diferencia de medias poblacionales. Considerando que las desviaciones estándar poblacionales son desconocidas e iguales. *Respuesta* $0.198 \leq \mu_B - \mu_A \leq 0.402$

Caso 5. Estimación de intervalo para la diferencia de medias con desviaciones estándar poblacional desconocidas y diferentes.

Las fórmulas correspondientes a este caso son.

$$\mu_1 - \mu_2 = (\bar{x}_1 - \bar{x}_2) \pm t \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$gl = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}}$$

En estas fórmulas se puede apreciar que son dos poblaciones y que de cada una de ellas se tiene información en cuanto a su media muestral, su desviación estándar muestral y su tamaño de muestra. Adicional a lo anterior se observa que se emplea el estadístico t de student.

Ejemplo 1. Un nutricionista está interesado en comparar la cantidad promedio de proteínas por cada 100 gramos de carne de pollo y carne de res. Se toman muestras aleatorias de 12 muestras de carne de pollo y 16 muestras de carne de res, y se mide la cantidad de proteínas en gramos. Si la cantidad promedio de proteínas en la muestra de carne de pollo es de 25 gramos con una desviación estándar muestral de 3 gramos, y la cantidad promedio de proteínas en la muestra de carne de res es de 22 gramos con una desviación estándar muestral de 4 gramos, construye un intervalo de confianza del 95% para la diferencia de medias poblacionales. Considerando que las desviaciones estándar poblacionales son desconocidas y diferentes.

		Método	
		Pollo	Res
Media muestral	$\bar{x}$	25	22
Desviación estándar	s	3	4
Varianza	s^2	9	16
Tamaño de muestra	n	12	16

Diferencia de medias	$\bar{x}_P - \bar{x}_R$	3
Nivel de confianza	NC	0.95
Nivel de significancia	α	0.05
Grados de libertad	gl	25.9968
Valor t de student	t	2.056
Lado derecho	LD	2.7192

$$\mu_1 - \mu_2 = (\bar{x}_1 - \bar{x}_2) \pm t \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Lado derecho (LD)

$$gl = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}} \quad gl = \frac{\left(\frac{9}{12} + \frac{16}{16}\right)^2}{\frac{\left(\frac{9}{12}\right)^2}{9 - 1} + \frac{\left(\frac{16}{16}\right)^2}{16 - 1}} = 25.99$$

26

$$\mu_1 - \mu_2 = 3 \pm 2.060 * \sqrt{\frac{9}{12} + \frac{16}{16}}$$

$$\bar{\mu}_1 - \mu_2 = 0.3 \pm 2.72$$

$$0.2808 \leq \mu_P - \mu_R \leq 5.7192$$

A un nivel de confianza del 95 % se tiene que la diferencia del promedio de proteínas de la carne de pollo menos la carne de res oscila de 0.2808 a 5.7192 gramos de proteína por cada 100 gramos de carne.

Ver video de solución dando clic en el siguiente link <https://youtu.be/5DoJ4ZUIzRc>

Ejemplo 2. Un investigador desea comparar la cantidad de vitamina C presente en naranjas y limones. Para ello, toma una muestra aleatoria de 10 naranjas y 13 limones, y mide la cantidad de vitamina C en miligramos por cada 100 gramos de fruta. Si la cantidad promedio de vitamina C en la muestra de naranjas es de 52 mg con una desviación estándar muestral de 5 mg, y la cantidad promedio de vitamina C en la muestra de limones es de 45 mg con una desviación estándar muestral de 7 mg, construye un intervalo de confianza del 90% para la diferencia de medias poblacionales. Considerando que las desviaciones estándar poblacionales son desconocidas y diferentes.

		Método	
		Naranjas	Limones
Media muestral	$\bar{x}$	52	45
Desviación estándar	s	5	7
Varianza	s^2	25	49
Tamaño de muestra	n	10	13

Diferencia de medias	$\bar{x}_N - \bar{x}_L$	7
Nivel de confianza	NC	0.9
Nivel de significancia	α	0.1
Grados de libertad	gl	20.9241
Valor t de student	t	1.721
Lado derecho	LD	4.3085

$$\mu_1 - \mu_2 = (\bar{x}_1 - \bar{x}_2) \pm t \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Lado derecho (LD)

$$gl = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}} \quad gl = \frac{\left(\frac{25}{10} + \frac{49}{13}\right)^2}{\frac{\left(\frac{25}{10}\right)^2}{10 - 1} + \frac{\left(\frac{49}{13}\right)^2}{13 - 1}} = 20.92$$

21

$$\mu_1 - \mu_2 = 3 \pm 2.060 * \sqrt{\frac{9}{12} + \frac{16}{16}}$$

$$\bar{\mu}_1 - \mu_2 = 0.3 \pm 2.72$$

2.6915	$\leq \mu_N - \mu_L \leq$	11.3085
--------	---------------------------	---------

A un nivel de confianza del 90 % se tiene que la diferencia del contenido promedio de vitamina C de las naranjas menos los limones oscila de 2.6915 a 11.3085 miligramos por cada 100 gramos de fruta.

Ejercicios propuestos.

Ejercicio 1. Una empresa de alimentos desea comparar el contenido de azúcar en dos tipos de cereal: "Cereal A" y "Cereal B". Se toman muestras aleatorias de 15 cajas de "Cereal A" y 10 cajas de "Cereal B", y se mide la cantidad de azúcar en gramos por porción. Si el contenido promedio de azúcar en la muestra de "Cereal A" es de 12 gramos con una desviación estándar muestral de 2 gramos, y el contenido promedio de azúcar en la muestra de "Cereal B" es de 10 gramos con una desviación estándar muestral de 1.5 gramos, construye un intervalo de confianza del 95% para la diferencia de medias poblacionales. Considerando que las desviaciones estándar poblacionales son desconocidas y diferentes.

Respuesta $0.548 \leq \mu_A - \mu_B \leq 3.452$

Ejercicio 2. Un consumidor desea comparar el peso promedio de las bolsas de papas fritas de dos marcas diferentes: "Marca X" y "Marca Y". Se toman muestras aleatorias de 8 bolsas de "Marca X" y 10 bolsas de "Marca Y", y se pesa el contenido de cada bolsa en gramos. Si el peso promedio de las papas fritas en la muestra de "Marca X" es de 150 gramos con una desviación estándar muestral de 5 gramos, y el peso promedio de las papas fritas en la muestra de "Marca Y" es de 145 gramos con una desviación estándar muestral de 4 gramos, construye un intervalo de confianza del 90% para la diferencia de medias poblacionales. Considerando que las desviaciones estándar poblacionales son desconocidas y diferentes.

$$\text{Respuesta } 1.157 \leq \mu_X - \mu_Y \leq 8.843$$

Ejercicio 3. Un nutricionista quiere comparar la cantidad de sodio en dos tipos de sopa enlatada: "Sopa de pollo" y "Sopa de vegetales". Se toman muestras aleatorias de 13 latas de "Sopa de pollo" y 15 latas de "Sopa de vegetales", y se mide la cantidad de sodio en miligramos por porción. Si la cantidad promedio de sodio en la muestra de "Sopa de pollo" es de 400 mg con una desviación estándar muestral de 30 mg, y la cantidad promedio de sodio en la muestra de "Sopa de vegetales" es de 350 mg con una desviación estándar muestral de 25 mg, construye un intervalo de confianza del 99% para la diferencia de medias poblacionales. Considerando que las desviaciones estándar poblacionales son desconocidas y diferentes. Respuesta $20.4921 \leq \mu_P - \mu_V \leq 79.5079$

Ejercicio 4. Un chef está probando dos marcas de pasta: "Marca A" y "Marca B". Se cocinan 8 muestras de pasta de cada marca y se mide el tiempo de cocción en minutos. Si el tiempo promedio de cocción de la muestra de "Marca A" es de 10 minutos con una desviación estándar muestral de 1 minuto, y el tiempo promedio de cocción de la muestra de "Marca B" es de 12 minutos con una desviación estándar muestral de 1.5 minutos, construye un intervalo de confianza del 95% para la diferencia de medias poblacionales. Considerando que las desviaciones estándar poblacionales son desconocidas y diferentes.

$$\text{Respuesta } 0.6137 \leq \mu_B - \mu_A \leq 3.3863$$

Ejercicio 5. Una empresa láctea desea comparar el contenido de grasa en dos tipos de leche: "Leche entera" y "Leche descremada". Se toman muestras aleatorias de 16 botellas de "Leche entera" y 12 botellas de "Leche descremada", y se mide el contenido de grasa en gramos por porción. Si el contenido promedio de grasa en la muestra de "Leche entera" es de 10 gramos con una desviación estándar muestral de 2 gramos, y el contenido promedio de grasa en la muestra de "Leche descremada" es de 2 gramos con una desviación estándar muestral de 0.5 gramos, construye un intervalo de confianza del 90% para la diferencia de medias poblacionales. Considerando que las desviaciones estándar poblacionales son desconocidas y diferentes. Respuesta $7.096 \leq \mu_E - \mu_D \leq 8.904$

Caso 6. Estimación de intervalo para la diferencia de medias para datos pareados.

Los datos pareados se utilizan cuando existe una relación natural entre las observaciones de los dos grupos. Esto ocurre, por ejemplo, en los siguientes casos:

- Mediciones repetidas en el mismo individuo: Se mide la presión arterial de un paciente antes y después de tomar un medicamento.
- Comparación de dos tratamientos en gemelos idénticos: Un gemelo recibe el tratamiento A y el otro gemelo recibe el tratamiento B.
- Evaluación de dos métodos de enseñanza en el mismo grupo de estudiantes: Se mide el rendimiento de los estudiantes después de recibir enseñanza con el método A y luego con el método B.

La fórmula correspondiente a este caso es.

$$\mu_1 - \mu_2 = \bar{d} \pm t \frac{s_d}{\sqrt{n}}$$

Donde:

$\bar{d}$ Corresponde al promedio de las diferencias de los datos pareados.

s_d Corresponde a la desviación estándar de las diferencias de los datos pareados.

Ejemplo 1. Una empresa quesera está evaluando dos métodos de elaboración de queso cheddar: el método tradicional (Método A) y un nuevo método (Método B) que se espera que acelere el proceso de maduración. Se elaboran siete lotes de queso, cada uno con ambos métodos, y se mide la acidez del queso (expresada en grados Dornic) después de 30 días de maduración. Los resultados se muestran en la siguiente tabla:

Lote	Método A	Método B
1	18.2	19.5
2	17.5	18.8
3	19.1	20.3
4	18.8	19.9
5	17.9	19.2
6	18.5	19.6
7	18	19.1

Construye un intervalo de confianza del 95% para la diferencia de medias poblacionales en la acidez del queso entre los dos métodos.

Lote	Método A	Método B	d
1	18.2	19.5	1.3
2	17.5	18.8	1.3
3	19.1	20.3	1.2
4	18.8	19.9	1.1
5	17.9	19.2	1.3
6	18.5	19.6	1.1
7	18	19.1	1.1

Diferencia promedio	$\bar{d}$	1.2
Desviación estándar	s_d	0.10
Nivel de confianza	NC	0.95
Nivel de significancia	α	0.05
Tamaño de muestra	n	7
Grados de libertad	gl	6
Valor t de student	t	2.447
Lado derecho	LD	0.09

$$\mu_1 - \mu_2 = \bar{d} \pm t \frac{s_d}{\sqrt{n}} \quad \text{Lado derecho (LD)}$$

$$\mu = 1.2 \pm 2.447 * \frac{0.1}{\sqrt{7}}$$

$$\mu = 1.2 \pm 0.09$$

1.11	$\leq \mu \leq$	1.29
------	-----------------	------

A un nivel de confianza del 95 % se tiene que la diferencia en la acidez promedio del método B menos el método A oscila de 1.11 a 1.29 grados Dornic

Ver video de solución dando clic en el siguiente link <https://youtu.be/2VXU7NPnyz0>

Ejercicio 2. Un restaurante está experimentando con dos métodos de cocción de carne de res: el método tradicional a la parrilla (Método A) y un nuevo método de cocción al vacío (Método B). Se cocinan 10 cortes de carne similares, cada uno con ambos métodos, y se mide la jugosidad de la carne (en una escala del 1 al 10) por un panel de catadores. Los resultados se muestran en la siguiente tabla:

Corte	Método A (Parrilla)	Método B (Vacío)
1	7	9
2	6	8
3	8	9
4	5	7
5	7	8
6	9	10
7	6	7
8	8	9
9	7	8
10	5	6

Construye un intervalo de confianza del 90% para la diferencia de medias poblacionales en la jugosidad de la carne entre los dos métodos.

Corte	Método A (Parrilla)	Método B (Vacío)	d
1	7	9	2
2	6	8	2
3	8	9	1
4	5	7	2
5	7	8	1
6	9	10	1
7	6	7	1
8	8	9	1
9	7	8	1
10	5	6	1

Diferencia promedio	$\bar{d}$	1.3
Desviación estándar	s_d	0.48
Nivel de confianza	NC	0.9
Nivel de significancia	α	0.1
Tamaño de muestra	n	10
Grados de libertad	gl	9
Valor t de student	t	1.833
Lado derecho	LD	0.28

Lado derecho (LD)

$$\mu_1 - \mu_2 = \bar{d} \pm t \frac{s_d}{\sqrt{n}}$$

$$\mu = 1.3 \pm 1.833 * \frac{0.48}{\sqrt{10}}$$

$$\mu = 1.3 \pm 0.28$$

1.02	$\leq \mu_B - \mu_A \leq$	1.58
-------------	---------------------------	-------------

A un nivel de confianza del 90 % se tiene que la diferencia en la jugosidad promedio de la carne del método de cocción al vacío menos el método de cocción a la parrilla oscila de 1.02 a 1.58.

Ejercicios propuestos.

Ejercicio 1. Una panadería está probando dos métodos de horneado de pan: el método tradicional en horno de leña (Método A) y un nuevo método en horno eléctrico (Método B). Se hornean 8 pares de panes, uno con cada método, y se mide el volumen de cada pan (en centímetros cúbicos). Los resultados se muestran en la siguiente tabla:

Par	Método A (Leña)	Método B (Eléctrico)
1	1200	1280
2	1150	1210
3	1300	1340
4	1250	1325

5	1100	1180
6	1200	1275
7	1350	1410
8	1150	1220

Construye un intervalo de confianza del 95% para la diferencia de medias poblacionales en el volumen del pan entre los dos métodos. *Respuesta* $58.37 \leq \mu_B - \mu_A \leq 76.63$

Ejercicio 2. Una empresa de alimentos está comparando dos tipos de harina de maíz: harina de maíz nixtamalizada (Harina A) y harina de maíz cruda (Harina B) en la elaboración de tortillas. Se elaboran 10 pares de tortillas, cada par con ambos tipos de harina, y se mide la textura de cada tortilla (en una escala del 1 al 10). Los resultados se muestran en la siguiente tabla:

Par	Harina A (Nixtamalizada)	Harina B (Cruda)
1	8	7
2	10	7
3	8	5
4	8	5
5	9	7
6	8	5
7	8	6
8	9	6
9	7	4
10	8	7

Construye un intervalo de confianza del 90% para la diferencia de medias poblacionales en la textura de las tortillas entre los dos tipos de harina. *Respuesta* $1.91 \leq \mu_A - \mu_B \leq 2.89$

Ejercicio 3. Una empresa de café está evaluando dos métodos de envasado: envasado al vacío (Método A) y envasado en atmósfera modificada (Método B). Se envasan 8 lotes de café, cada lote con ambos métodos, y se mide la frescura del café (en una escala del 1 al 100) después de un mes de almacenamiento. Los resultados se muestran en la siguiente tabla:

Lote	Método A (Vacío)	Método B (Modificada)
1	95	90
2	92	88
3	98	95
4	90	85
5	93	89
6	96	92

7	91	87
8	94	91

Construye un intervalo de confianza del 99% para la diferencia de medias poblacionales en la frescura del café entre los dos métodos de envasado. *Respuesta* $3.49 \leq \mu_A - \mu_B \leq 4.51$

Ejercicio 4. Una cervecería artesanal está experimentando con dos tipos de levadura: levadura Ale (Levadura A) y levadura Lager (Levadura B). Se elaboran 9 lotes de cerveza, cada lote con ambos tipos de levadura, y se mide el contenido de alcohol de cada cerveza (en porcentaje). Los resultados se muestran en la siguiente tabla:

Lote	Levadura A (Ale)	Levadura B (Lager)
1	5.2	4.8
2	5.5	5.3
3	5	4.8
4	5.3	4.6
5	5.1	4.9
6	5.4	5.1
7	5.2	4.9
8	5.6	5
9	5.3	5.1

Construye un intervalo de confianza del 95% para la diferencia de medias poblacionales en el contenido de alcohol de la cerveza entre los dos tipos de levadura.

Respuesta $0.23 \leq \mu_A - \mu_B \leq 0.46$

Ejercicio 5. Un productor de aguacate está comparando dos métodos de maduración: maduración natural (Método A) y maduración controlada con etileno (Método B). Se seleccionan 9 pares de aguacates similares, cada par con ambos métodos de maduración, y se mide la firmeza de cada aguacate (en una escala del 1 al 10) después de una semana. Los resultados se muestran en la siguiente tabla:

Par	Método A (Natural)	Método B (Etileno)
1	7.5	9
2	6.9	8
3	8	9.8
4	5.4	7
5	7.3	9
6	9.1	10
7	6.4	8
8	8.1	9
9	7.3	9

Construye un intervalo de confianza del 90% para la diferencia de medias poblacionales en la firmeza del aguacate entre los dos métodos de maduración.

Respuesta $1.20 \leq \mu_B - \mu_A \leq 1.64$

Caso 7. Estimación de intervalo para la proporción de una muestra grande.

La fórmula correspondiente a este caso es.

$$P = \hat{p} \pm z \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

$$\hat{p} = \frac{x}{n} \quad \hat{q} = 1 - \hat{p}$$

Como el tamaño de muestra utilizado para estimar el intervalo de la proporción es grande ($n \geq 30$), sigue una distribución normal y eso se denota en la fórmula por el valor tipificado z .

Ejemplo1. Una empresa de helados está lanzando un nuevo sabor y quiere estimar la proporción de consumidores que lo prefieren. Realizan una encuesta a 500 personas y 320 dicen que les gusta el nuevo sabor. Construye un intervalo de confianza del 95% para la proporción poblacional de consumidores que prefieren el nuevo sabor.

Tamaño de muestra	n	500
Exitos en la muestra	x	320
Proporción muestral	$\hat{p}$	0.64
	$\hat{q}$	0.36
Nivel de confianza	NC	0.95
Valor tipificado	z	1.96

Lado derecho	LD	0.04
--------------	----	------

$$P = \hat{p} \pm z \sqrt{\frac{\hat{p}\hat{q}}{n}} \quad \text{Lado derecho (LD)}$$

$$P = 0.64 \pm 1.96 \sqrt{\frac{0.64 * 0.36}{500}}$$

$$P = 0.64 \pm 0.04$$

0.60	$\leq P \leq$	0.68
-------------	---------------------------------	-------------

A un nivel de confianza del 95 % se tiene que la proporción de personas a las que les gusta el nuevo sabor oscila de 0.60 a 0.68.

Ver video de solución dando clic en el siguiente link <https://youtu.be/NDDbjopYItw>

Ejemplo 2. Una empresa de alimentos enlatados tiene una línea de producción donde se envasan miles de latas al día. Se toma una muestra aleatoria de 200 latas y se encuentran 12 defectuosas. Construye un intervalo de confianza del 90% para la proporción de latas defectuosas en toda la producción.

Tamaño de muestra	n	200
Éxitos en la muestra	x	12
Proporción muestral	$\hat{p}$	0.06
	$\hat{q}$	0.94
Nivel de confianza	NC	0.9
Valor tipificado	z	1.64

Lado derecho	LD	0.03
--------------	----	------

$$P = \hat{p} \pm z \sqrt{\frac{\hat{p}\hat{q}}{n}} \quad \text{Lado derecho (LD)}$$

$$P = 0.06 \pm 1.64 \sqrt{\frac{0.06 * 0.94}{200}}$$

$$P = 0.06 \pm 0.03$$

0.03	≤ P ≤	0.09
-------------	--------------	-------------

A un nivel de confianza del 90 % se tiene que la proporción de latas defectuosas oscila de 0.03 a 0.09.

Ejercicios propuestos.

Ejercicio 1. Una empresa de investigación de mercado quiere estimar la proporción de personas que consumen productos orgánicos regularmente. Realizan una encuesta a 800 personas y 240 responden que sí consumen productos orgánicos. Construye un intervalo de confianza del 95% para la proporción poblacional de personas que consumen productos orgánicos. *Respuesta* $0.2682 \leq P \leq 0.3318$

Ejercicio 2. Un inspector de salud realiza una inspección sorpresa a 150 restaurantes en una ciudad y encuentra que 120 cumplen con las normas de higiene. Construye un intervalo de confianza del 90% para la proporción de restaurantes que cumplen con las normas de higiene en toda la ciudad. *Respuesta* $0.7463 \leq P \leq 0.8537$

Ejercicio 3. Una empresa de café realiza una prueba de sabor con 300 consumidores. 180 personas dicen que prefieren la marca A sobre la marca B. Construye un intervalo de confianza del 95% para la proporción poblacional de consumidores que prefieren la marca A. *Respuesta* $0.5446 \leq P \leq 0.6554$

Ejercicio 4. Una empresa de alimentos está lanzando un nuevo producto y quiere estimar la proporción de supermercados que lo venden. Visitan 200 supermercados y encuentran que 160 lo tienen en sus estantes. Construye un intervalo de confianza del 90% para la proporción de supermercados que venden el nuevo producto en toda la región. *Respuesta* $0.7535 \leq P \leq 0.8465$

Ejercicio 5. Una empresa de alimentos quiere estimar la proporción de personas que tienen alergia a un ingrediente específico que utilizan en sus productos. Realizan una encuesta a 1000 personas y 50 responden que tienen alergia al ingrediente. Construye un intervalo de confianza del 99% para la proporción poblacional de personas con alergia al ingrediente. *Respuesta* $0.0322 \leq P \leq 0.0678$

Caso 8. Estimación de intervalo para la diferencia de proporciones.

La fórmula correspondiente a este caso es.

$$P_1 - P_2 = (\hat{p}_1 - \hat{p}_2) \pm z \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}$$

Como los tamaños de muestra utilizados para estimar el intervalo de la diferencia de proporciones son grandes ($n \geq 30$), sigue una distribución normal y eso se denota en la fórmula por el valor tipificado z .

Ejemplo 1. Una empresa productora de latas de frijoles desea comparar la proporción de latas defectuosas producidas por dos líneas de producción diferentes. Se toman muestras aleatorias de 200 latas de la línea A y 250 latas de la línea B. Se encuentran 10 latas defectuosas en la línea A y 15 en la línea B. Estima el intervalo de confianza del 95% para la diferencia en la proporción de latas defectuosas entre las dos líneas.

		Línea	
		1	2
Tamaño de muestra	n	200	250
Éxitos en la muestra	x	10	15
Proporción muestral	$\hat{p}$	0.05	0.06
	$\hat{q}$	0.95	0.94

$$P_1 - P_2 = (\hat{p}_1 - \hat{p}_2) \pm z \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}$$

Lado derecho (LD)

Dif. de proporciones	$\hat{p}_2 - \hat{p}_1$	0.01
Nivel de confianza	NC	0.95
Valor tipificado	z	1.96
Lado derecho	LD	0.04

$$P_2 - P_1 = 0.01 \pm 1.96 \sqrt{\frac{0.05 * 0.95}{200} + \frac{0.06 * 0.94}{250}}$$

$$P_2 - P_1 = 0.01 \pm 0.04$$

$$-0.0322 \leq P_2 - P_1 \leq 0.0522$$

A un nivel de confianza del 95 % se tiene que la diferencia en la proporción de latas defectuosas producidas por la línea de producción 2 menos la 1 oscila de -0.0322 a 0.0522.

Ver video de solución dando clic en el siguiente link <https://youtu.be/2zDKDAYaA8c>

Ejemplo 2. Una empresa de cereales realiza una encuesta para comparar la preferencia de dos marcas de cereal en niños. Se encuestan 300 niños y se encuentra que 150 prefieren la marca X y 100 prefieren la marca Y. Estima el intervalo de confianza del 90% para la diferencia en la proporción de niños que prefieren cada marca.

		Marca	
		X	Y
Tamaño de muestra	n	300	300
Éxitos en la muestra	x	150	100
Proporción muestral	$\hat{p}$	0.5	0.3333
	$\hat{q}$	0.5	0.6667

$$P_1 - P_2 = (\hat{p}_1 - \hat{p}_2) \pm z \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}$$

Lado derecho (LD)

Dif. de proporciones	$\hat{p}_X - \hat{p}_Y$	0.1667
Nivel de confianza	NC	0.9
Valor tipificado	z	1.64
Lado derecho	LD	0.07

$$P_X - P_Y = 0.1667 \pm 1.64 \sqrt{\frac{0.5 * 0.5}{300} + \frac{0.333 * 0.666}{300}}$$

$$P_X - P_Y = 0.1667 \pm 0.07$$

$$0.1014 \leq P_X - P_Y \leq 0.2319$$

A un nivel de confianza del 90 % se tiene que la diferencia en la proporción de niños que prefieren la marca de cereal X menos los que prefieren la marca de cereal Y oscila de - 0.1014 a 0.2319.

Ejercicios propuestos.

Ejercicio 1. Un fabricante de alimentos está probando dos conservantes diferentes para prolongar la vida útil de un producto. Se toman muestras aleatorias de 150 unidades de producto con cada conservante. Se encuentra que 20 unidades se echan a perder con el conservante A y 12 unidades se echan a perder con el conservante B. Estima el intervalo de confianza del 99% para la diferencia en la proporción de unidades que se echan a perder con cada conservante. *Respuesta* $-0.0381 \leq P_A - P_B \leq 0.1448$

Ejercicio 2. Una cadena de restaurantes desea comparar la satisfacción del cliente en dos de sus sucursales. Se encuestan 500 clientes en cada sucursal y se encuentra que 400 están satisfechos en la sucursal 1 y 350 están satisfechos en la sucursal 2. Estima el intervalo de confianza del 95% para la diferencia en la proporción de clientes satisfechos entre las dos sucursales. *Respuesta* $0.0467 \leq P_1 - P_2 \leq 0.1533$

Ejercicio 3. Se realiza una encuesta para comparar la opinión sobre alimentos transgénicos entre dos grupos de personas. Se encuestan 400 personas en el grupo 1 y 350 personas en el grupo 2. Se encuentra que 200 personas en el grupo 1 están a favor de los alimentos transgénicos y 150 personas en el grupo 2 también lo están. Estima el intervalo de confianza del 90% para la diferencia en la proporción de personas a favor de los alimentos transgénicos entre los dos grupos. *Respuesta* $0.0116 \leq P_1 - P_2 \leq 0.1313$

Ejercicio 4. Un fabricante de alimentos desea comparar la calidad de dos proveedores de un ingrediente clave. Se toman muestras aleatorias de 100 unidades del proveedor A y 120 unidades del proveedor B. Se encuentra que 5 unidades del proveedor A no cumplen con los estándares de calidad y 8 unidades del proveedor B tampoco los cumplen. Estima el intervalo de confianza del 95% para la diferencia en la proporción de unidades que no cumplen con

los estándares de calidad entre los dos proveedores. *Respuesta* $-0.0685 \leq P_A - P_B \leq 0.0352$

Ejercicio 5. Una empresa de alimentos está lanzando un nuevo sabor de helado y quiere comparar la aceptación del producto en dos ciudades diferentes. Se realizan pruebas de sabor con 400 personas en la ciudad A y 350 personas en la ciudad B. Se observa que 280 personas en la ciudad A aprueban el nuevo sabor, mientras que en la ciudad B lo aprueban 200 personas. Estima el intervalo de confianza del 95% para la diferencia en la proporción de personas que aprueban el nuevo sabor de helado entre las dos ciudades. *Respuesta* $0.06 \leq P_A - P_B \leq 0.1972$

Caso 9 Estimación de intervalo para la varianza.

La fórmula para este caso es.

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}$$

En esta fórmula se puede observar que se utiliza el estadístico chi cuadrada, su tabla y como consultarla se presenta a continuación.

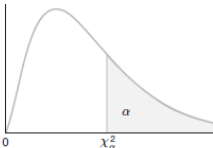
Tabla A.5 Valores críticos de la distribución chi cuadrada									
									
Tabla A.5 Valores críticos de la distribución chi cuadrada									
v	alpha								
	0.995	0.99	0.98	0.975	0.95	0.90	0.80	0.75	0.70
1	0.00393	0.0157	0.0628	0.0982	0.00393	0.0158	0.0642	0.102	0.148
2	0.0100	0.0201	0.0404	0.0506	0.103	0.211	0.446	0.575	0.713
3	0.0717	0.115	0.185	0.216	0.352	0.584	1.005	1.213	1.424
4	0.207	0.297	0.429	0.484	0.711	1.064	1.649	1.923	2.195
5	0.412	0.554	0.752	0.831	1.145	1.610	2.343	2.675	3.000
6	0.676	0.872	1.134	1.237	1.635	2.204	3.070	3.455	3.828
7	0.989	1.239	1.564	1.690	2.167	2.833	3.822	4.255	4.671
8	1.344	1.647	2.032	2.180	2.733	3.490	4.594	5.071	5.527
9	1.735	2.088	2.532	2.700	3.325	4.168	5.380	5.899	6.393
10	2.156	2.558	3.059	3.247	3.940	4.865	6.179	6.737	7.267
11	2.603	3.053	3.609	3.816	4.575	5.578	6.989	7.584	8.148
12	3.074	3.571	4.178	4.404	5.226	6.304	7.807	8.438	9.034
13	3.565	4.107	4.765	5.009	5.892	7.041	8.634	9.299	9.926
14	4.075	4.660	5.368	5.629	6.571	7.790	9.467	10.165	10.821
15	4.601	5.229	5.985	6.262	7.261	8.547	10.307	11.037	11.721
16	5.142	5.812	6.614	6.908	7.962	9.312	11.152	11.912	12.624
17	5.697	6.408	7.255	7.564	8.672	10.085	12.002	12.792	13.531
18	6.265	7.015	7.906	8.231	9.390	10.865	12.857	13.675	14.440
19	6.844	7.633	8.567	8.907	10.117	11.651	13.716	14.562	15.352
20	7.434	8.260	9.237	9.591	10.851	12.443	14.578	15.452	16.266
21	8.034	8.897	9.915	10.283	11.591	13.240	15.445	16.344	17.182
22	8.643	9.542	10.600	10.982	12.338	14.041	16.314	17.240	18.101
23	9.260	10.196	11.293	11.689	13.091	14.848	17.187	18.137	19.021
24	9.886	10.856	11.992	12.401	13.848	15.659	18.062	19.037	19.943
25	10.520	11.524	12.697	13.120	14.611	16.473	18.940	19.939	20.867
26	11.160	12.198	13.409	13.844	15.379	17.292	19.820	20.843	21.792
27	11.808	12.878	14.125	14.573	16.151	18.114	20.703	21.749	22.719
28	12.461	13.565	14.847	15.308	16.928	18.939	21.588	22.657	23.647
29	13.121	14.256	15.574	16.047	17.708	19.768	22.475	23.567	24.577
30	13.787	14.953	16.306	16.791	18.493	20.599	23.364	24.478	25.508
40	20.707	22.164	23.838	24.433	26.509	29.051	32.345	33.66	34.872
50	27.991	29.707	31.664	32.357	34.764	37.689	41.449	42.942	44.313
60	35.534	37.485	39.699	40.482	43.188	46.459	50.641	52.294	53.809

Tabla A.5 (continuación) Valores críticos de la distribución chi cuadrada									
v	alpha								
	0.30	0.25	0.20	0.10	0.05	0.025	0.02	0.01	0.005
1	1.074	1.323	1.642	2.706	3.841	5.024	5.412	6.635	7.879
2	2.408	2.773	3.219	4.605	5.991	7.378	7.824	9.210	10.597
3	3.665	4.108	4.642	6.251	7.815	9.348	9.837	11.345	12.838
4	4.878	5.385	5.989	7.779	9.488	11.143	11.668	13.277	14.860
5	6.064	6.626	7.289	9.236	11.070	12.832	13.388	15.086	16.750
6	7.231	7.841	8.558	10.645	12.592	14.449	15.033	16.812	18.548
7	8.383	9.037	9.803	12.017	14.067	16.013	16.622	18.475	20.278
8	9.524	10.219	11.030	13.362	15.507	17.535	18.168	20.090	21.955
9	10.656	11.389	12.242	14.684	16.919	19.023	19.679	21.666	23.589
10	11.781	12.549	13.442	15.987	18.307	20.483	21.161	23.209	25.188
11	12.899	13.701	14.631	17.275	19.675	21.920	22.618	24.725	26.757
12	14.011	14.845	15.812	18.549	21.026	23.337	24.054	26.217	28.300
13	15.119	15.984	16.985	19.812	22.362	24.736	25.471	27.688	29.819
14	16.222	17.117	18.151	21.064	23.685	26.119	26.873	29.141	31.319
15	17.322	18.245	19.311	22.307	24.996	27.488	28.259	30.578	32.801
16	18.418	19.369	20.465	23.542	26.296	28.845	29.633	32.000	34.267
17	19.511	20.489	21.615	24.769	27.587	30.191	30.995	33.409	35.718
18	20.601	21.605	22.760	25.989	28.869	31.526	32.346	34.805	37.156
19	21.689	22.718	23.900	27.204	30.144	32.852	33.687	36.191	38.582
20	22.775	23.828	25.038	28.412	31.410	34.170	35.020	37.566	39.997
21	23.858	24.935	26.171	29.615	32.671	35.479	36.343	38.932	41.401
22	24.939	26.039	27.301	30.813	33.924	36.781	37.659	40.289	42.796
23	26.018	27.141	28.429	32.007	35.172	38.076	38.968	41.638	44.181
24	27.096	28.241	29.553	33.196	36.415	39.364	40.270	42.980	45.558
25	28.172	29.339	30.675	34.382	37.652	40.646	41.566	44.314	46.928
26	29.246	30.435	31.795	35.563	38.885	41.923	42.856	45.642	48.290
27	30.319	31.528	32.912	36.741	40.113	43.195	44.140	46.963	49.645
28	31.391	32.620	34.027	37.916	41.337	44.461	45.419	48.278	50.994
29	32.461	33.711	35.139	39.087	42.557	45.722	46.693	49.588	52.335
30	33.530	34.800	36.250	40.256	43.773	46.979	47.962	50.892	53.672
40	44.165	45.616	47.269	51.805	55.758	59.342	60.436	63.691	66.766
50	54.723	56.334	58.164	63.167	67.505	71.420	72.613	76.154	79.490
60	65.226	66.981	68.972	74.397	79.082	83.298	84.58	88.379	91.952

Un ejemplo de cómo se consultar la tabla de chi cuadrada se presenta a continuación.

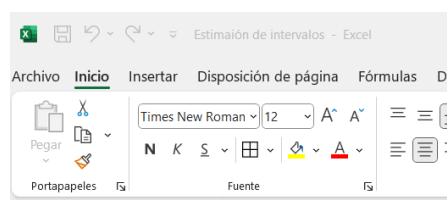
Cuál es el valor de chi cuadrada para un nivel de chi cuadrada para un nivel de significancia (α) de 0.05 y un tamaño de muestra de 20.

Como el tamaño de muestra es 20, los grados de libertad son 19.

ν	α							
	0.30	0.25	0.20	0.10	0.05	0.025	0.02	0.01
1	1.074	1.323	1.642	2.706	3.841	5.024	5.412	6.635
2	2.408	2.773	3.219	4.605	5.991	7.378	7.824	9.210
3	3.665	4.108	4.642	6.251	7.879	9.348	9.837	11.345
4	4.878	5.385	5.989	7.779	9.488	11.143	11.668	13.277
5	6.064	6.626	7.289	9.236	11.070	12.832	13.388	15.086
6	7.231	7.841	8.558	10.645	12.592	14.449	15.033	16.812
7	8.383	9.037	9.803	12.017	14.067	16.013	16.622	18.475
8	9.524	10.219	11.030	13.362	15.507	17.535	18.168	20.090
9	10.656	11.389	12.242	14.684	16.919	19.023	19.679	21.666
10	11.781	12.549	13.442	15.987	18.307	20.483	21.161	23.209
11	12.899	13.701	14.631	17.275	19.675	21.920	22.618	24.725
12	14.011	14.845	15.812	18.549	21.026	23.337	24.054	26.217
13	15.119	15.984	16.985	19.812	22.362	24.736	25.471	27.688
14	16.222	17.117	18.151	21.064	23.685	26.119	26.873	29.141
15	17.322	18.245	19.311	22.307	24.996	27.488	28.259	30.578
16	18.418	19.369	20.465	23.542	26.296	28.845	29.633	32.000
17	19.511	20.489	21.615	24.769	27.597	30.191	30.995	33.409
18	20.601	21.605	22.760	25.989	28.869	31.526	32.346	34.805
19	21.689	22.718	23.906	27.204	30.144	32.852	33.687	36.191
20	22.775	23.828	25.038	28.412	31.410	34.170	35.020	37.566
21	23.858	24.935	26.171	29.615	32.671	35.479	36.343	38.932
22	24.939	26.039	27.301	30.813	33.924	36.781	37.659	40.289
23	26.018	27.141	28.429	32.007	35.172	38.076	38.968	41.638
24	27.096	28.241	29.553	33.196	36.415	39.364	40.270	42.980
25	28.172	29.339	30.675	34.382	37.652	40.646	41.566	44.314
26	29.246	30.435	31.795	35.563	38.885	41.923	42.856	45.642
27	30.319	31.528	32.912	36.741	40.113	43.195	44.140	46.963
28	31.391	32.620	34.027	37.916	41.337	44.461	45.419	48.278
29	32.461	33.711	35.139	39.087	42.557	45.722	46.693	49.588
30	33.530	34.800	36.250	40.256	43.773	46.979	47.962	50.892
40	44.165	45.616	47.269	51.805	55.758	59.342	60.436	63.691
50	54.723	56.334	58.164	63.167	67.505	71.420	72.613	76.154
60	65.226	66.981	68.972	74.397	79.082	83.298	84.58	88.379

De la consulta de la tabla para un $\alpha = 0.05$ y grados de libertad = 19 el valor de chi cuadrada es de 30.144.

La consulta de este valor puede ser realizado en Excel mediante la función INV.CHICUAD.CD, tal como se aprecia en la siguiente imagen.



	A	B	C	D
1	Nivel de significancia	α	0.05	
2	Tamaño de muestra	n	20	
3	Grados de libertad	v	19	
4				
5	Valor de chi cuadrada	χ^2	30.144	

Ejemplo 1. Control de calidad en el envasado de jugo: Una empresa envasadora de jugo de manzana desea asegurarse de que la variabilidad en el volumen de llenado de sus botellas sea mínima. Se toma una muestra aleatoria de 25 botellas y se mide su contenido, obteniendo una varianza muestral de 4 ml². Estima el intervalo de confianza del 90% para la varianza del volumen de llenado de todas las botellas.

Tamaño de muestra	n	25
Grados de libertad	gl	24
Varianza muestral	s ²	4
Nivel de confianza	NC	0.9
Nivel de significancia	α	0.1
Chi cuadrado izquierdo	$\chi^2_{\alpha/2}$	36.42
Chi cuadrado derecho	$\chi^2_{1-\alpha/2}$	13.85

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}$$

$$\frac{(25-1) * 4}{36.42} \leq \sigma^2 \leq \frac{(25-1) * 4}{13.85}$$

2.64	$\leq \sigma^2 \leq$	6.93
-------------	----------------------	-------------

1.62	$\leq \sigma \leq$	2.63
-------------	--------------------	-------------

A un nivel de confianza del 90 % se tiene que la varianza en el volumen de llenado de las botellas de jugo de manzana oscila de 2.64 a 6.93 mililitros cuadrados.

A un nivel de confianza del 90 % se tiene que la desviación estandar en el volumen de llenado de las botellas de jugo de manzana oscila de 1.62 a 2.63 mililitros.

Recuerde que la desviación estándar es la raíz cuadrada de la varianza, por lo tanto, si se tiene el intervalo de confianza para la varianza, se puede tener el intervalo de confianza para la desviación estándar.

Ver video de solución dando clic en el [siguiente link https://youtu.be/vdE1xsUHmxU](https://youtu.be/vdE1xsUHmxU)

Ejemplo 2. Homogeneidad de la mezcla de ingredientes: Un fabricante de alimentos para mascotas necesita garantizar que la mezcla de ingredientes sea homogénea para asegurar la calidad nutricional del producto. Se toman 15 muestras de la mezcla y se mide el contenido de proteína en cada muestra, obteniendo una varianza muestral de 2.5 (unidades de proteína)². Estima el intervalo de confianza del 95% para la varianza del contenido de proteína en toda la mezcla.

Tamaño de muestra	n	15
Grados de libertad	gl	14
Varianza muestral	s^2	2.5
Nivel de confianza	NC	0.95
Nivel de significancia	α	0.05
Chi cuadrado izquierdo	$\chi^2_{\alpha/2}$	26.12
Chi cuadrado derecho	$\chi^2_{1-\alpha/2}$	5.63

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}$$

$$\frac{(15-1) * 2.5}{26.12} \leq \sigma^2 \leq \frac{(15-1) * 2.5}{5.63}$$

1.34	$\leq \sigma^2 \leq$	6.22
-------------	----------------------	-------------

1.16	$\leq \sigma \leq$	2.49
-------------	--------------------	-------------

A un nivel de confianza del 95 % se tiene que la varianza en el contenido de proteína en el alimento para mascotas oscila de 1.34 a 6.22 unidades de proteína al cuadrado.

A un nivel de confianza del 95 % se tiene que la desviación estándar en el contenido de proteína en el alimento para mascotas oscila de 1.16 a 2.49 unidades de proteína.

Ejercicios propuestos.

Ejercicio 1. Variabilidad en el peso de las galletas: Una empresa galletera busca reducir la variabilidad en el peso de sus galletas para cumplir con los estándares de empaque y evitar pérdidas. Se toman 30 galletas de la línea de producción y se pesan, obteniendo una varianza muestral de 0.2 g². Estima el intervalo de confianza del 99% para la varianza del peso de todas las galletas producidas. *Respuesta* $0.11 \leq \sigma^2 \leq 0.44$ $0.33 \leq \sigma \leq 0.66$

Ejercicio 2. Control de la temperatura en cámaras frigoríficas: Una empresa de almacenamiento de alimentos necesita mantener una temperatura estable en sus cámaras frigoríficas para garantizar la conservación de los productos. Se registran las temperaturas en 15 diferentes puntos de la cámara durante un período de tiempo y se calcula una varianza muestral de 0.5 °C². Estima el intervalo de confianza del 90% para la varianza de la temperatura en toda la cámara. *Respuesta* $0.3 \leq \sigma^2 \leq 1.07$ $0.54 \leq \sigma \leq 1.03$

Ejercicio 3. Variabilidad en el tiempo de cocción de alimentos: Un fabricante de alimentos preparados desea reducir la variabilidad en el tiempo de cocción de sus productos para asegurar una calidad consistente. Se toman 20 muestras del producto y se mide el tiempo de cocción, obteniendo una varianza muestral de 1.2 minutos². Estima el intervalo de confianza del 95% para la varianza del tiempo de cocción de todos los productos. *Respuesta* $0.69 \leq \sigma^2 \leq 2.56$ $0.83 \leq \sigma \leq 1.60$

Ejercicio 4. Control de la calidad de la leche: Una empresa láctea necesita mantener una variabilidad baja en la calidad de la leche que utiliza para sus productos. Se toman 12 muestras de leche de diferentes proveedores y se mide el contenido de grasa en cada muestra, obteniendo una varianza muestral de 0.8 (unidades de grasa)². Estima el intervalo de

confianza del 99% para la varianza del contenido de grasa en la leche de todos los proveedores. *Respuesta* $0.33 \leq \sigma^2 \leq 3.38$ $0.57 \leq \sigma \leq 1.84$

Ejercicio 5. Variabilidad en el tamaño de las frutas: Un productor de frutas desea reducir la variabilidad en el tamaño de sus productos para facilitar el empaque y la comercialización. Se toman 40 frutas de la cosecha y se miden sus diámetros, obteniendo una varianza muestral de 1.5 cm². Estima el intervalo de confianza del 90% para la varianza del diámetro de todas las frutas cosechadas. *Respuesta* $1.07 \leq \sigma^2 \leq 2.28$ $1.04 \leq \sigma \leq 1.51$

Caso 10. Intervalo de confianza para la razón de varianzas.

La fórmula para este caso es.

$$\frac{s_1^2}{s_2^2} \frac{1}{f(v_1, v_2)} \leq \frac{\sigma_1^2}{\sigma_2^2} \leq \frac{s_1^2}{s_2^2} f(v_2, v_1)$$

En esta fórmula se aprecia que son dos poblaciones, cada una de ella con su varianza, su tamaño de muestra y su grado de libertad. Adicional a ello se observa que el estadístico de prueba es la distribución F de Fisher.

La tabla de Fisher y como de consulta se presenta a continuación, asumiendo un nivel de confianza del 90% y dos poblaciones de las cuales se tomaron 10 y 12 elementos.

Los grados de libertad son igual al tamaño de muestra menos 1, por lo tanto $v_1 = 9$ y $v_2 = 11$.

El nivel de significancia α es igual a 10%, que equivale a 0.1. Como la prueba es bilateral, ya que se calcula un límite inferior y uno superior, el nivel de significancia se divide entre 2, quedando así un valor de .05.

Con los datos de grados de libertad y nivel de significancia, se tiene que el valor de F es de 2.9, tal como se puede observar en la tabla F de Fisher.

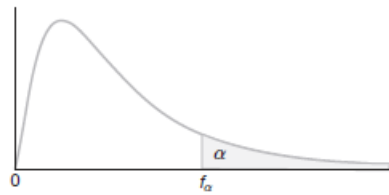


Tabla A.6 Valores críticos de la distribución F

		$f_{0.05}(v_1, v_2)$								
		v_1								
v_2		1	2	3	4	5	6	7	8	9
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.39
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.78
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.97
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.07
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.65
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.36
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.15
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.99
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.87
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.77
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.68
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.62
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.56
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.51
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.46
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.43
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.39
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.36
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.34
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.31
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.29
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.27
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.25
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.24
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.22
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.21
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.19
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.18
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.09
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	2.01
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.93
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.85

Reproducida de la tabla 18 de *Biometrika Tables for Statisticians*, Vol. I, con autorización de E.S. Pearson y Biometrika Trustees.

La obtención del valor de F de Fisher mediante Excel con la función INV.F.CD se presenta en la siguiente imagen, en la cual se puede apreciar que se obtiene el mismo valor.

Estimación de intervalos - Excel

Archivo Inicio Insertar Disposición de página Fórmulas

Normal Ver salt. Diseño Vistas Regla
Pág. de página personalizadas Líneas de cuadrícula Barra de fórmula

Vistas de libro

C10 : X ✓ fx =+INV.F.CD(C8,C4,D4)

	A	B	C	D
1			Población	
2			1	2
3	Tamaño de muestra	n	10	12
4	Grados de libertad	v	9	11
5				
6	Nivel de confianza	NC	0.9	
7	Nivel de significancia	α	0.1	
8		$\alpha/2$	0.05	
9				
10	Valor F de Fisher	F	2.90	

Ejemplo 1. Comparación de la variabilidad en el llenado de botellas de refresco: Una empresa refresquera tiene dos máquinas llenadoras y desea comparar la variabilidad en el volumen de llenado de las botellas. Se toman muestras aleatorias de 20 botellas de la máquina A y 25 botellas de la máquina B. La varianza muestral del volumen de llenado es de 2 ml² para la máquina A y de 1.5 ml² para la máquina B. Estima el intervalo de confianza del 95% para la razón de las varianzas del volumen de llenado entre las dos máquinas.

		Máquina	
		A	B
Tamaño de muestra	n	20	25
Grados de libertad	gl	19	24
Varianza muestral	s ²	2	1.5

Nivel de confianza	NC	0.95
Nivel de significancia	α	0.05
Valor F izquierdo	$f(v_1, v_2)$	2.35
Valor F derecho	$f(v_2, v_1)$	2.45

$$\frac{s_1^2}{s_2^2} \frac{1}{f(v_1, v_2)} \leq \frac{\sigma_1^2}{\sigma_2^2} \leq \frac{s_1^2}{s_2^2} f(v_2, v_1)$$

$$\frac{2}{1.5} * \frac{1}{2.35} \leq \frac{\sigma_1^2}{\sigma_2^2} \leq \frac{2}{1.5} * 2.45$$

0.57	$\leq \frac{\sigma_1^2}{\sigma_2^2} \leq$	3.27
------	---	------

A un nivel de confianza del 95 % se tiene que la razón de varianzas de la máquina A sobre la máquina B en cuanto al nivel de llenado oscila de 0.57 a 3.27 mililitos cuadrados.

Ver video de solución dando clic en el siguiente link <https://youtu.be/kBWv4YE5LNU>

Ejemplo 2. Comparación de la homogeneidad de la mezcla de ingredientes para alimentos para mascotas: Un fabricante de alimentos para mascotas utiliza dos mezcladoras diferentes y quiere comparar la homogeneidad de la mezcla de ingredientes. Se toman 8 muestras aleatorias de la mezcla de cada mezcladora y se mide el contenido de proteína en cada muestra. La varianza muestral del contenido de proteína es de 1.8 (unidades de proteína) ² para la mezcladora 1 y de 2.2 (unidades de proteína) ² para la mezcladora 2. Estima el intervalo de confianza del 90% para la razón de las varianzas del contenido de proteína entre las dos mezcladoras.

		Mezcladora	
		A	B
Tamaño de muestra	n	8	8
Grados de libertad	gl	7	7
Varianza muestral	s ²	1.8	2.2

Nivel de confianza	NC	0.9
Nivel de significancia	α	0.1
Valor F izquierdo	f(v ₁ , v ₂)	3.79
Valor F derecho	f(v ₂ , v ₁)	3.79

$$\frac{s_1^2}{s_2^2} \frac{1}{f(v_1, v_2)} \leq \frac{\sigma_1^2}{\sigma_2^2} \leq \frac{s_1^2}{s_2^2} f(v_2, v_1)$$

$$\frac{1.8}{2.2} * \frac{1}{3.79} \leq \frac{\sigma_1^2}{\sigma_2^2} \leq \frac{1.8}{2.2} * 3.79$$

0.22	$\leq \frac{\sigma_1^2}{\sigma_2^2} \leq$	3.10
------	---	------

A un nivel de confianza del 90 % se tiene que la razón de varianzas del contenido de proteína de la mezcladora A sobre la mezcladora B oscila de 0.22 a 3.10 unidades de proteína al cuadrado.

Ejercicios propuestos.

Ejercicio 1. Comparación de la variabilidad en el peso de las barras de chocolate: Una empresa chocolatera produce barras de chocolate y desea comparar la variabilidad en el peso de las barras producidas por dos líneas de producción diferentes. Se toman muestras aleatorias de 30 barras de cada línea y se pesan. La varianza muestral del peso es de 0.3 g² para la línea 1 y de 0.25 g² para la línea 2. Estima el intervalo de confianza del 99% para la razón de las varianzas del peso entre las dos líneas. *Respuesta* $0.45 \leq \frac{\sigma_1^2}{\sigma_2^2} \leq 3.21$

Ejercicio 2. Comparación de la variabilidad en el tiempo de cocción de alimentos preparados: Un fabricante de alimentos preparados desea comparar la variabilidad en el tiempo de cocción de sus productos utilizando dos métodos de cocción diferentes. Se toman muestras aleatorias de los productos cocinados con cada método y se mide el tiempo de cocción. La

varianza muestral del tiempo de cocción es de 1.5 minutos² para 12 muestras con el método 1 y de 1 minuto² para 15 muestras con el método 2. Estima el intervalo de confianza del 90% para la razón de las varianzas del tiempo de cocción entre los dos métodos.

Respuesta $0.58 \leq \frac{\sigma_1^2}{\sigma_2^2} \leq 4.11$

Ejercicio 3. Comparación de la variabilidad en el contenido de azúcar en dos marcas de yogur: Una empresa láctea produce dos marcas de yogur y desea comparar la variabilidad en el contenido de azúcar entre ambas marcas. Se toman muestras aleatorias de 15 yogures de la marca A y 20 yogures de la marca B. La varianza muestral del contenido de azúcar es de 0.8 g² para la marca A y de 1.2 g² para la marca B. Estima el intervalo de confianza del 95% para la razón de las varianzas del contenido de azúcar entre las dos marcas.

Respuesta $0.25 \leq \frac{\sigma_A^2}{\sigma_B^2} \leq 1.91$

Ejercicio 4. Comparación de la variabilidad en el peso de los paquetes de papas fritas de dos proveedores: Una empresa de snacks utiliza dos proveedores diferentes de papas y desea comparar la variabilidad en el peso de los paquetes de papas fritas de cada proveedor. Se toman muestras aleatorias de 25 paquetes de papas fritas del proveedor 1 y 30 paquetes del proveedor 2. La varianza muestral del peso es de 2 g² para el proveedor 1 y de 1.5 g² para el proveedor 2. Estima el intervalo de confianza del 90% para la razón de las varianzas del peso entre los dos proveedores.

Respuesta $0.70 \leq \frac{\sigma_1^2}{\sigma_2^2} \leq 2.59$

Ejercicio 5. Comparación de la variabilidad en la calidad de los huevos de dos granjas: Una empresa de alimentos utiliza huevos de dos granjas diferentes y necesita comparar la variabilidad en la calidad de los huevos de cada granja. Se toman muestras aleatorias de 40 huevos de cada granja y se mide el peso y la frescura de cada huevo. La varianza muestral del índice de frescura es de 0.5 (unidades de frescura)² para la granja 1 y de 0.8 (unidades de frescura)² para la granja 2. Estima el intervalo de confianza del 99% para la razón de las varianzas del índice de frescura entre las dos granjas.

Respuesta $0.27 \leq \frac{\sigma_1^2}{\sigma_2^2} \leq 1.45$

4.6. Pruebas de Hipótesis.

Pruebas de Hipótesis en la Industria Alimentaria

La industria alimentaria es un sector crítico que impacta directamente la salud y el bienestar de las personas. En este contexto, las pruebas de hipótesis se erigen como herramientas fundamentales para garantizar la calidad, la seguridad y la eficiencia en la producción de alimentos. A través de métodos estadísticos rigurosos, las empresas pueden tomar decisiones informadas y optimizar sus procesos.

¿Qué son las Pruebas de Hipótesis?

Las pruebas de hipótesis son procedimientos estadísticos que permiten evaluar la validez de una afirmación o hipótesis sobre una población, utilizando datos de una muestra. En la industria alimentaria, estas pruebas se aplican en diversas áreas, desde el control de calidad de los productos hasta la evaluación de nuevas tecnologías y procesos.

Aplicaciones en la Industria Alimentaria

1. Control de Calidad: Las pruebas de hipótesis son esenciales para verificar que los productos cumplen con los estándares de calidad establecidos. Por ejemplo, se pueden utilizar para determinar si el contenido promedio de un ingrediente en un lote de alimentos se encuentra dentro de los límites aceptables o si existe alguna diferencia significativa en la calidad de un producto entre diferentes proveedores.

2. Seguridad Alimentaria: Estas pruebas también juegan un papel crucial en la identificación y prevención de riesgos para la seguridad alimentaria. Se pueden emplear para detectar la presencia de contaminantes en alimentos, evaluar la eficacia de los procesos de cocción para eliminar bacterias dañinas o determinar si un nuevo ingrediente o proceso es seguro para el consumo.

3. Desarrollo de Nuevos Productos: Las pruebas de hipótesis son valiosas en el desarrollo de nuevos productos alimenticios. Permiten a los investigadores comparar diferentes formulaciones, evaluar la aceptación de los consumidores y determinar si un nuevo producto es superior a los existentes en términos de sabor, textura u otros atributos relevantes.

4. Optimización de Procesos: Las empresas pueden utilizar pruebas de hipótesis para optimizar sus procesos de producción. Por ejemplo, se pueden comparar diferentes métodos de procesamiento para identificar el que produce alimentos de mejor calidad o el que reduce los costos de producción.

Ejemplos Concretos

Ejemplo 1: Una empresa productora de leche desea determinar si un nuevo método de pasteurización es más efectivo para reducir la carga bacteriana en comparación con el método tradicional. Se toman muestras de leche pasteurizada con ambos métodos y se realiza una prueba de hipótesis para comparar las medias de las cargas bacterianas.

Ejemplo 2: Un fabricante de snacks quiere evaluar si un nuevo sabor es más popular entre los consumidores que el sabor original. Se realiza una encuesta y se utiliza una prueba de hipótesis para determinar si existe una diferencia significativa en la preferencia de los consumidores entre ambos sabores.

Importancia y Beneficios

El uso de pruebas de hipótesis en la industria alimentaria ofrece numerosos beneficios:

- **Toma de Decisiones Informadas:** Permiten a las empresas tomar decisiones basadas en datos y evidencia sólida, en lugar de intuiciones o conjeturas.
- **Mejora Continua:** Facilitan la identificación de áreas de mejora en los procesos productivos y la evaluación de nuevas tecnologías y métodos.
- **Garantía de Calidad y Seguridad:** Contribuyen a asegurar que los productos alimenticios cumplan con los más altos estándares de calidad y seguridad.
- **Eficiencia y Reducción de Costos:** Ayudan a optimizar los procesos y a reducir los costos asociados a la producción.

En conclusión, se tiene que las pruebas de hipótesis son herramientas poderosas que permiten a la industria alimentaria tomar decisiones informadas, mejorar la calidad y la seguridad de los productos, y optimizar los procesos de producción. Su aplicación adecuada es esencial para garantizar la competitividad y el éxito en un mercado cada vez más exigente.

Tipos de Errores en Pruebas de Hipótesis.

Las pruebas de hipótesis son herramientas fundamentales en la investigación científica y la toma de decisiones. Sin embargo, como todo proceso inferencial, están sujetas a la posibilidad de cometer errores. Comprender y gestionar estos errores es crucial para garantizar la validez y confiabilidad de los resultados obtenidos.

El Error de Tipo I.

El error de Tipo I, también conocido como falso positivo, ocurre cuando se rechaza la hipótesis nula (H_0) cuando en realidad es verdadera. En otras palabras, afirmamos que existe un efecto o diferencia significativa cuando en realidad no la hay. La probabilidad de cometer un error de Tipo I se denota por alfa (α) y generalmente se fija en un valor pequeño, como 0.05 o 0.01. Este valor representa el nivel de significancia de la prueba.

Las consecuencias de cometer un error de Tipo I pueden ser graves, especialmente en campos como la medicina o la industria alimentaria. Por ejemplo, si en un ensayo clínico se concluye erróneamente que un nuevo fármaco es efectivo, esto podría llevar a su aprobación y comercialización, poniendo en riesgo la salud de los pacientes.

El Error de Tipo II.

El error de Tipo II, también conocido como falso negativo, ocurre cuando no se rechaza la hipótesis nula (H_0) cuando en realidad es falsa. En otras palabras, no se detecta un efecto o diferencia significativa que realmente existe. La probabilidad de cometer un error de Tipo II se denota por beta (β).

Las consecuencias de cometer un error de Tipo II pueden ser igualmente problemáticas. Por ejemplo, si en un estudio no se detecta un efecto adverso de un producto químico, esto podría llevar a su uso continuado, causando daños a largo plazo.

Relación entre los Errores de Tipo I y Tipo II

Existe una relación inversa entre los errores de Tipo I y Tipo II. Al disminuir la probabilidad de cometer un error de Tipo I (α), generalmente aumenta la probabilidad de cometer un error de Tipo II (β), y viceversa. Esto se debe a que al ser más estrictos en la aceptación de resultados (disminuyendo α), es más difícil detectar efectos pequeños pero reales (aumentando β).

Estrategias para Minimizar los Errores

Existen diversas estrategias para minimizar los errores en las pruebas de hipótesis:

1. Aumentar el tamaño de la muestra: Un tamaño de muestra mayor proporciona más información y reduce la variabilidad de los resultados, lo que disminuye la probabilidad de cometer ambos tipos de errores (Field, 2013).

2. Controlar el nivel de significancia (α): Ajustar el nivel de significancia (α) permite controlar la probabilidad de cometer un error de Tipo I. Sin embargo, es importante recordar que disminuir α aumenta la probabilidad de cometer un error de Tipo II.

3. Aumentar la potencia de la prueba ($1-\beta$): La potencia de una prueba es la probabilidad de detectar un efecto real cuando existe. Aumentar la potencia de la prueba reduce la probabilidad de cometer un error de Tipo II. Esto se puede lograr aumentando el tamaño de la muestra, utilizando pruebas estadísticas más potentes o mejorando la precisión de las mediciones (Howell, 2017).

4. Considerar el contexto y las consecuencias: Al interpretar los resultados de una prueba de hipótesis, es fundamental considerar el contexto y las posibles consecuencias de cometer cada tipo de error. En situaciones donde las consecuencias de un falso positivo son graves, puede ser preferible disminuir α , incluso si esto aumenta la probabilidad de un falso negativo.

En conclusión, se tiene que los errores de Tipo I y Tipo II son inherentes a las pruebas de hipótesis. Comprender su naturaleza, sus implicaciones y las estrategias para minimizarlos es esencial para garantizar la validez y confiabilidad de los resultados obtenidos. Al tomar decisiones basadas en pruebas de hipótesis, es fundamental considerar el equilibrio entre ambos tipos de errores y elegir el enfoque que mejor se adapte al contexto y a las consecuencias de cada posible resultado.

Pasos del Procedimiento de Prueba de Hipótesis

El procedimiento de prueba de hipótesis consta de varios pasos bien definidos:

1.- Formulación de Hipótesis:

Hipótesis Nula (H_0): Es la afirmación que se busca refutar. Generalmente, es una afirmación de no efecto, no diferencia o no relación.

Hipótesis Alternativa (H_1): Es la afirmación que se acepta si se rechaza la hipótesis nula. Es la afirmación de un efecto, diferencia o relación.

2.- Selección del Nivel de Significancia (α):

El nivel de significancia (α) es la probabilidad de rechazar la hipótesis nula cuando es verdadera (error tipo I).

Generalmente, se utiliza un nivel de significancia de 0.05 (5%), lo que significa que hay una probabilidad del 5% de cometer un error de tipo I.

3.- Selección del Estadístico de Prueba:

El estadístico de prueba es una medida que se calcula a partir de los datos de la muestra y se utiliza para evaluar la evidencia en contra de la hipótesis nula.

La elección del estadístico de prueba depende del tipo de datos y de la hipótesis que se está probando.

4.- Establecimiento de la Región Crítica:

La región crítica es el conjunto de valores del estadístico de prueba para los cuales se rechaza la hipótesis nula.

La región crítica se determina en función del nivel de significancia y de la distribución del estadístico de prueba.

5.- Cálculo del Estadístico de Prueba:

Se calcula el valor del estadístico de prueba utilizando los datos de la muestra.

6.- Toma de Decisión:

Se compara el valor del estadístico de prueba con la región crítica.

Si el valor del estadístico de prueba cae dentro de la región crítica, se rechaza la hipótesis nula.

Si el valor del estadístico de prueba no cae dentro de la región crítica, no se rechaza la hipótesis nula.

7.- Conclusión:

Se interpreta la decisión en el contexto del problema de investigación.

Se establece si hay suficiente evidencia para apoyar la hipótesis alternativa o si no hay suficiente evidencia para rechazar la hipótesis nula.

Ejemplos de Hipótesis Nula y Alternativa en la Industria Alimentaria

Prueba Unilateral por la Izquierda

Contexto: Una empresa de alimentos para bebés está desarrollando una nueva fórmula infantil y quiere asegurarse de que el contenido promedio de sodio por porción sea menor a 150 mg, el límite recomendado por las autoridades sanitarias.

- Hipótesis Nula (H_0): El contenido promedio de sodio en la nueva fórmula infantil es mayor o igual a 150 mg por porción ($\mu \geq 150$).
- Hipótesis Alternativa (H_1): El contenido promedio de sodio en la nueva fórmula infantil es menor a 150 mg por porción ($\mu < 150$).

Explicación: En este caso, la empresa está interesada en demostrar que el contenido de sodio es menor al límite recomendado. Por lo tanto, la hipótesis alternativa se formula con un signo "menor que" ($<$), lo que indica una prueba unilateral por la izquierda.

Prueba Bilateral

Contexto: Un fabricante de pan está evaluando un nuevo proceso de horneado para reducir la variabilidad en el peso de sus hogazas. El peso ideal de cada hogaza es de 400 g.

- Hipótesis Nula (H_0): El peso promedio de las hogazas con el nuevo proceso es igual a 400 g ($\mu = 400$).
- Hipótesis Alternativa (H_1): El peso promedio de las hogazas con el nuevo proceso es diferente de 400 g ($\mu \neq 400$).

Explicación: En este caso, el fabricante está interesado en determinar si el nuevo proceso produce un cambio en el peso promedio, ya sea un aumento o una disminución. Por lo tanto, la hipótesis alternativa se formula con un signo "diferente de" ($\neq$), lo que indica una prueba bilateral.

Prueba Unilateral por la Derecha

Contexto: Una empresa productora de jugos de fruta quiere aumentar la concentración de vitamina C en su producto. Actualmente, la concentración promedio es de 60 mg por cada 100 ml.

- Hipótesis Nula (H_0): La concentración promedio de vitamina C en el jugo es menor o igual a 60 mg por cada 100 ml ($\mu \leq 60$).
- Hipótesis Alternativa (H_1): La concentración promedio de vitamina C en el jugo es mayor a 60 mg por cada 100 ml ($\mu > 60$).

Explicación: En este caso, la empresa está interesada en demostrar que la concentración de vitamina C es mayor al valor actual. Por lo tanto, la hipótesis alternativa se formula con un signo "mayor que" ($>$), lo que indica una prueba unilateral por la derecha.

De los ejemplos anteriores de hipótesis se puede apreciar que:

- la hipótesis nula y la alternativa deben ser mutuamente excluyentes, lo que significa que no pueden ser verdaderas al mismo tiempo. También deben ser exhaustivas, lo que significa que cubren todas las posibles situaciones.
- Lo que se desea comprobar se encuentra planteado en la hipótesis alternativa (H_1).
- El signo de igual siempre se encuentra en la hipótesis nula (H_0).

Caso 1. Prueba de hipótesis para la media con desviación estándar poblacional conocida.

La fórmula para este caso es.

$$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

Ejemplo 1. Control de calidad en el envasado de cereales. Una empresa de cereales afirma que el peso promedio de sus cajas es de 500 gramos. Se toma una muestra de 25 cajas y se encuentra que el peso promedio es de 495 gramos. La desviación estándar poblacional es de 10 gramos. ¿Hay evidencia suficiente para dudar de la afirmación de la empresa? Utiliza un nivel de significancia del 5%.

1. Formulación de hipótesis.

$$H_0 \mu = 500$$

$$H_1 \mu \neq 500$$

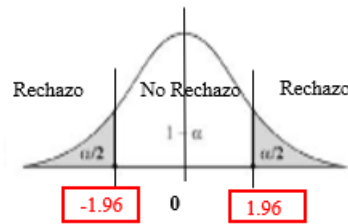
2. Nivel de significancia. $\alpha = 0.05$

3. Estadístico de prueba. $Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$

4. Establecimiento de la región crítica.

Nivel de significancia (α) 0.05

Z critico 1.96



5. Cálculo del estadístico de prueba

Media hipotética	μ	500
Media muestral	$\bar{x}$	495
Desv. Std. Poblacional	σ	10
Tamaño de muestra	n	25

$$Z = \frac{495 - 500}{\frac{10}{\sqrt{25}}} = -2.5$$

Z calculada	-2.5
-------------	------

6. Toma de decisión.

Como el valor de Z calculado es menor al valor crítico por la izquierda, la decisión es rechazar la hipótesis nula

7. Conclusión.

A un nivel de significancia de 0.05 se puede afirmar que el contenido promedio de las cajas de cereales es diferente a 500 gramos.

Ver video de solución en Excel dando clic en el siguiente link https://youtu.be/78hCV8_qe5M

Ejemplo 2. Una empresa que produce barras de granola afirma que cada barra contiene un promedio de 20 gramos de proteína. Se sabe que la desviación estándar de la cantidad de proteína por barra es de 3 gramos. Un grupo de consumidores sospecha que las barras contienen menos proteína de la que afirma la empresa. Para probar esta sospecha, se toma una muestra aleatoria de 40 barras y se encuentra que su contenido promedio de proteína es de 19 gramos. ¿Hay suficiente evidencia para apoyar la sospecha del grupo de consumidores de que las barras de granola contienen menos de 20 gramos de proteína?

1. Formulación de hipótesis.

$$H_0 \mu \geq 20$$

$$H_1 \mu < 20$$

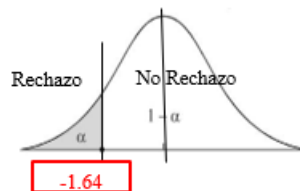
2. Nivel de significancia. $\alpha = 0.05$

3. Estadístico de prueba. $Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$

4. Establecimiento de la región crítica.

Nivel de significancia (α) 0.05

Z crítico -1.64



5. Cálculo del estadístico de prueba

Media hipotética	μ	20
Media muestral	$\bar{x}$	19
Desv. Std. Poblacional	σ	3
Tamaño de muestra	n	40

$$Z = \frac{19 - 20}{\frac{3}{\sqrt{40}}} = -2.11$$

Z calculada	-2.11
-------------	-------

6. Toma de decisión.

Como el valor de Z calculado es menor al valor crítico, la decisión es rechazar la hipótesis nula

7. Conclusión.

A un nivel de significancia de 0.05 se puede afirmar que el contenido promedio de proteína en las barras de granola es menor a 20 gramos.

Ver video de solución en Excel dando clic en el siguiente link <https://youtu.be/ZJqK-TdtK6o>

Ejemplo 3. Una empresa que produce jugos de fruta envasados afirma que cada botella contiene un promedio de 300 ml de jugo. Se sabe que la desviación estándar de la cantidad de jugo por botella es de 5 ml. Un grupo de consumidores sospecha que las botellas contienen más jugo del que afirma la empresa. Para probar esta sospecha, se toma una muestra aleatoria de 35 botellas y se encuentra que su contenido promedio de jugo es de 302 ml. ¿Para un nivel de significancia de 0.01, hay suficiente evidencia para apoyar la sospecha del grupo de consumidores de que las botellas de jugo contienen más de 300 ml?

1. Formulación de hipótesis.

$$H_0 \mu \leq 300$$

$$H_1 \mu > 300$$

2. Nivel de significancia. $\alpha = 0.01$

3. Estadístico de prueba. $Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$

4. Establecimiento de la región crítica.

Nivel de significancia (α) 0.01
Z crítico 2.33



5. Cálculo del estadístico de prueba

Media hipotética	μ	300
Media muestral	$\bar{x}$	302
Desv. Std. Poblacional	σ	5
Tamaño de muestra	n	30

$$Z = \frac{302 - 300}{\frac{5}{\sqrt{30}}} = 2.19$$

Z calculada	2.19
-------------	------

6. Toma de decisión.

Como el valor de Z calculado es menor al valor crítico, la decisión es no rechazar la hipótesis nula

7. Conclusión.

A un nivel de significancia de 0.01 al parecer el contenido promedio de las botellas de jugo de fruta envasado es igual a 300 mililitros.

Ver video de solución en Excel dando clic en el siguiente link <https://youtu.be/-TWgxusKcR0>

En los ejercicios anteriores se puede observar lo siguiente:

- Existen pruebas bilaterales, pruebas unilaterales por la izquierda y pruebas unilaterales por la derecha.
- En la hipótesis nula siempre va el signo de igual y en la hipótesis alternativa se expresa lo que se desea comprobar.
- De ubicar el valor calculado en el diagrama que contiene el o los valores críticos se observa si este cae en zona de rechazo o no rechazo de la hipótesis nula.

- Si se rechaza la hipótesis nula, significa que se acepta la hipótesis alternativa.
- En la conclusión se dice “al parecer” en los casos en que la hipótesis nula no se rechazó, y para los casos en que se rechaza la hipótesis nula se dice que “se puede afirmar”.

Ejercicios propuestos.

Ejercicio 1. Nivel de azúcar en bebidas deportivas. Un fabricante de bebidas deportivas afirma que su producto tiene un contenido promedio de azúcar de 20 gramos por porción. Se analizan 32 porciones y se encuentra que el contenido promedio es de 22 gramos. Considerando que la desviación estándar poblacional es de 3 gramos. ¿Hay evidencia suficiente para apoyar la afirmación del fabricante? Utiliza un nivel de significancia del 1%. $H_0 = 20$ $H_1 \neq 20$, $Z_{calc.} = 3.7712$, *Rechazar H_0*

Ejercicio 2. Calidad de la leche. Una cooperativa lechera afirma que la leche que produce tiene un contenido promedio de grasa del 3.5%. Se analizan 40 muestras de leche y se encuentra que el contenido promedio de grasa es del 3.2%. La desviación estándar poblacional es de 0.3%. ¿Hay evidencia suficiente para dudar de la afirmación de la cooperativa? Utiliza un nivel de significancia del 5%. $H_0 = 20$ $H_1 < 20$, $Z_{calc.} = -6.32$, *Rechazar H_0*

Ejercicio 3. Una fábrica de galletas asegura que el peso promedio de cada galleta es de 15 gramos. La desviación estándar poblacional del peso es de 3 gramos. Un grupo de control de calidad sospecha que el peso real de las galletas es superior al declarado. Para verificar esta sospecha, se toma una muestra aleatoria de 32 galletas y se encuentra que su peso promedio es de 16 gramos. ¿Hay suficiente evidencia para apoyar la sospecha del grupo de control de calidad de que las galletas pesan más de 15 gramos? Utiliza un nivel de significancia del 4%. $H_0 = 15$ $H_1 > 15$, $Z_{calc.} = 1.8856$, *Rechazar H_0*

Ejercicio 4. Una empresa productora de leche asegura que cada envase contiene un promedio de 1 litro de leche. La desviación estándar de la cantidad de leche por envase es de 0,02 litros. Un grupo de consumidores sospecha que los envases contienen menos leche de la declarada. Para probar esta sospecha, se toma una muestra aleatoria de 20 envases y se encuentra que su contenido promedio de leche es de 0,99 litros. ¿Hay suficiente evidencia para apoyar la sospecha del grupo de consumidores de que los envases de leche contienen menos de 1 litro? Considere que el volumen se distribuye de forma normal y utilice un nivel de significancia del 1%. $H_0 = 1$ $H_1 < 1$, $Z_{calc.} = -2.2361$, *No Rechazar H_0*

Ejemplo 5. Una empresa que produce mantequilla de maní afirma que cada frasco contiene un promedio de 10 gramos de grasa saturada por cada 100 gramos de producto. Se sabe que la desviación estándar poblacional de la cantidad de grasa saturada es de 2 gramos. Un grupo de nutricionistas sospecha que los frascos contienen más grasa saturada de la declarada. Para probar esta sospecha, se toma una muestra aleatoria de 30 frascos y se encuentra que su contenido promedio de grasa saturada es de 10.8 gramos por cada 100 gramos de producto.

¿Hay suficiente evidencia para apoyar la sospecha del grupo de nutricionistas de que los frascos de mantequilla de maní contienen más de 10 gramos de grasa saturada por cada 100 gramos de producto? Utiliza un nivel de significancia del 5%. $H_0 = 10$ $H_1 > 10$, $Z_{calc.} = 2.1919$, *Rechazar H_0*

Caso 2. Prueba de hipótesis para la media con desviación estándar poblacional desconocida.

La fórmula para este caso es.

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

$$gl = n - 1$$

Ejemplo 1. Una panadería afirma que sus hogazas de pan integral pesan un promedio de 500 gramos. Un grupo de consumidores sospecha que las hogazas pesan menos de lo declarado. Para probar esta sospecha, se toma una muestra aleatoria de 20 hogazas y se encuentra que su peso promedio es de 490 gramos con una desviación estándar de 15 gramos. ¿Hay suficiente evidencia para apoyar la sospecha del grupo de consumidores de que las hogazas de pan integral pesan menos de 500 gramos? Considere un nivel de significancia de 0.01.

1. Formulación de hipótesis.

$$H_0: \mu = 500$$

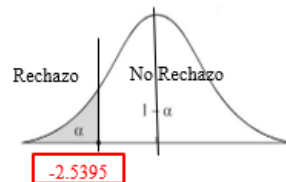
$$H_1: \mu < 500$$

2. Nivel de significancia. $\alpha = 0.01$

3. Estadístico de prueba. $t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$

4. Establecimiento de la región crítica.

Nivel de significancia (α)	0.01
Tamaño de muestra (n)	20
grados de libertad (gl)	19
t crítico	-2.5395



5. Cálculo del estadístico de prueba

Media hipotética	μ	500
Media muestral	$\bar{x}$	490
Desv. Std. Muestral	s	15
Tamaño de muestra	n	20

$$t = \frac{490 - 500}{\frac{15}{\sqrt{20}}} = -2.9814$$

t calculada	-2.9814
-------------	---------

6. Toma de decisión.

Como el valor de t calculado es menor al valor crítico por la izquierda, la decisión es rechazar la hipótesis nula

7. Conclusión.

A un nivel de significancia de 0.01 se puede afirmar que el peso promedio de las hogazas de pan es menor a 500 gramos.

Ver video de solución en Excel dando clic en el siguiente link <https://youtu.be/0JSh3X3heic>

Ejemplo 2. Una empresa que produce latas de atún afirma que cada lata contiene un promedio de 150 gramos de atún. Un grupo de consumidores sospecha que las latas no contienen la cantidad de atún declarada. Para probar esta sospecha, se toma una muestra aleatoria de 25 latas y se encuentra que su contenido promedio de atún es de 148 gramos con una desviación estándar de 5 gramos. ¿Hay suficiente evidencia para apoyar la sospecha del grupo de consumidores de que las latas de atún no contienen 150 gramos de atún? Considere un nivel de significancia de 0.01.

1. Formulación de hipótesis.

$$H_0 \mu = 150$$

$$H_1 \mu \neq 150$$

2. Nivel de significancia. $\alpha = 0.01$

3. Estadístico de prueba.
$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

4. Establecimiento de la región crítica.

Nivel de significancia (α)	0.01
Tamaño de muestra (n)	25
grados de libertad (gl)	24
t crítico	2.7969



5. Cálculo del estadístico de prueba

Media hipotética	μ	150
Media muestral	$\bar{x}$	148
Desv. Std. Muestral	s	5
Tamaño de muestra	n	25

$$t = \frac{148 - 150}{\frac{5}{\sqrt{25}}} = -2.0$$

t calculada	-2.0000
-------------	---------

6. Toma de decisión.

Como el valor de t calculado se encuentra entre los valores críticos, la decisión es no rechazar la hipótesis nula

7. Conclusión.

A un nivel de significancia de 0.01 al parecer el peso promedio de las latas de atún es igual a 150 gramos.

Ver video de solución en Excel dando clic en el siguiente link <https://youtu.be/l8DQHzQfmKc>

Ejemplo 3. Una empresa que produce bolsas de papas fritas afirma que cada bolsa contiene un promedio de 180 gramos de papas. Un grupo de consumidores sospecha que las bolsas contienen más papas de lo declarado. Para probar esta sospecha, se toma una muestra aleatoria de 27 bolsas y se encuentra que su contenido promedio de papas es de 183 gramos con una desviación estándar de 8 gramos. ¿Hay suficiente evidencia para apoyar la sospecha del grupo de consumidores de que las bolsas de papas fritas contienen más de 180 gramos de papas? Considere un nivel de significancia de 0.04.

1. Formulación de hipótesis.

$$H_0: \mu = 180$$

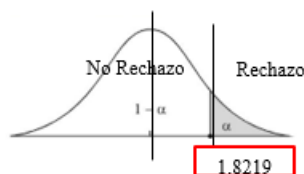
$$H_1: \mu > 180$$

2. Nivel de significancia. $\alpha = 0.04$

3. Estadístico de prueba.
$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

4. Establecimiento de la región crítica.

Nivel de significancia (α)	0.04
Tamaño de muestra (n)	27
grados de libertad (gl)	26
t crítico	1.8219



5. Cálculo del estadístico de prueba

Media hipotética	μ	180
Media muestral	$\bar{x}$	183
Desv. Std. Muestral	s	8
Tamaño de muestra	n	27

$$t = \frac{183 - 180}{\frac{8}{\sqrt{27}}} = 1.9486$$

t calculada	1.9486
-------------	--------

6. Toma de decisión.

Como el valor de t calculado es mayor al valor crítico, la decisión es rechazar la hipótesis nula

7. Conclusión.

A un nivel de significancia de 0.01 se puede afirmar que el peso promedio de las bolsas de papas fritas es mayor a 180 gramos

Ver video de solución en Excel dando clic en el siguiente link
<https://youtu.be/eCWAqHIN4zo>

Ejercicios propuestos.

Ejercicio 1. Una empresa que produce botellas de agua mineral afirma que cada botella contiene un promedio de 500 ml de agua. Un grupo de consumidores sospecha que las botellas contienen menos agua de lo declarado. Para probar esta sospecha, se toma una muestra aleatoria de 25 botellas y se encuentra que su contenido promedio de agua es de 498 ml con una desviación estándar de 8 ml. ¿Hay suficiente evidencia para apoyar la sospecha del grupo de consumidores de que las botellas de agua mineral contienen menos de 500 ml? Considere un nivel de significancia de 0.02. $H_0 \mu = 500$ $H_1 \mu < 500$, $t_{calc.} = -1.25$, *No Rechazar H_0*

Ejercicio 2. Una fábrica de chocolate afirma que el peso promedio de cada tableta de chocolate es de 100 gramos. Un grupo de control de calidad sospecha que el peso real de las tabletas de chocolate es diferente al declarado. Para verificar esta sospecha, se toma una muestra aleatoria de 20 tabletas y se encuentra que su peso promedio es de 98 gramos con una desviación estándar de 4 gramos. ¿Hay suficiente evidencia para apoyar la sospecha del grupo de control de calidad de que las tabletas de chocolate no pesan 100 gramos? Considere un nivel de significancia de 0.05. $H_0 \mu = 100$ $H_1 \mu \neq 100$, $t_{calc.} = -2.2361$, *Rechazar H_0*

Ejercicio 3. Una empresa que produce bolsas de café en grano afirma que cada bolsa contiene un promedio de 250 gramos de café. Un grupo de consumidores sospecha que las bolsas contienen más café de lo declarado. Para probar esta sospecha, se toma una muestra aleatoria de 18 bolsas y se encuentra que su contenido promedio de café es de 255 gramos con una desviación estándar de 10 gramos. ¿Hay suficiente evidencia para apoyar la sospecha del grupo de consumidores de que las bolsas de café en grano contienen más de 250 gramos? Considere un nivel de significancia de 0.06. $H_0 \mu = 250$ $H_1 \mu > 250$, $t_{calc.} = 2.1213$, *Rechazar H_0*

Ejercicio 4. Una empresa que produce latas de conservas de duraznos afirma que cada lata contiene un promedio de 400 gramos de duraznos. Un grupo de consumidores sospecha que las latas contienen menos duraznos de lo declarado. Para probar esta sospecha, se toma una muestra aleatoria de 28 latas y se encuentra que su contenido promedio de duraznos es de 395 gramos con una desviación estándar de 12 gramos. ¿Hay suficiente evidencia para apoyar la sospecha del grupo de consumidores de que las latas de conservas de duraznos contienen menos de 400 gramos? Considere un nivel de significancia de 0.02. $H_0 \mu = 400$ $H_1 \mu < 400$, $t_{calc.} = -2.0352$, *No Rechazar H_0*

Ejercicio 5. Una fábrica de mermelada afirma que cada frasco de mermelada contiene un promedio de 350 gramos de mermelada. Un grupo de control de calidad sospecha que el peso real de los frascos de mermelada es diferente al declarado. Para verificar esta sospecha, se toma una muestra aleatoria de 17 frascos y se encuentra que su peso promedio es de 355 gramos con una desviación estándar de 8 gramos. ¿Hay suficiente evidencia para apoyar la

sospecha del grupo de control de calidad de que los frascos de mermelada no pesan 350 gramos? Considere un nivel de significancia de 0.05. $H_0 \mu = 350$ $H_1 \mu \neq 350$, $t_{calc.} = 2.5769$, *Rechazar H_0*

Caso 3. Prueba de hipótesis para la diferencia de medias con desviaciones estándar poblacional conocidas.

Para este caso la fórmula es.

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Ejemplo 1. Dos empresas producen jugo de naranja en botellas de 500 mililitros. La empresa A afirma que el volumen de sus botellas de jugo tiene una desviación estándar de 5 ml, mientras que la empresa B afirma que el volumen de sus botellas tiene una desviación estándar de 6 ml. Un grupo de consumidores sospecha que la empresa A en realidad vende botellas con un contenido promedio menor al de la empresa B. Para probar esta sospecha, se toman muestras aleatorias de 30 botellas de cada empresa y se encuentra que el contenido promedio de la empresa A es de 498 ml y el de la empresa B es de 501 ml. ¿Hay suficiente evidencia para apoyar la sospecha del grupo de consumidores de que las botellas de jugo de la empresa A contienen menos jugo que las de la empresa B? Considere un nivel de significancia de 0.05.

1. Formulación de hipótesis.

$$H_0 \mu_A - \mu_B \geq 0$$

$$H_1 \mu_A - \mu_B < 0$$

2. Nivel de significancia. $\alpha = 0.05$

3. Estadístico de prueba.

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

4. Establecimiento de la región crítica.

Nivel de significancia (α)	0.05
Z crítico	-1.64



5. Cálculo del estadístico de prueba

		Empresa	
		A	B
Media muestral	$\bar{x}$	498	501
Desv. Std. Poblacional	σ	5	6
Varianza Poblacional	σ^2	25	36
Tamaño de muestra	n	30	30

Dif. Medias Muestrales	$\bar{x}_1 - \bar{x}_2$	-3
Valor calculado de A	$Z_{calc.}$	-2.1039

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{-10 - 0}{\sqrt{\frac{25}{30} + \frac{36}{30}}} = -2.1039$$

6. Toma de decisión.

Como el valor de Z calculado es menor al valor crítico, la decisión es rechazar la hipótesis nula

7. Conclusión.

A un nivel de significancia de 0.05 se puede afirmar que el volumen promedio de las botellas de jugo de naranja de la empresa A es menor que el de la empresa B.

Ver video de solución en Excel dando clic en el siguiente link <https://youtu.be/bMjzksCnjKs>

Ejercicio 2. Dos empresas producen barras de cereal. La empresa A afirma que sus barras de cereal, en relación a su contenido calórico, tienen una desviación estándar de 10 calorías, mientras que la empresa B afirma que sus barras de cereal tienen una desviación estándar de 12 calorías. Un grupo de consumidores sospecha que el contenido calórico de las barras de la empresa A es diferente al de la empresa B. Para probar esta sospecha, se toman muestras aleatorias de 40 barras de cada empresa y se encuentra que el contenido promedio de calorías

de la empresa A es de 203 calorías y el de la empresa B es de 198 calorías. ¿Hay suficiente evidencia para apoyar la sospecha del grupo de consumidores de que el contenido calórico de las barras de cereal de la empresa A es diferente al de la empresa B? Considere un nivel de significancia de 0.02.

1. Formulación de hipótesis.

$$H_0: \mu_A - \mu_B = 0$$

$$H_1: \mu_A - \mu_B \neq 0$$

2. Nivel de significancia. $\alpha = 0.02$

3. Estadístico de prueba.

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

4. Establecimiento de la región crítica.

Nivel de significancia (α)	0.02
Z crítico	2.33



5. Cálculo del estadístico de prueba

		Empresa	
		A	B
Media muestral	$\bar{x}$	203	198
Desv. Std. Poblacional	σ	10	12
Varianza Poblacional	σ^2	100	144
Tamaño de muestra	n	40	40

Dif. Medias Muestrales	$\bar{x}_1 - \bar{x}_2$	5
Valor calculado de A	$Z_{calc.}$	2.0244

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{5 - 0}{\sqrt{\frac{100}{40} + \frac{144}{40}}} = 2.0244$$

6. Toma de decisión.

Como el valor de Z calculado se encuentra entre los valores críticos, la decisión es no rechazar la hipótesis nula

7. Conclusión.

A un nivel de significancia de 0.02 al parecer el contenido calórico de las barras de cereal de ambas empresas es igual.

Ver video de solución en Excel dando clic en el siguiente link <https://youtu.be/lf08enzAS10>

Ejercicio 3. Dos empresas, "A" y "B", producen salsa de tomate. La empresa "A" afirma que su salsa tiene un sabor más intenso, medido en unidades de sabor, que la salsa de la empresa "B". Se sabe que la desviación estándar poblacional de las unidades de sabor para la salsa "A" es de 3 unidades, y para la salsa "B" es de 2 unidades. Para verificar esta afirmación, se toman muestras aleatorias de 25 botellas de salsa de tomate de la empresa A y 32 botellas de salsa de tomate de la empresa B. Los resultados muestran que la salsa "A" tiene una media de 28 unidades de sabor, mientras que la salsa "B" tiene una media de 25 unidades. ¿Hay

suficiente evidencia para apoyar la afirmación de la empresa A en cuanto a que su salsa de tomate tiene un sabor más intenso en que la salsa de tomate de la empresa B? Considere un nivel de significancia de 0.04.

1. Formulación de hipótesis.

$$H_0: \mu_A - \mu_B \leq 0$$

$$H_1: \mu_A - \mu_B > 0$$

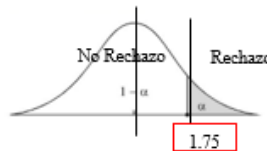
2. Nivel de significancia. $\alpha = 0.04$

3. Estadístico de prueba.

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

4. Establecimiento de la región crítica.

Nivel de significancia (α)	0.04
Z crítico	1.75



5. Cálculo del estadístico de prueba

		Empresa	
		A	B
Media muestral	$\bar{x}$	28	26
Desv. Std. Poblacional	σ	3	2
Varianza Poblacional	σ^2	9	4
Tamaño de muestra	n	25	32

Dif. Medias Muestrales	$\bar{x}_1 - \bar{x}_2$	2
Valor calculado de A	$Z_{calc.}$	2.8718

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \quad Z = \frac{2 - 0}{\sqrt{\frac{9}{25} + \frac{4}{32}}} = 2.8718$$

6. Toma de decisión.

Como el valor de Z calculado es mayor al valor crítico, la decisión es rechazar la hipótesis nula

7. Conclusión.

A un nivel de significancia de 0.04 se puede afirmar que la salsa de tomate de la empresa A tiene un sabor más intenso que la salsa de tomate de la empresa B.

Ver video de solución en Excel dando clic en el siguiente link <https://youtu.be/jPYFsOmE-Yk>

Ejercicios propuestos.

Ejercicio 1: Comparación de la calidad de dos marcas de aceite de oliva. Dos empresas, A y B, producen aceite de oliva. La empresa A afirma que su aceite tiene una acidez menor que la de la empresa B. Se sabe que la desviación estándar de la acidez del aceite de la empresa A es de 0.2%, y la de la empresa B es de 0.3%. Se toman muestras de 30 y 40 botellas

respectivamente de cada marca y se mide su acidez. Los resultados muestran una acidez promedio de 0.5% para la marca A y 0.6% para la marca B. ¿Existe evidencia estadística suficiente para apoyar la afirmación de la empresa A? Considere un nivel de significancia de 0.05. $H_0 \mu_A - \mu_B \geq 0$ $H_1 \mu_A - \mu_B < 0$, $Z_{calc.} = -1.6705$, *Rechazar H_0*

Ejercicio 2. Diferencia en el contenido de azúcar de dos tipos de yogur. Una empresa produce dos tipos de yogur: natural y con frutas. Se afirma que ambos yogures tienen el mismo contenido de azúcar. Se sabe que la desviación estándar del contenido de azúcar del yogur natural es de 2 gramos, y la del yogur con frutas es de 2.5 gramos. Se toman muestras de 25 y 28 yogures respectivamente de cada tipo y se mide su contenido de azúcar. Los resultados muestran un contenido promedio de 10 gramos de azúcar para el yogur natural y 11.5 gramos para el yogur con frutas. ¿Existe evidencia estadística suficiente para rechazar la afirmación de la empresa? Considere un nivel de significancia de 0.05. $H_0 \mu_N - \mu_F = 0$ $H_1 \mu_N - \mu_F \neq 0$, $Z_{calc.} = -2.2431$ y 2.2431 , *Rechazar H_0*

Ejercicio 3. Comparación del peso de dos tipos de galletas. Dos empresas, X e Y, producen galletas. La empresa X afirma que sus galletas superan en peso a las galletas de la empresa Y. Se sabe que la desviación estándar del peso de las galletas de la empresa X es de 1.5 gramos, y la de la empresa Y es de 1 gramo. Se toman muestras de 50 galletas de cada empresa y se pesan. Los resultados muestran un peso promedio de 20 gramos para las galletas de la empresa X y 19.5 gramos para las de la empresa Y. ¿Existe evidencia estadística suficiente para apoyar la afirmación de la empresa X? Considere un nivel de significancia de 0.01. $H_0 \mu_X - \mu_Y \leq 0$ $H_1 \mu_X - \mu_Y > 0$, $Z_{calc.} = 1.9612$, *No Rechazar H_0*

Ejercicio 4. Diferencia en la cantidad de sal de dos marcas de salsa de soya. Dos marcas de salsa de soya, A y B, compiten en el mercado. Se afirma que ambas salsas tienen la misma cantidad de sal. Se sabe que la desviación estándar de la cantidad de sal de la salsa A es de 0.5 gramos, y la de la salsa B es de 0.6 gramos. Se toman muestras de 25 y 32 botellas de salsa de soya respectivamente y se mide su contenido de sal. Los resultados muestran un contenido promedio de 10.1 gramos de sal para la salsa A y 10.35 gramos para la salsa B. ¿Existe evidencia estadística suficiente para rechazar la afirmación de que ambas salsas tienen la misma cantidad de sal? Considere un nivel de significancia de 0.05. $H_0 \mu_A - \mu_B = 0$ $H_1 \mu_A - \mu_B \neq 0$, $Z_{calc.} = 1.715$ y -1.715 , *No Rechazar H_0*

Ejercicio 5. Comparación del tiempo de cocción de dos marcas de pasta. Dos empresas, R y S, producen pasta. La empresa R afirma que su pasta se cocina en menos tiempo que la de la empresa S. Se sabe que la desviación estándar del tiempo de cocción de la pasta de la empresa R es de 2 minutos, y la de la empresa S es de 2.5 minutos. Se cocinan muestras de 30 y 35 paquetes de pasta de cada marca respectivamente y se registra el tiempo de cocción. Los resultados muestran un tiempo promedio de cocción de 8.9 minutos para la pasta de la empresa R y 10 minutos para la de la empresa S. ¿Existe evidencia estadística suficiente para

apoyar la afirmación de la empresa R? Considere un nivel de significancia de 0.04. $H_0: \mu_R - \mu_S \geq 0$ $H_1: \mu_R - \mu_S < 0$, $Z_{calc.} = -1.9696$, Rechazar H_0

Caso 4. Prueba de hipótesis para la diferencia de medias con desviaciones estándar poblacional desconocidas e iguales.

Para este caso la fórmula es.

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$S_p = \sqrt{\frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{n_1 + n_2 - 2}}$$

$$gl = n_1 + n_2 - 2$$

Ejemplo 1. Una empresa de alimentos está comparando dos métodos de producción de salsa de chile jalapeño. El método A es el método actual, mientras que el método B es un nuevo método que se espera que reduzca el tiempo de cocción. Se han tomado muestras aleatorias de los tiempos de cocción (en minutos) de ambos métodos:

Método A: 25, 28, 22, 24, 27, 26, 29, 23, 25, 28, 21, 24, 26, 27, 22

Método B: 24, 23, 23, 21, 22, 25, 24, 24, 23, 26, 22, 24

¿Existe evidencia suficiente para concluir que el método B reduce significativamente el tiempo de cocción en comparación con el método A? Utiliza un nivel de significancia del 5% y considera que las desviaciones estándar poblacional son desconocidas e iguales.

1. Formulación de hipótesis.

$$H_0 \mu_B - \mu_A \geq 0$$

$$H_1 \mu_B - \mu_A < 0$$

2. Nivel de significancia. $\alpha = 0.05$

3. Estadístico de prueba.

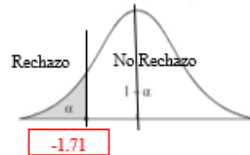
$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad s_p^2 = \frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{n_1 + n_2 - 2}$$

$$gl = n_1 + n_2 - 2$$

4. Establecimiento de la región crítica.

Nivel de significancia (α)	0.05
Grados de libertad (gl)	25
t crítico	-1.71

$$gl = 15 + 12 - 2 = 25$$



5. Cálculo del estadístico de prueba

		Método	
		A	B
Media muestral	$\bar{x}$	25.13	23.42
Varianza Muestral	s^2	5.98	1.90
Tamaño de muestra	n	15	12

Dif. Medias Muestrales	$\bar{x}_B - \bar{x}_A$	-1.72
Desv. Std. Conjunta	s_p	2.0460
Valor calculado de t	$t_{calc.}$	-2.1664

$$s_p^2 = \frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{n_1 + n_2 - 2} \quad s_p = \sqrt{\frac{5.98(15 - 1) + 1.9(12 - 1)}{15 + 12 - 2}} = 2.0460$$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad t = \frac{-1.72 - 0}{2.0460 * \sqrt{\frac{1}{15} + \frac{1}{12}}} = -2.1664$$

6. Toma de decisión.

Como el valor de t calculado es menor al valor crítico, la decisión es rechazar la hipótesis nula

7. Conclusión.

A un nivel de significancia de 0.05 se puede afirmar que el método B es un nuevo método que si reduce el tiempo de cocción en la producción de salsa de chile jalapeño.

Ver video de solución en Excel dando clic en el siguiente link <https://youtu.be/j-2Y9Xm1IVQ>

Ejemplo 2. Una empresa agrícola desea comparar la efectividad de dos tipos de fertilizantes (A y B) en el crecimiento de plantas de tomate. Se aplica cada fertilizante a un grupo de plantas y se mide la altura (en centímetros) de las plantas después de 30 días. Los resultados son los siguientes:

Fertilizante A: 35, 38, 40, 36, 39, 41, 37, 38, 40, 39

Fertilizante B: 32, 34, 36, 33, 35, 37, 34, 36, 38, 35, 33, 37

¿Existe evidencia suficiente para concluir que no existe una diferencia significativa en la altura promedio de las plantas de tomate entre los dos tipos de fertilizantes? Utiliza un nivel de significancia del 5% y considera que las desviaciones estándar poblacional son desconocidas e iguales.

1. Formulación de hipótesis.

$$H_0: \mu_A - \mu_B = 0$$

$$H_1: \mu_A - \mu_B \neq 0$$

2. Nivel de significancia. $\alpha = 0.05$

3. Estadístico de prueba.

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad s_p^2 = \frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{n_1 + n_2 - 2}$$

$$gl = n_1 + n_2 - 2$$

4. Establecimiento de la región crítica.

Nivel de significancia (α)	0.05
Grados de libertad (gl)	20
t crítico	2.09

$$gl = 10 + 12 - 2 = 20$$



5. Cálculo del estadístico de prueba

		Método	
		A	B
Media muestral	$\bar{x}$	38.30	35.00
Varianza Muestral	s^2	3.57	3.45
Tamaño de muestra	n	10	12

Dif. Medias Muestrales	$\bar{x}_A - \bar{x}_B$	3.30
Desv. Std. Conjunta	s_p	1.8722
Valor calculado de t	$t_{calc.}$	4.1167

$$s_p^2 = \frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{n_1 + n_2 - 2} \quad s_p = \sqrt{\frac{3.57(10 - 1) + 3.45(12 - 1)}{10 + 12 - 2}} = 1.8722$$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad t = \frac{3.33 - 0}{1.8722 * \sqrt{\frac{1}{10} + \frac{1}{12}}} = 4.1167$$

6. Toma de decisión.

Como el valor de t calculado es mayor al valor crítico, la decisión es rechazar la hipótesis nula

7. Conclusión.

A un nivel de significancia de 0.05 se puede afirmar que los fertilizantes A y B son diferentes en cuanto al crecimiento promedio de las plantas.

Ver video de solución en Excel dando clic en el siguiente link
<https://youtu.be/uUTtKgBiOCM>

Ejemplo 3. Una empresa de alimentos está investigando dos métodos de envasado para sus productos. El método A es el método tradicional, mientras que el método B es un nuevo método que se espera que prolongue la frescura del producto. Se envasan muestras aleatorias del mismo producto utilizando ambos métodos y se mide el tiempo (en días) que tarda el producto en mostrar signos de deterioro. Los resultados son los siguientes:

Método A: 19, 20, 19, 18, 19, 18, 21, 19, 20, 19, 19, 20.

Método B: 20, 18, 19, 21, 21, 20, 21, 20, 20, 19.

¿Existe evidencia suficiente para concluir que existe una diferencia significativa en el tiempo promedio que tarda el producto en deteriorarse entre los dos métodos de envasado? Utiliza un nivel de significancia del 4% y considera que las desviaciones estándar poblacional son desconocidas e iguales.

1. Formulación de hipótesis.

$$H_0: \mu_B - \mu_A \leq 0$$

$$H_1: \mu_B - \mu_A > 0$$

2. Nivel de significancia. $\alpha = 0.04$

3. Estadístico de prueba.

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad s_p^2 = \frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{n_1 + n_2 - 2}$$

$$gl = n_1 + n_2 - 2$$

4. Establecimiento de la región crítica.

Nivel de significancia (α)	0.04
Grados de libertad (gl)	20
t crítico	1.84

$$gl = 12 + 10 - 2 = 20$$



5. Cálculo del estadístico de prueba

		Método	
		A	B
Media muestral	$\bar{x}$	19.25	19.90
Varianza Muestral	s^2	0.75	0.99
Tamaño de muestra	n	12	10

Dif. Medias Muestrales	$\bar{x}_B - \bar{x}_A$	0.65
Dev. Std. Conjunta	s_p	0.9260
Valor calculado de t	$t_{calc.}$	1.6394

$$s_p^2 = \frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{n_1 + n_2 - 2} \quad S_p = \sqrt{\frac{0.75(12 - 1) + 0.99(10 - 1)}{12 + 10 - 2}} = 0.9260$$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad t = \frac{0.65 - 0}{0.9260 * \sqrt{\frac{1}{12} + \frac{1}{10}}} = 1.6394$$

6. Toma de decisión.

Como el valor de t calculado es menor al valor crítico, la decisión es no rechazar la hipótesis nula

7. Conclusión.

A un nivel de significancia de 0.04 al parecer los dos métodos de envasado son iguales en cuanto al tiempo promedio de conservación de los productos.

Ver video de solución en Excel dando clic en el siguiente link
<https://youtu.be/2XTwuVFGNOU>

Ejercicios propuestos.

Ejercicio 1. Una empresa de snacks está comparando dos máquinas de llenado de bolsas de papas fritas. La máquina A es la actual y la máquina B es una nueva que se espera que reduzca el tiempo en segundos en llenar cada bolsa. Se toman muestras aleatorias y se mide el tiempo de llenado con ambas máquinas:

Máquina A: 25, 28, 26, 27, 29, 24, 27, 26, 28, 25, 29, 27.

Máquina B: 26, 25, 25, 27, 26, 24, 25, 26.

¿Existe evidencia suficiente para concluir que la máquina B llena las bolsas con menos papas que la máquina A? Utiliza un nivel de significancia del 5% y considera que las desviaciones estándar poblacional son desconocidas e iguales. $H_0 \mu_B - \mu_A \geq 0$ $H_1 \mu_B - \mu_A < 0$, $t_{calc.} = -1.9377$, *Rechazar H_0*

Ejercicio 2. Una empresa de lácteos está comparando dos métodos de producción de yogur. El método A es el tradicional y el método B es uno nuevo que se espera que modifique el sabor del yogur. Se toman muestras aleatorias de la calificación de sabor (en una escala de 1 a 10) de consumidores que probaron yogur elaborado con ambos métodos:

Método A: 7, 8, 6, 9, 7, 8, 7, 6, 8, 9, 7, 6, 8, 7, 9.

Método B: 8, 9, 7, 6, 8, 9, 8, 7, 6, 8.

¿Existe evidencia suficiente para concluir que hay una diferencia en la calificación de sabor promedio entre los dos métodos de producción de yogur? Utiliza un nivel de significancia del 6% y considera que las desviaciones estándar poblacional son desconocidas e iguales.

$H_0 \mu_A - \mu_B = 0$ $H_1 \mu_A - \mu_B \neq 0$, $t_{calc.} = -0.9957$ y 0.9957 , *No Rechazar H_0*

Ejercicio 3. Una empresa de bebidas está comparando dos fórmulas de su bebida energética. La fórmula A es la actual y la fórmula B es una nueva que se espera que aumente el tiempo de reacción de los consumidores. Se miden los tiempos de reacción (en segundos) de consumidores que probaron ambas fórmulas:

Fórmula A: 15, 18, 16, 17, 19, 14, 18, 17, 16, 15, 19, 17, 16, 18, 17, 19, 15, 16.

Fórmula B: 17, 18, 19, 19, 17, 16, 19, 18, 17, 17, 19, 18, 19

¿Existe evidencia suficiente para concluir que la fórmula B aumenta el tiempo de reacción en comparación con la fórmula A? Utiliza un nivel de significancia del 1% y considera que las desviaciones estándar poblacional son desconocidas e iguales. $H_0 \mu_B - \mu_A \leq 0$ $H_1 \mu_B - \mu_A > 0$, $t_{calc.} = 2.3491$, *No Rechazar H_0*

Ejercicio 4. Una empresa de conservas está comparando dos proveedores de tomates. El proveedor A es el actual y el proveedor B es uno nuevo que se espera que tenga tomates con menor acidez. Se mide la acidez (en porcentaje) de tomates de ambos proveedores:

Proveedor A: 0.5, 0.6, 0.4, 0.5, 0.7, 0.3, 0.6, 0.5, 0.4, 0.5, 0.7, 0.3

Proveedor B: 0.3, 0.4, 0.2, 0.3, 0.5, 0.1, 0.4, 0.3, 0.2

¿Existe evidencia suficiente para concluir que los tomates del proveedor B tienen menor acidez que los tomates del proveedor A? Utiliza un nivel de significancia del 10% y considera que las desviaciones estándar poblacional son desconocidas e iguales. $H_0: \mu_B - \mu_A \geq 0$ $H_1: \mu_B - \mu_A < 0$, $t_{calc.} = -3.2941$, *Rechazar H_0*

Ejercicio 5. Una panadería artesanal está evaluando un nuevo tipo de levadura (Levadura B) que promete acelerar el proceso de fermentación de la masa de pan en comparación con su levadura tradicional (Levadura A). Para compararlas, preparan varias masas con cada tipo de levadura y miden el tiempo de fermentación en minutos:

Levadura A: 52, 55, 54, 52, 52, 56, 53, 55, 53, 54

Levadura B: 50, 53, 52, 54, 51, 53, 55, 52, 54, 50, 53, 51

¿Existe evidencia suficiente para concluir que la levadura B reduce el tiempo de fermentación en comparación con la levadura A? Utiliza un nivel de significancia del 5% y considera que las desviaciones estándar poblacional son desconocidas e iguales. $H_0: \mu_B - \mu_A \geq 0$ $H_1: \mu_B - \mu_A < 0$, $t_{calc.} = -1.9285$, *Rechazar H_0*

Caso 5. Prueba de hipótesis para la diferencia de medias con desviaciones estándar poblacional desconocidas y diferentes.

Para este caso la fórmula es.

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$gl = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}}$$

Ejemplo 1. Una empresa láctea está probando un nuevo proceso de pasteurización (Proceso B) que afirman que reduce el tiempo necesario para alcanzar la temperatura de pasteurización en comparación con su proceso tradicional (Proceso A). Para verificar esta afirmación, midieron el tiempo en minutos que tardó la leche en alcanzar la temperatura de pasteurización en varios lotes con cada proceso.

Proceso A: 12.5, 13.2, 12.8, 13.5, 12.9, 13.1, 12.7, 13.3, 13.0, 12.6, 13.4, 12.3, 13.6, 12.2, 13.7

Proceso B: 11.8, 12.1, 11.5, 12.4, 11.9, 12.2, 11.7, 12.0, 11.6, 12.3, 11.4, 12.5, 11.3, 12.6, 11.2

¿Existe evidencia suficiente para concluir que el Proceso B reduce el tiempo necesario para alcanzar la temperatura de pasteurización en comparación con el Proceso A? Utiliza un nivel de significancia del 5% y considera que las desviaciones estándar poblacionales son desconocidas y diferentes.

1. Formulación de hipótesis.

$$H_0 \mu_B - \mu_A \geq 0$$

$$H_1 \mu_B - \mu_A < 0$$

2. Nivel de significancia. $\alpha = 0.05$

3. Estadístico de prueba.

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$gl = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}}$$

4. Establecimiento de la región crítica.

Nivel de significancia (α)	0.05
Grados de libertad (gl)	27.94
t crítico	-1.70

28



5. Cálculo del estadístico de prueba

		Método	
		A	B
Media muestral	$\bar{x}$	12.99	11.90
Desv. Std. muestral	s	0.47	0.45
Varianza Muestral	s^2	0.22	0.20
Tamaño de muestra	n	15	15

Dif. Medias Muestrales	$\bar{x}_B - \bar{x}_A$	-1.09
Valor calculado de A	$t_{calc.}$	-6.4955

$$gl = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}}$$

$$gl = \frac{\left(\frac{0.22}{15} + \frac{0.20}{15}\right)^2}{\frac{\left(\frac{0.22}{15}\right)^2}{15 - 1} + \frac{\left(\frac{0.20}{15}\right)^2}{15 - 1}} = 27.94$$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$t = \frac{-1.09 - 0}{\sqrt{\frac{0.22}{15} + \frac{0.20}{15}}} = -6.4955$$

6. Toma de decisión.

Como el valor de t calculado es menor al valor crítico, la decisión es rechazar la hipótesis nula

7. Conclusión.

A un nivel de significancia de 0.05 se puede afirmar que el nuevo proceso de pasteurización (proceso B) reduce el tiempo promedio necesario para alcanzar la temperatura de pasteurización en comparación con el proceso tradicional (proceso A).

Ver video de solución en Excel dando clic en el siguiente link <https://youtu.be/3kKTvlrrqns>

Ejemplo 2. Una empresa de alimentos está comparando dos métodos de cocción para su nueva línea de productos de pollo frito: el método tradicional (MT) y un nuevo método de cocción al vacío (CV). La empresa quiere determinar si hay una diferencia significativa en el tiempo promedio de cocción entre los dos métodos.

Se tomaron muestras aleatorias de los tiempos de cocción (en minutos) para ambos métodos, resultando lo siguiente:

		Método	
		Tradicional	Al vacío
Media muestral	$\bar{x}$	20	18
Desviación estándar muestral	s	5.2	4.8
Varianza muestral	s^2	27.04	23.04
Tamaño de muestra	n	12	15

¿Existe evidencia suficiente para concluir que hay una diferencia significativa en el tiempo promedio de cocción entre el método tradicional y el método al vacío? Utiliza un nivel de significancia del 5% y considera que las desviaciones estándar poblacionales son desconocidas y diferentes.

1. Formulación de hipótesis.

$$H_0 \mu_T - \mu_V = 0$$

$$H_1 \mu_T - \mu_V \neq 0$$

2. Nivel de significancia. $\alpha = 0.05$

3. Estadístico de prueba.

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$gl = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}}$$

4. Establecimiento de la región crítica.

Nivel de significancia (α)	0.05
Grados de libertad (gl)	22.79
t crítico	2.07

23



5. Cálculo del estadístico de prueba

		Método	
		Tradicional	Al vacío
Media muestral	$\bar{x}$	20	18
Desv. Std. muestral	s	5.2	4.8
Varianza Muestral	s^2	27.04	23.04
Tamaño de muestra	n	12	15

Dif. Medias Muestrales	$\bar{x}_T - \bar{x}_V$	2.00
Valor calculado de A	$t_{calc.}$	1.0274

$$gl = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}}$$

$$gl = \frac{\left(\frac{27.04}{12} + \frac{23.04}{15}\right)^2}{\frac{\left(\frac{27.04}{12}\right)^2}{12 - 1} + \frac{\left(\frac{23.04}{15}\right)^2}{15 - 1}} = 22.79$$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$t = \frac{2 - 0}{\sqrt{\frac{27.04}{12} + \frac{23.04}{15}}} = 1.0274$$

6. Toma de decisión.

Como el valor de t calculado se encuentra entre los valores críticos, la decisión es no rechazar la hipótesis nula

7. Conclusión.

A un nivel de significancia de 0.05 al parecer ambos métodos de cocción tienen el mismo tiempo promedio de cocción.

Ver video de solución en Excel dando clic en el siguiente link
<https://youtu.be/O7fSNdEFXxc>

Ejemplo 3. Un fabricante de mermeladas está investigando un nuevo conservador químico (NC) que promete prolongar la vida útil de sus productos en comparación con el conservador tradicional (CT) que utiliza actualmente. Para evaluar su efectividad, se prepararon lotes de mermelada de fresa, con cada tipo de conservador, y se midió el tiempo (en días) que tardaron en mostrar signos de deterioro.

Se obtuvieron los siguientes datos:

		Conservador	
		Tradicional	Nuevo
Media muestral	$\bar{x}$	60	65
Desviación estándar muestral	s	8	7
Varianza muestral	s^2	64	49
Tamaño de muestra	n	8	10

¿Existe evidencia suficiente para concluir que el nuevo conservador prolonga la vida útil de la mermelada en comparación con el conservador tradicional? Utiliza un nivel de significancia del 1% y considera que las desviaciones estándar poblacionales son desconocidas y diferentes.

1. Formulación de hipótesis.

$$H_0 \mu_N - \mu_T \leq 0$$

$$H_1 \mu_N - \mu_T > 0$$

2. Nivel de significancia. $\alpha = 0.01$

3. Estadístico de prueba.

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \quad gl = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}}$$

4. Establecimiento de la región crítica.

Nivel de significancia (α)	0.01
Grados de libertad (gl)	14.09
t crítico	2.62

14



5. Cálculo del estadístico de prueba

		Conservador	
		Tradicional	Nuevo
Media muestral	$\bar{x}$	60	65
Desv. Std. muestral	s	8	7
Varianza Muestral	s^2	64	49
Tamaño de muestra	n	8	10

Dif. Medias Muestrales	$\bar{x}_N - \bar{x}_T$	5.00
Valor calculado de A	$t_{calc.}$	1.3921

$$gl = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}} \quad gl = \frac{\left(\frac{64}{8} + \frac{49}{10}\right)^2}{\frac{\left(\frac{64}{8}\right)^2}{8-1} + \frac{\left(\frac{49}{10}\right)^2}{10-1}} = 14.09$$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \quad t = \frac{5.00 - 0}{\sqrt{\frac{64}{8} + \frac{49}{10}}} = 1.3921$$

6. Toma de decisión.

Como el valor de t calculado es menor al valor crítico, la decisión no es rechazar la hipótesis nula

7. Conclusión.

A un nivel de significancia de 0.01 al parecer ambos conservadores tienen el mismo efecto en la duración promedio que la mermelada de fresa se empieza a deteriorar.

Ver video de solución en Excel dando clic en el siguiente link <https://youtu.be/jBZSBXqyxy0>

Ejercicios propuestos.

Ejercicio 1. Eficacia de un Nuevo Conservante. Una empresa de embutidos está probando un nuevo conservante (NC) que afirman que reduce la cantidad de bacterias en sus productos en

comparación con el conservante tradicional (CT). Para verificar esta afirmación, midieron el recuento de bacterias (en UFC/g) en muestras de embutidos con cada conservante.

Conservante T

150, 160, 140, 170, 150, 180, 160, 150, 170, 140, 160

Conservante X

130, 140, 120, 150, 130, 140, 130, 120, 140, 130, 150, 120, 140

¿Existe evidencia suficiente para concluir que el Conservante X reduce el recuento de bacterias en comparación con el Conservante T? Utiliza un nivel de significancia del 5% y considera que las desviaciones estándar poblacionales son desconocidas y diferentes.

$$H_0 \mu_X - \mu_T \geq 0 \quad H_1 \mu_X - \mu_T < 0, \quad t_{calc.} = -4.8736, \text{ Rechazar } H_0$$

Ejercicio 2. Comparación de dos Métodos de Horneado. Una panadería está evaluando dos métodos de horneado para sus panes: un método tradicional (Método A) y un nuevo método con horno de convección (Método B). Quieren determinar si hay una diferencia significativa en el peso promedio (gramos) de los panes horneados con cada método.

Método A

500, 520, 510, 530, 520, 510, 500, 520, 510, 530, 520, 510, 500, 520, 510

Método B

520, 540, 530, 550, 540, 530, 520, 540, 530, 550, 540

¿Existe evidencia suficiente para concluir que hay una diferencia significativa en el peso promedio de los panes entre el Método A y el Método B? Utiliza un nivel de significancia del 5% y considera que las desviaciones estándar poblacionales son desconocidas y diferentes. $H_0 \mu_A - \mu_B = 0 \quad H_1 \mu_A - \mu_B \neq 0, \quad t_{calc.} = -5.323, \text{ Rechazar } H_0$

Ejercicio 3. Efecto de un Nuevo Ingrediente. Una empresa de alimentos está desarrollando una nueva barra energética con un ingrediente adicional que afirman que aumenta el rendimiento físico de los atletas. Quieren determinar si esta nueva barra (Barra N) es más efectiva que la barra actual (Barra A) que están vendiendo. Para ello, midieron el tiempo de carrera en una prueba de resistencia en atletas que consumieron cada barra.

Barra A

25, 28, 26, 29, 27, 28, 25, 27, 26, 29, 28

Barra N

26, 29, 25, 28, 29, 29, 27, 29, 28, 28, 30, 28, 27, 26

¿Existe evidencia suficiente para concluir que la Barra N mejora el rendimiento físico en mayor medida que la Barra A? Utiliza un nivel de significancia del 5% y considera que las desviaciones estándar poblacionales son desconocidas y diferentes.

$$H_0 \mu_N - \mu_A \leq 0 \quad H_1 \mu_N - \mu_A > 0, \quad t_{calc.} = 1.2091, \quad \text{No Rechazar } H_0$$

Ejercicio 4. Preferencia de los Consumidores. Una empresa de bebidas está lanzando dos nuevas variedades de su producto estrella: una versión con sabor a limón (Versión L) y otra con sabor a naranja (Versión N). Quieren determinar si hay una diferencia significativa en la preferencia de los consumidores por una u otra versión. Para ello, realizaron una encuesta y registraron cuántos consumidores prefirieron cada versión.

Versión L

120, 130, 110, 140, 120, 130, 110, 140, 120, 130, 110, 140, 120

Versión N

100, 110, 90, 120, 100, 110, 90, 120, 100, 110, 90, 120

¿Existe evidencia suficiente para concluir que hay una diferencia significativa en la preferencia de los consumidores entre la Versión L y la Versión N? Utiliza un nivel de significancia del 5% y considera que las desviaciones estándar poblacionales son desconocidas y diferentes.

$$H_0 \mu_L - \mu_N = 0 \quad H_1 \mu_L - \mu_N \neq 0, \quad t_{calc.} = 4.2674, \quad \text{Rechazar } H_0$$

Ejercicio 5. Duración de un Nuevo Proceso de Envasado. Una empresa de alimentos está probando un nuevo proceso de envasado que afirman que prolonga la vida útil de sus productos en comparación con el proceso tradicional. Quieren determinar si el nuevo proceso (Proceso N) realmente aumenta la vida útil de los productos. Para ello, midieron el tiempo (en días) que tardaron los productos envasados con cada proceso en mostrar signos de deterioro.

Proceso T

30, 35, 28, 32, 31, 34, 29, 33, 30, 36, 27, 31, 35

Proceso N

35, 40, 38, 42, 39, 41, 36, 40, 37, 43, 34, 39

¿Existe evidencia suficiente para concluir que el Proceso N prolonga la vida útil de los productos en comparación con el Proceso T? Utiliza un nivel de significancia del 5% y considera que las desviaciones estándar poblacionales son desconocidas y diferentes.

$$H_0 \mu_N - \mu_T \leq 0 \quad H_1 \mu_N - \mu_T > 0, \quad t_{calc.} = 6.281, \quad \text{Rechazar } H_0$$

Caso 6. Prueba de hipótesis para la diferencia de medias para datos pareados.

Para este caso la fórmula es.

$$t = \frac{\bar{d} - \mu_d}{\frac{s_d}{\sqrt{n}}}$$

$$gl = n - 1$$

Ejemplo 1. Una empresa de investigación en nutrición está estudiando el efecto de una nueva dieta rica en fibra sobre los niveles de colesterol LDL (el "colesterol malo") en adultos. Para ello, reclutaron a 7 participantes y midieron sus niveles de colesterol LDL antes de comenzar la dieta y después de 8 semanas de seguirla.

Los resultados obtenidos fueron los siguientes (en mg/dL):

Participante	Antes de la dieta	Después de la dieta
1	160	159
2	180	176
3	155	150
4	200	196
5	175	172
6	190	186
7	165	162

¿Existe evidencia suficiente para concluir que la nueva dieta rica en fibra reduce los niveles de colesterol LDL en adultos? Utiliza un nivel de significancia del 1%.

1. Formulación de hipótesis.

$$H_0 \mu_d \geq 0$$

$$H_1 \mu_d < 0$$

2. Nivel de significancia. $\alpha = 0.01$

3. Estadístico de prueba. $t = \frac{\bar{d} - \mu_d}{\frac{s_d}{\sqrt{n}}}$

4. Establecimiento de la región crítica.

Nivel de significancia (α)	0.01
Tamaño de muestra (n)	7
grados de libertad (gl)	6
t crítico	-3.14



5. Cálculo del estadístico de prueba

Media muestral	$\bar{d}$	-3.43
Desv. Std. Muestral	s	1.27
Tamaño de muestra	n	7

$$t = \frac{-3.43 - 0}{\frac{1.27}{\sqrt{7}}} = -7.13$$

t calculada	-7.13
-------------	-------

6. Toma de decisión.

Como el valor de t calculado es menor al valor crítico, la decisión es rechazar la hipótesis nula

7. Conclusión.

A un nivel de significancia de 0.01 se puede afirmar que la nueva dieta rica en fibra reduce los niveles de colesterol LDL.

Ver video de solución en Excel dando clic en el siguiente link <https://youtu.be/5lY-zOCxPpE>

Ejemplo 2. Un apicultor quiere evaluar si un nuevo tipo de colmena (Colmena N) produce más miel en comparación con su colmena tradicional (Colmena T). Para ello, colocó 4 colmenas de cada tipo en su apiario (en lugares similares para minimizar las diferencias en el entorno).

Colmena	Colmena T	Colmena N
1	20	21.9
2	18	20.8
3	22	23.5
4	19	21.9

¿Existe evidencia suficiente para concluir que la nueva colmena produce más miel? Utiliza un nivel de significancia del 1%.

1. Formulación de hipótesis.

$$H_0: \mu_d \geq 0$$

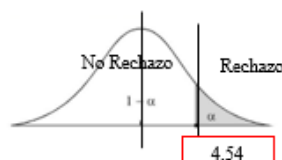
$$H_1: \mu_d < 0$$

2. Nivel de significancia. $\alpha = 0.01$

3. Estadístico de prueba. $t = \frac{\bar{d} - \mu_d}{\frac{s_d}{\sqrt{n}}}$

4. Establecimiento de la región crítica.

Nivel de significancia (α)	0.01
Tamaño de muestra (n)	4
grados de libertad (gl)	3
t crítico	4.54



5. Cálculo del estadístico de prueba

Media muestral	$\bar{d}$	2.28
Desv. Std. Muestral	s	0.68
Tamaño de muestra	n	4

$$t = \frac{2.28 - 0}{\frac{0.68}{\sqrt{4}}} = -7.13$$

t calculada	6.64
-------------	------

6. Toma de decisión.

Como el valor de t calculado es mayor al valor crítico, la decisión es rechazar la hipótesis nula

7. Conclusión.

A un nivel de significancia de 0.01 se puede afirmar que el nuevo tipo de colmena produce más miel que la colmena tradicional.

Ejercicios propuestos.

Ejercicio 1. Una empresa láctea está probando un nuevo tipo de fermento para su yogurt y quiere evaluar si este nuevo fermento (Fermento N) mejora la textura del yogurt en comparación con el fermento tradicional (Fermento T). Para ello, realizaron una prueba con 9 personas. Cada persona probó ambos yogurts (en orden aleatorio) y calificó la textura en una escala del 1 al 10, donde 1 es "Muy mala" y 10 es "Excelente".

Persona	Fermento T	Fermento N
1	7	8
2	6	9
3	8	7
4	5	6
5	9	8
6	7	9
7	6	7
8	8	8
9	5	7

¿Existe evidencia suficiente para concluir que el nuevo fermento mejora la textura del yogurt en comparación con el fermento tradicional? Utiliza un nivel de significancia del 4%.

$$H_0 \mu_d \leq 0 \quad H_1 \mu_d > 0, \quad t_{calc.} = 1.95, \quad \text{No Rechazar } H_0$$

Ejercicio 2. Un productor de café quiere evaluar si un nuevo método de tostado (Tostado N) produce un café con un sabor más intenso en comparación con su método tradicional (Tostado T). Para ello, realizó una cata con 8 expertos. Cada experto probó ambos cafés (en orden aleatorio) y calificó la intensidad del sabor en una escala del 1 al 10, donde 1 es "Muy suave" y 10 es "Muy intenso".

Experto	Tostado T	Tostado N
1	7	8
2	6	9
3	8	7
4	5	6
5	9	8
6	7	9
7	6	7
8	8	8

¿Existe evidencia suficiente para concluir que el nuevo método de tostado produce un café con un sabor más intenso? Utiliza un nivel de significancia del 5%.

$$H_0 \mu_d \leq 0 \quad H_1 \mu_d > 0, \quad t_{calc.} = 1.53, \quad \text{No Rechazar } H_0$$

Ejercicio 3. Una empresa de alimentos está desarrollando una nueva barra energética y quiere evaluar si esta nueva barra (Barra N) proporciona más energía en comparación con su barra tradicional (Barra T). Para ello, realizaron una prueba con 7 atletas. Cada atleta consumió ambas barras (en días diferentes) y se midió su rendimiento en una prueba física.

Atleta	Barra T	Barra N
1	70	71
2	65	63
3	75	76
4	60	62
5	80	82
6	70	71
7	65	66

¿Existe evidencia suficiente para concluir que la nueva barra energética proporciona más energía? Utiliza un nivel de significancia del 5%.

$$H_0 \mu_d \leq 0 \quad H_1 \mu_d > 0, \quad t_{calc.} = 1.69, \quad \text{No Rechazar } H_0$$

Ejercicio 4. Un granjero quiere evaluar si un nuevo fertilizante (Fertilizante N) aumenta el rendimiento de su cosecha de maíz en comparación con el fertilizante tradicional (Fertilizante T). Para ello, sembró 6 parcelas de maíz con cada fertilizante (en parcelas adyacentes para minimizar las diferencias en el suelo).

Parcela	Fertilizante T	Fertilizante N
1	80	83
2	75	76
3	85	88
4	70	72
5	90	91
6	80	85

¿Existe evidencia suficiente para concluir que el nuevo fertilizante aumenta el rendimiento de la cosecha de maíz? Utiliza un nivel de significancia del 2%.

$$H_0 \mu_d \leq 0 \quad H_1 \mu_d > 0, \quad t_{calc.} = 4.04, \quad \text{Rechazar } H_0$$

Ejercicio 5. Una empresa pesquera está probando una nueva red de pesca (Red N) que afirman que captura más peces en comparación con su red tradicional (Red T). Para ello, realizaron 5 pruebas de pesca con cada red (en diferentes días y en la misma zona de pesca).

Prueba	Red T	Red N
1	50	56
2	45	49
3	55	57
4	40	48
5	60	62

¿Existe evidencia suficiente para concluir que la nueva red de pesca captura más peces? Utiliza un nivel de significancia del 5%.

$$H_0 \mu_d \leq 0 \quad H_1 \mu_d > 0, \quad t_{calc.} = 3.77, \quad \text{Rechazar } H_0$$

Caso 7. Prueba de hipótesis para la proporción.

Para este caso la fórmula es.

$$z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$$

Ejemplo 1. Una empresa de producción de alimentos enlatados afirma que menos del 5% de sus latas tienen algún tipo de defecto (abolladuras, óxido, etc.). Para verificar esta afirmación, un inspector de calidad toma una muestra aleatoria de 200 latas y encuentra que 7 de ellas tienen defectos. ¿Existe evidencia suficiente para respaldar la afirmación de la empresa de que la proporción de latas defectuosas es inferior al 5%? Utiliza un nivel de significancia del 10%.

1. Formulación de hipótesis.

$$H_0: p \geq 0.05$$

$$H_1: p < 0.05$$

2. Nivel de significancia.

$$\alpha = 0.10$$

3. Estadístico de prueba.

$$z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$$

4. Establecimiento de la región crítica.

Nivel de significancia (α)	0.1
Z crítico	-1.28



5. Cálculo del estadístico de prueba

Proporción de éxito hipotética	p	0.05
Proporción de fracaso hipotética	q	0.95
Éxitos en la muestra	x	7
Tamaño de muestra	n	200
Proporción muestral	$\hat{p}$	0.035

Z calculada	Z_{calc}	-0.97
-------------	------------	-------

$$z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$$

$$z = \frac{0.035 - 0.05}{\sqrt{\frac{0.05 * 0.95}{200}}} = -0.97$$

6. Toma de decisión.

Como el valor de Z calculada es mayor al valor crítico, la decisión es no rechazar la hipótesis nula

7. Conclusión.

A un nivel de significancia de 0.10 al parecer la proporción de latas defectuosas es mayor o igual a 0.05

Ver video de solución en Excel dando clic en el siguiente link
[https://youtu.be/ PwGHX9kkAA](https://youtu.be/PwGHX9kkAA)

Ejemplo 2. Una cadena de restaurantes de comida rápida afirma que el 80% de sus clientes están "muy satisfechos" con la experiencia en sus establecimientos. Un estudio de mercado independiente busca verificar esta afirmación. Para ello, encuestan a 300 clientes seleccionados al azar, y 225 de ellos reportan estar "muy satisfechos". Con un nivel de significancia del 5%, ¿existe evidencia suficiente para refutar la afirmación de la cadena de restaurantes sobre el porcentaje de clientes "muy satisfechos"?

1. Formulación de hipótesis.

$$H_0: p = 0.80$$

$$H_1: p \neq 0.80$$

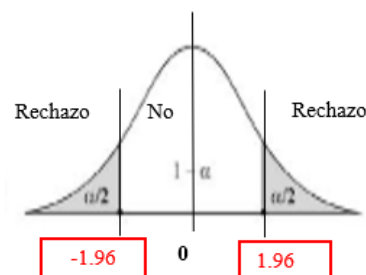
2. Nivel de significancia. $\alpha = 0.05$

3. Estadístico de prueba.

$$z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$$

4. Establecimiento de la región crítica.

Nivel de significancia (α)	0.05
Z crítico	1.96



5. Cálculo del estadístico de prueba

Proporción de éxito hipotética	p	0.8
Proporción de fracaso hipotética	q	0.2
Éxitos en la muestra	x	225
Tamaño de muestra	n	300
Proporción muestral	$\hat{p}$	0.75

Z calculada	Z_{calc}	-2.17
-------------	------------	-------

$$z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$$

$$z = \frac{0.75 - 0.8}{\sqrt{\frac{0.80 * 0.20}{300}}} = -2.17$$

6. Toma de decisión.

Como el valor de Z calculada se encuentra fuera de los valores críticos, la decisión es rechazar la hipótesis nula

7. Conclusión.

A un nivel de significancia de 0.05 se puede afirmar que la proporción de clientes muy satisfechos de la cadena de restaurantes de comida rápida es diferente de 0.80.

Ver video de solución en Excel dando clic en el siguiente link
<https://youtu.be/UP0WEjOoqFQ>

Ejemplo 3. Una empresa de producción de vinagre de manzana afirma que al menos el 95% de sus botellas contienen la cantidad de vinagre de manzana especificada en la etiqueta. Un inspector de calidad sospecha que la proporción real podría ser menor. Para ello, toma una muestra aleatoria de 150 botellas y mide la cantidad de vinagre de manzana en cada una. Encuentra que 145 botellas cumplen con la especificación de la etiqueta. Con un nivel de significancia del 5%, ¿existe evidencia suficiente para respaldar la afirmación de la empresa de que la proporción de botellas que cumplen con la especificación es superior al 95%?

1. Formulación de hipótesis.

$$H_0: p \leq 0.95$$

$$H_1: p > 0.95$$

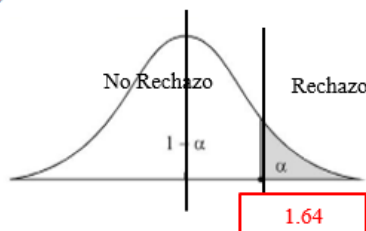
2. Nivel de significancia. $\alpha = 0.05$

3. Estadístico de prueba.

$$z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$$

4. Establecimiento de la región crítica.

Nivel de significancia (α)	0.05
Z crítico	1.64



5. Cálculo del estadístico de prueba

Proporción de éxito hipotética	p	0.9
Proporción de fracaso hipotética	q	0.1
Éxitos en la muestra	x	145
Tamaño de muestra	n	150
Proporción muestral	$\hat{p}$	0.97

Z calculada	Z_{calc}	2.72
-------------	------------	------

$$z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$$

$$z = \frac{0.97 - 0.90}{\sqrt{\frac{0.90 * 0.10}{150}}} = 2.72$$

6. Toma de decisión.

Como el valor de Z calculada es mayor al valor crítico, la decisión es rechazar la hipótesis nula

7. Conclusión.

A un nivel de significancia de 0.05 se puede afirmar que más del 95% de las botellas contienen la cantidad de vinagre de manzana especificada en la etiqueta.

Ver video de solución en Excel dando clic en el siguiente link
<https://youtu.be/8dy5DoLX8fU>

Ejercicios propuestos

Ejercicio 1. Una empresa productora de jugos afirma que menos del 3% de sus botellas tienen un volumen inferior al declarado en la etiqueta. Se toma una muestra aleatoria de 250 botellas y se encuentra que 5 de ellas tienen un volumen inferior al declarado. ¿Existe evidencia suficiente, con un nivel de significancia del 1%, para refutar la afirmación de la empresa?

$$H_0 \ p \geq 0.03 \quad H_1 \ p < 0.03, \quad Z_{calc.} = -1.9269, \quad \text{No rechazar } H_0$$

Ejercicio 2. Un estudio de mercado afirma que el 60% de los consumidores prefieren un nuevo sabor de helado. Se realiza una encuesta a 200 consumidores y se encuentra que 110 prefieren el nuevo sabor. ¿Existe evidencia suficiente, con un nivel de significancia del 5%, para afirmar que la proporción de consumidores que prefieren el nuevo sabor es diferente del 60%?

$$H_0 \ p = 0.60 \quad H_1 \ p \neq 0.60, \quad Z_{calc.} = -1.4434, \quad \text{No rechazar } H_0$$

Ejercicio 3. Un agricultor afirma que más del 90% de su cosecha de tomates cumple con los estándares de calidad para ser exportada. Se toma una muestra aleatoria de 150 tomates y se encuentra que 140 cumplen con los estándares. ¿Existe evidencia suficiente, con un nivel de significancia del 5%, para respaldar la afirmación del agricultor?

$$H_0 \ p \leq 0.90 \quad H_1 \ p > 0.90, \quad Z_{calc.} = 1.9052, \quad \text{Rechazar } H_0$$

Ejercicio 4. Un fabricante de galletas afirma que menos del 8% de sus paquetes contienen menos de la cantidad de galletas indicada en el empaque. Se toman 100 paquetes de galletas al azar y se encuentra que 6 de ellos contienen menos galletas de lo indicado. ¿Existe evidencia suficiente, con un nivel de significancia del 10%, para refutar la afirmación del fabricante?

$$H_0 \ p \geq 0.08 \quad H_1 \ p < 0.08, \quad Z_{calc.} = -1.4744, \quad \text{Rechazar } H_0$$

Ejercicio 5. Un nutricionista afirma que el 55% de los adultos consumen la cantidad recomendada de frutas y verduras al día. Se toma una muestra aleatoria de 300 adultos y se encuentra que 170 consumen la cantidad recomendada. ¿Existe evidencia suficiente, con un nivel de significancia del 5%, para afirmar que la proporción de adultos que consumen la cantidad recomendada de frutas y verduras es diferente del 55%?

$$H_0 \ p = 0.55 \quad H_1 \ p \neq 0.55, \quad Z_{calc.} = 1.7438, \quad \text{No rechazar } H_0$$

Caso 8. Prueba de hipótesis para la diferencia de proporciones.

Para este caso la fórmula es.

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}}$$

Ejemplo 1. Una empresa de alimentos lanza dos nuevas campañas publicitarias (A y B) para un nuevo producto. Se quiere determinar si la campaña A tiene una tasa de conversión (proporción de personas que compran el producto después de ver la campaña) significativamente menor que la campaña B. Se muestra la campaña A a 250 personas, de las cuales 80 compran el producto. La campaña B se muestra a 200 personas, de las cuales 75 compran el producto. Con un nivel de significancia del 5%, ¿existe evidencia suficiente para afirmar que la campaña A tiene una tasa de conversión menor que la campaña B?

1. Formulación de hipótesis.

$$H_0: p_A - p_B \geq 0$$

$$H_1: p_A - p_B < 0$$

2. Nivel de significancia.

$$\alpha = 0.05$$

3. Estadístico de prueba.

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}}$$

4. Establecimiento de la región crítica.

Nivel de significancia (α)	0.05
Z critico	-1.64



5. Cálculo del estadístico de prueba

		Campaña	
		A	B
Exitos en la muestra	x	80	75
Tamaño de muestra	n	250	200
Proporción de éxito muestral	$\hat{p}$	0.32	0.375
Proporción de fracaso muestral	$\hat{q}$	0.68	0.625

Diferencia de proporciones hipotética	$p_A - p_B$	0.00
Z calculada	$\hat{p}_A - \hat{p}_B$	-0.06
Z calculada	Z_{calc}	-1.22

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}} \quad Z = \frac{-0.06 - 0}{\sqrt{\frac{0.32 * 0.68}{250} + \frac{0.375 * 0.625}{200}}} = -1.22$$

6. Toma de decisión.

Como el valor de Z calculada es mayor al valor crítico, la decisión es no rechazar la hipótesis nula

7. Conclusión.

A un nivel de significancia de 0.05 al parecer la campaña A tiene una tasa de conversión igual o mayor que la campaña B.

Ver video de solución en Excel dando clic en el siguiente link <https://youtu.be/1J8U2BbdUfg>

Ejemplo 2. Una empresa de comida rápida introduce dos nuevos combos promocionales (Combo 1 y Combo 2) con la expectativa de que ambos tengan la misma aceptación entre los clientes. Para evaluar su popularidad, se ofrece el Combo 1 a 300 clientes, de los cuales 150 lo eligen. El Combo 2 se ofrece a 280 clientes, y 130 lo seleccionan. Con un nivel de significancia del 10%, ¿existe evidencia suficiente para afirmar que la proporción de clientes que prefieren el Combo 1 es diferente de la proporción que prefiere el Combo 2?

1. Formulación de hipótesis.

$$H_0: p_1 - p_2 = 0$$

$$H_1: p_1 - p_2 \neq 0$$

2. Nivel de significancia.

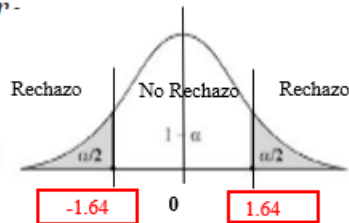
$$\alpha = 0.10$$

3. Estadístico de prueba.

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}}$$

4. Establecimiento de la región crítica.

Nivel de significancia (α)	0.1
Z crítico	1.64



5. Cálculo del estadístico de prueba

		Combo	
		1	2
Éxitos en la muestra	x	150	130
Tamaño de muestra	n	300	280
Proporción de éxito muestral	$\hat{p}$	0.5	0.46
Proporción de fracaso muestral	$\hat{q}$	0.5	0.54

Diferencia de proporciones hipotética	$p_1 - p_2$	0.00
Z calculada	$\hat{p}_1 - \hat{p}_2$	0.04
Z calculada	Z_{calc}	0.86

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}} = \frac{0.04 - 0}{\sqrt{\frac{0.5 * 0.5}{300} + \frac{0.46 * 0.54}{280}}} = 0.86$$

6. Toma de decisión.

Como el valor de Z calculada se encuentra entre los valores críticos, la decisión es no rechazar la hipótesis nula

7. Conclusión.

A un nivel de significancia de 0.10 al parecer ambos combos de comida rápida tienen el mismo grado de aceptación entre los clientes.

Ver video de solución en Excel dando clic en el siguiente link <https://youtu.be/6KPy3ul6Ed8>

Ejemplo 3. Dos empresas de alimentos compiten en el mercado de barras energéticas. La empresa A afirma que su nueva barra tiene una mayor aceptación entre deportistas que la barra de la empresa B. Para verificar esto, se realiza una encuesta a 400 deportistas, de los cuales 220 prefieren la barra A. La empresa B, por su parte, encuesta a 350 deportistas y 160 prefieren su barra. Con un nivel de significancia del 5%, ¿existe evidencia suficiente para afirmar que la proporción de deportistas que prefieren la barra A es mayor a la proporción que prefiere la barra B?

1. Formulación de hipótesis.

$$H_0: p_A - p_B \leq 0$$

$$H_1: p_A - p_B > 0$$

2. Nivel de significancia.

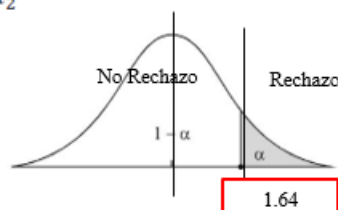
$$\alpha = 0.05$$

3. Estadístico de prueba.

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}}$$

4. Establecimiento de la región crítica.

Nivel de significancia (α)	0.05
Z crítico	1.64



5. Cálculo del estadístico de prueba

		Campaña	
		A	B
Éxitos en la muestra	x	220	160
Tamaño de muestra	n	400	350
Proporción de éxito muestral	$\hat{p}$	0.55	0.46
Proporción de fracaso muestral	$\hat{q}$	0.45	0.54

Diferencia de proporciones hipotética	$p_A - p_B$	0.00
Z calculada	$\hat{p}_A - \hat{p}_B$	0.09
Z calculada	Z_{calc}	2.55

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}} \quad Z = \frac{0.09 - 0}{\sqrt{\frac{0.55 * 0.45}{400} + \frac{0.46 * 0.54}{350}}} = 2.55$$

6. Toma de decisión.

Como el valor de Z calculada es mayor al valor crítico, la decisión es rechazar la hipótesis nula

7. Conclusión.

A un nivel de significancia de 0.05 se puede afirmar que la proporción de deportistas que prefieren la barra A es mayor a la proporción que prefiere la barra B

Ver video de solución en Excel dando clic en el siguiente link <https://youtu.be/etiPaXtZ8Tw>

Ejercicios propuestos.

Ejercicio 1. Una empresa de snacks está probando dos nuevas recetas de papas fritas: la receta A y la receta B. Creen que la receta A, que es más económica, será menos popular que la receta B. Se realiza una encuesta a 200 personas que prueban la receta A y 150 personas que prueban la receta B. 70 personas prefieren la receta A y 90 personas prefieren la receta B. Con un nivel de significancia del 5%, ¿existe evidencia suficiente para afirmar que la proporción de personas que prefieren la receta A es menor que la proporción de personas que prefieren la receta B?

$$H_0 \ p_A - p_B \geq 0 \quad H_1 \ p_A - p_B < 0, \quad Z_{calc.} = -2.0424, \quad \text{Rechazar } H_0$$

Ejercicio 2. Dos cadenas de supermercados, Supermercado X y Supermercado Y, afirman tener la misma proporción de clientes que compran productos orgánicos. Se realiza un estudio y se encuentra que, de 300 clientes del Supermercado X, 150 compran productos orgánicos, mientras que de 250 clientes del Supermercado Y, 100 compran productos orgánicos. Con un nivel de significancia del 1%, ¿existe evidencia suficiente para afirmar que la proporción de clientes que compran productos orgánicos es diferente entre los dos supermercados?

$$H_0 \ p_X - p_Y = 0 \quad H_1 \ p_X - p_Y \neq 0, \quad Z_{calc.} = 2.3452, \quad \text{No Rechazar } H_0$$

Ejercicio 3. Una empresa de bebidas energéticas lanza dos nuevos sabores: sabor Limón y sabor Frutos Rojos. La empresa cree que el sabor Limón tendrá una mayor aceptación entre los deportistas. Se realiza una encuesta a 180 deportistas que prueban el sabor Limón y 160 deportistas que prueban el sabor Frutos Rojos. 110 deportistas prefieren el sabor Limón y 80 deportistas prefieren el sabor Frutos Rojos. Con un nivel de significancia del 10%, ¿existe evidencia suficiente para afirmar que la proporción de deportistas que prefieren el sabor Limón es mayor que la proporción de deportistas que prefieren el sabor Frutos Rojos?

$$H_0 \ p_L - p_F \leq 0 \quad H_1 \ p_L - p_F > 0, \quad Z_{calc.} = 2.0595, \quad \text{Rechazar } H_0$$

Ejercicio 4. Una empresa de alimentos para mascotas está probando dos nuevas fórmulas: fórmula Premium y fórmula Económica. Creen que la fórmula Económica será menos popular entre los dueños de mascotas. Se realiza una encuesta a 220 dueños de mascotas que prueban la fórmula Premium y 250 dueños de mascotas que prueban la fórmula Económica. 150 dueños de mascotas prefieren la fórmula Premium y 150 dueños de mascotas prefieren la fórmula Económica. Con un nivel de significancia del 5%, ¿existe evidencia suficiente para afirmar que la proporción de dueños de mascotas que prefieren la fórmula Económica es menor que la proporción de dueños de mascotas que prefieren la fórmula Premium?

$$H_0 \ p_E - p_P \geq 0 \quad H_1 \ p_E - p_P < 0, \quad Z_{calc.} = -1.842, \quad \text{Rechazar } H_0$$

Ejercicio 5. Dos empresas de café, Café A y Café B, compiten en el mercado. Café A afirma que su café tiene una mayor proporción de cafeína que el café de la competencia. Se toman muestras de ambos cafés y se encuentra que de 200 tazas de Café A, 180 tienen un alto contenido de cafeína, mientras que, de 250 tazas de Café B, 200 tienen un alto contenido de cafeína. Con un nivel de significancia del 1%, ¿existe evidencia suficiente para afirmar que la proporción de tazas con alto contenido de cafeína es diferente entre las dos marcas de café?

$$H_0 \ p_A - p_B = 0 \quad H_1 \ p_A - p_B \neq 0, \quad Z_{calc.} = 2.9084, \quad \text{Rechazar } H_0$$

Caso 9. Prueba de hipótesis para la varianza.

Para este caso la fórmula es.

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$

$$gl = n - 1$$

Esta prueba de hipótesis implica el estadístico de probabilidad de chi cuadrada, cuya tabla de probabilidades se presenta a continuación.

Tabla A.5 Valores críticos de la distribución chi cuadrada

739

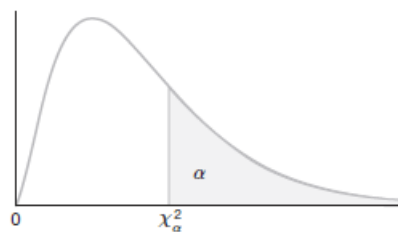


Tabla A.5 Valores críticos de la distribución chi cuadrada

v	α									
	0.995	0.99	0.98	0.975	0.95	0.90	0.80	0.75	0.70	0.50
1	0.004393	0.01157	0.01628	0.01982	0.02393	0.03158	0.0442	0.052	0.061	0.105
2	0.0100	0.0201	0.0270	0.0318	0.0370	0.0468	0.0600	0.069	0.080	0.136
3	0.0717	0.115	0.159	0.185	0.216	0.260	0.319	0.354	0.394	0.484
4	0.207	0.297	0.429	0.484	0.558	0.675	0.831	0.935	1.064	1.385
5	0.412	0.554	0.752	0.831	0.935	1.145	1.357	1.509	1.676	2.078
6	0.676	0.872	1.134	1.237	1.357	1.604	1.848	2.045	2.246	2.798
7	0.989	1.239	1.564	1.690	1.833	2.167	2.445	2.693	2.938	3.599
8	1.344	1.647	2.032	2.180	2.333	2.700	3.007	3.247	3.485	4.317
9	1.735	2.088	2.532	2.700	2.875	3.248	3.591	3.828	4.065	5.024
10	2.156	2.558	3.059	3.247	3.438	3.815	4.191	4.428	4.661	5.791
11	2.603	3.053	3.609	3.816	4.013	4.390	4.807	5.044	5.277	6.581
12	3.074	3.571	4.178	4.404	4.600	5.000	5.437	5.674	5.907	7.278
13	3.565	4.107	4.765	5.009	5.212	5.622	6.079	6.316	6.548	8.034
14	4.075	4.660	5.368	5.629	5.841	6.261	6.737	6.974	7.206	8.690
15	4.601	5.229	5.985	6.262	6.481	6.911	7.397	7.634	7.866	9.348
16	5.142	5.812	6.614	6.908	7.135	7.575	8.071	8.308	8.539	10.090
17	5.697	6.408	7.255	7.564	7.799	8.249	8.755	8.992	9.223	10.831
18	6.265	7.015	7.906	8.231	8.474	8.933	9.449	9.686	9.917	11.572
19	6.844	7.633	8.567	8.907	9.158	9.627	10.153	10.390	10.621	12.312
20	7.434	8.260	9.237	9.591	9.851	10.329	10.865	11.102	11.333	13.051
21	8.034	8.897	9.915	10.283	10.551	11.039	11.585	11.822	12.053	13.791
22	8.643	9.542	10.600	10.982	11.250	11.748	12.294	12.531	12.762	14.531
23	9.260	10.196	11.293	11.689	11.957	12.455	13.001	13.238	13.469	15.271
24	9.886	10.856	11.992	12.401	12.669	13.167	13.713	13.950	14.181	16.011
25	10.520	11.524	12.697	13.120	13.388	13.886	14.432	14.669	14.900	16.751
26	11.160	12.198	13.409	13.844	14.112	14.610	15.156	15.393	15.624	17.491
27	11.808	12.878	14.125	14.573	14.841	15.339	15.885	16.122	16.353	18.231
28	12.461	13.565	14.847	15.308	15.576	16.074	16.620	16.857	17.088	18.971
29	13.121	14.256	15.574	16.047	16.315	16.813	17.359	17.596	17.827	19.711
30	13.787	14.953	16.306	16.791	17.059	17.557	18.103	18.340	18.571	20.451
40	20.707	22.164	23.838	24.433	25.009	25.585	26.131	26.368	26.600	27.488
50	27.991	29.707	31.664	32.357	33.037	33.613	34.159	34.396	34.628	35.566
60	35.534	37.485	39.699	40.482	41.254	41.830	42.376	42.613	42.845	43.801

Tabla A.5 (continuación) Valores críticos de la distribución chi cuadrada

ν	α									
	0.30	0.25	0.20	0.10	0.05	0.025	0.02	0.01	0.005	0.001
1	1.074	1.323	1.642	2.706	3.841	5.024	5.412	6.635	7.879	10.827
2	2.408	2.773	3.219	4.605	5.991	7.378	7.824	9.210	10.597	13.815
3	3.665	4.108	4.642	6.251	7.815	9.348	9.837	11.345	12.838	16.266
4	4.878	5.385	5.989	7.779	9.488	11.143	11.668	13.277	14.860	18.466
5	6.064	6.626	7.289	9.236	11.070	12.832	13.388	15.086	16.750	20.515
6	7.231	7.841	8.558	10.645	12.592	14.449	15.033	16.812	18.548	22.457
7	8.383	9.037	9.803	12.017	14.067	16.013	16.622	18.475	20.278	24.321
8	9.524	10.219	11.030	13.362	15.507	17.535	18.168	20.090	21.955	26.124
9	10.656	11.389	12.242	14.684	16.919	19.023	19.679	21.666	23.589	27.877
10	11.781	12.549	13.442	15.987	18.307	20.483	21.161	23.209	25.188	29.588
11	12.899	13.701	14.631	17.275	19.675	21.920	22.618	24.725	26.757	31.264
12	14.011	14.845	15.812	18.549	21.026	23.337	24.054	26.217	28.300	32.909
13	15.119	15.984	16.985	19.812	22.362	24.736	25.471	27.688	29.819	34.527
14	16.222	17.117	18.151	21.064	23.685	26.119	26.873	29.141	31.319	36.124
15	17.322	18.245	19.311	22.307	24.996	27.488	28.259	30.578	32.801	37.698
16	18.418	19.369	20.465	23.542	26.296	28.845	29.633	32.000	34.267	39.252
17	19.511	20.489	21.615	24.769	27.587	30.191	30.995	33.409	35.718	40.791
18	20.601	21.605	22.760	25.989	28.869	31.526	32.346	34.805	37.156	42.312
19	21.689	22.718	23.900	27.204	30.144	32.852	33.687	36.191	38.582	43.819
20	22.775	23.828	25.038	28.412	31.410	34.170	35.020	37.566	39.997	45.314
21	23.858	24.935	26.171	29.615	32.671	35.479	36.343	38.932	41.401	46.796
22	24.939	26.039	27.301	30.813	33.924	36.781	37.659	40.289	42.796	48.268
23	26.018	27.141	28.429	32.007	35.172	38.076	38.968	41.638	44.181	49.728
24	27.096	28.241	29.553	33.196	36.415	39.364	40.270	42.980	45.558	51.179
25	28.172	29.339	30.675	34.382	37.652	40.646	41.566	44.314	46.928	52.619
26	29.246	30.435	31.795	35.563	38.885	41.923	42.856	45.642	48.290	54.051
27	30.319	31.528	32.912	36.741	40.113	43.195	44.140	46.963	49.645	55.475
28	31.391	32.620	34.027	37.916	41.337	44.461	45.419	48.278	50.994	56.892
29	32.461	33.711	35.139	39.087	42.557	45.722	46.693	49.588	52.335	58.301
30	33.530	34.800	36.250	40.256	43.773	46.979	47.962	50.892	53.672	59.702
40	44.165	45.616	47.269	51.805	55.758	59.342	60.436	63.691	66.766	73.403
50	54.723	56.334	58.164	63.167	67.505	71.420	72.613	76.154	79.490	86.660
60	65.226	66.981	68.972	74.397	79.082	83.298	84.58	88.379	91.952	99.608

Para su consulta se requiere del nivel de significancia α y los grados de libertad gl los cuales son igual al tamaño de muestra n menos uno ($gl = n - 1$).

Ejemplo de consulta.

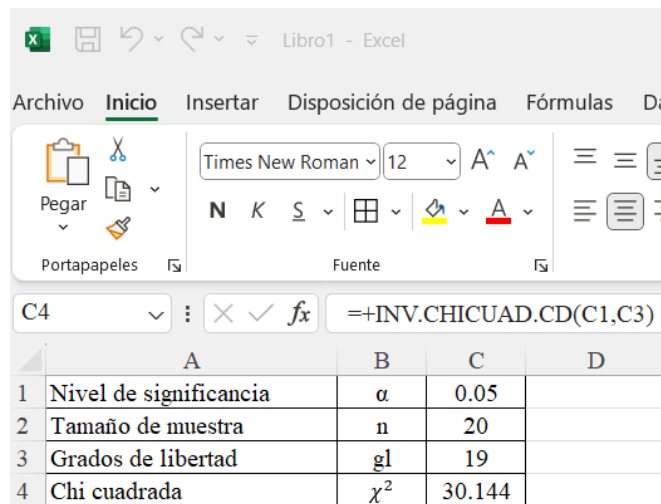
Cual será el valor de chi cuadrada para un nivel de significancia de 0.05 y un tamaño de muestra de 20

Para $\alpha = 0.05$, $n = 20$, $gl = 19$. El valor es de 30.144, tal como se puede observar a continuación.

Tabla A.5 (continuación) Valores críticos de la distribución chi cuadrada

ν	α									
	0.30	0.25	0.20	0.10	0.05	0.025	0.02	0.01	0.005	0.001
1	1.074	1.323	1.642	2.706	3.841	5.024	5.412	6.635	7.879	10.827
2	2.408	2.773	3.219	4.605	5.991	7.378	7.824	9.210	10.597	13.815
3	3.665	4.108	4.642	6.251	7.815	9.348	9.837	11.345	12.838	16.266
4	4.878	5.385	5.989	7.779	9.488	11.143	11.668	13.277	14.860	18.466
5	6.064	6.626	7.289	9.236	11.070	12.832	13.388	15.086	16.750	20.515
6	7.231	7.841	8.558	10.645	12.592	14.449	15.033	16.812	18.548	22.457
7	8.383	9.037	9.803	12.017	14.067	16.013	16.622	18.475	20.278	24.321
8	9.524	10.219	11.030	13.362	15.507	17.535	18.168	20.090	21.955	26.124
9	10.656	11.389	12.242	14.684	16.919	19.023	19.679	21.666	23.589	27.877
10	11.781	12.549	13.442	15.987	18.307	20.483	21.161	23.209	25.188	29.588
11	12.899	13.701	14.631	17.275	19.675	21.920	22.618	24.725	26.757	31.264
12	14.011	14.845	15.812	18.549	21.026	23.337	24.054	26.217	28.300	32.909
13	15.119	15.984	16.985	19.812	22.362	24.736	25.471	27.688	29.819	34.527
14	16.222	17.117	18.151	21.064	23.685	26.119	26.873	29.141	31.319	36.124
15	17.322	18.245	19.311	22.307	24.996	27.488	28.259	30.578	32.801	37.698
16	18.418	19.369	20.465	23.542	26.296	28.845	29.633	32.000	34.267	39.252
17	19.511	20.489	21.615	24.769	27.597	30.191	30.995	33.409	35.718	40.791
18	20.601	21.605	22.760	25.989	28.869	31.526	32.346	34.805	37.156	42.312
19	21.689	22.718	23.900	27.204	30.144	32.852	33.687	36.191	38.582	43.819
20	22.775	23.828	25.038	28.412	31.410	34.170	35.020	37.566	39.997	45.314
21	23.858	24.935	26.171	29.615	32.671	35.479	36.343	38.932	41.401	46.796
22	24.939	26.039	27.301	30.813	33.924	36.781	37.659	40.289	42.796	48.268
23	26.018	27.141	28.429	32.007	35.172	38.076	38.968	41.638	44.181	49.728
24	27.096	28.241	29.553	33.196	36.415	39.364	40.270	42.980	45.558	51.179
25	28.172	29.339	30.675	34.382	37.652	40.646	41.566	44.314	46.928	52.619
26	29.246	30.435	31.795	35.563	38.885	41.923	42.856	45.642	48.290	54.051
27	30.319	31.528	32.912	36.741	40.113	43.195	44.140	46.963	49.645	55.475
28	31.391	32.620	34.027	37.916	41.337	44.461	45.419	48.278	50.994	56.892
29	32.461	33.711	35.139	39.087	42.557	45.722	46.693	49.588	52.335	58.301
30	33.530	34.800	36.250	40.256	43.773	46.979	47.962	50.892	53.672	59.702
40	44.165	45.616	47.269	51.805	55.758	59.342	60.436	63.691	66.766	73.403
50	54.723	56.334	58.164	63.167	67.505	71.420	72.613	76.154	79.490	86.660
60	65.226	66.981	68.972	74.397	79.082	83.298	84.58	88.379	91.952	99.608

Obtener el valor de chi cuadrada mediante Excel es mediante la función INV.CHICUAD.CD, tal como se puede apreciar en la siguiente imagen.



Ejemplo 1. Una empresa productora de alimentos enlatados afirma que la varianza en el peso de sus latas de frijoles es menor que 0.5 onzas cuadradas. Se toma una muestra aleatoria de 25 latas y se encuentra que la varianza muestral es de 0.35 onzas cuadradas. ¿Existe evidencia suficiente para respaldar la afirmación de la empresa a un nivel de significancia del 5%?

1. Formulación de hipótesis.

$$H_0 \sigma^2 \geq 0.5$$

$$H_1 \sigma^2 < 0.5$$

2. Nivel de significancia. $\alpha = 0.05$

3. Estadístico de prueba. $\chi^2 = \frac{(n-1)s^2}{\sigma^2}$

4. Establecimiento de la región crítica.

Nivel de significancia (α)	0.05
Tamaño de muestra (n)	25
grados de libertad (gl)	24
χ^2 crítica	13.85



5. Cálculo del estadístico de prueba

Varianza hipotética	σ^2	0.5
Varianza muestral	s^2	0.35
Chi cuadrada calculada	χ^2 Calc.	16.8

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$

$$\chi^2 = \frac{(25-1) * 0.35}{0.5} = 16.8$$

6. Toma de decisión.

Como el valor de chi cuadrado calculado es mayor al valor crítico, la decisión es no rechazar la hipótesis nula

7. Conclusión.

A un nivel de significancia de 0.05 al parecer la varianza en el peso de las latas de frijoles es mayor o igual a 0.5 onzas cuadradas.

Ver video de solución en Excel dando clic en el siguiente link <https://youtu.be/gzBc6dr9kXk>

Ejemplo 2. Una empresa produce barras de cereal y busca mantener un peso consistente en sus productos. La máquina envasadora está configurada para producir barras con un peso promedio de 50 gramos y una varianza de 4 gramos². Se toma una muestra aleatoria de 20 barras y se encuentra que la varianza muestral es de 5 gramos². ¿Existe evidencia suficiente, con un nivel de significancia del 5%, para afirmar que la varianza del peso de las barras es diferente de la varianza objetivo?

1. Formulación de hipótesis.

$$H_0 \sigma^2 = 4$$

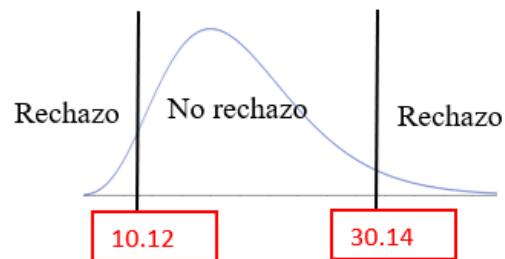
$$H_1 \sigma^2 \neq 4$$

2. Nivel de significancia. $\alpha = 0.05$

3. Estadístico de prueba. $\chi^2 = \frac{(n-1)s^2}{\sigma^2}$

4. Establecimiento de la región crítica.

Nivel de significancia (α)	0.1
Tamaño de muestra (n)	20
grados de libertad (gl)	19
χ^2 crítica izquierda	10.12
χ^2 crítica derecha	30.14



5. Cálculo del estadístico de prueba

Varianza hipotética	σ^2	4
Varianza muestral	s^2	5
Chi cuadrada calculada	χ^2 Calc.	23.75

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$

$$\chi^2 = \frac{(20-1) * 5}{4} = 23.75$$

6. Toma de decisión.

Como el valor de chi cuadrado calculado se encuentra entre los valores críticos, la decisión es no rechazar la hipótesis nula

7. Conclusión.

A un nivel de significancia de 0.05 al parecer la varianza en el peso de barras de cereal es de 4 gramos cuadrados.

Ver video de solución en Excel dando clic en el siguiente link <https://youtu.be/1hn4C-GFDf0>

Ejemplo 3. Una empresa que produce bolsas de papas fritas afirma que la varianza en el peso de sus bolsas es mayor a 25 gramos cuadrados. Se toma una muestra aleatoria de 30 bolsas y se encuentra que la varianza muestral es de 32 gramos cuadrados. ¿Existe evidencia suficiente para respaldar la afirmación de la empresa a un nivel de significancia del 1%?

1. Formulación de hipótesis.

$$H_0 \sigma^2 \leq 25$$

$$H_1 \sigma^2 > 25$$

2. Nivel de significancia. $\alpha = 0.01$

3. Estadístico de prueba. $\chi^2 = \frac{(n-1)s^2}{\sigma^2}$

4. Establecimiento de la región crítica.

Nivel de significancia (α)	0.01
Tamaño de muestra (n)	30
grados de libertad (gl)	29
χ^2 crítica	49.59



5. Cálculo del estadístico de prueba

Varianza hipotética	σ^2	25
Varianza muestral	s^2	32
Chi cuadrada calculada	χ^2 Calc.	37.12

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$

$$\chi^2 = \frac{(30-1) * 32}{25} = 37.12$$

6. Toma de decisión.

Como el valor de chi cuadrado calculado es menor al valor crítico, la decisión es no rechazar la hipótesis nula

A un nivel de significancia de 0.01 al parecer la varianza en el peso de las bolsas de papas fritas es menor o igual a 25 gramos cuadrados.

Ver video de solución en Excel dando clic en el siguiente link <https://youtu.be/GHUr3l3-nv8>

Ejercicios propuestos.

Ejercicio 1. Una empresa productora de leche afirma que la varianza en el contenido de grasa de sus botellas de leche descremada es menor a 0.05%. Se toma una muestra aleatoria de 20 botellas y se encuentra que la varianza muestral es de 0.03%. ¿Existe evidencia suficiente para respaldar la afirmación de la empresa a un nivel de significancia del 5%?

$$H_0 \sigma^2 \geq 0.05 \quad H_1 \sigma^2 < 0.05, \quad \chi^2_{calc.} = 11.4, \quad \text{Rechazar } H_0$$

Ejercicio 2. Un matadero de bovinos busca mantener la uniformidad en el peso de sus cortes de carne. Idealmente, la varianza en el peso de los cortes de carne empacados debería ser de 100 gramos cuadrados. Se selecciona una muestra aleatoria de 40 paquetes de carne y se encuentra una varianza muestral de 120 gramos cuadrados. ¿Existe evidencia suficiente a un nivel de significancia del 5% para concluir que la varianza en el peso de los cortes de carne es diferente de la deseada?

$$H_0 \sigma^2 = 100 \quad H_1 \sigma^2 \neq 100, \quad \chi_{calc.}^2 = 46.8, \quad \text{Rechazar } H_0$$

Ejercicio 3. Una empresa que produce bolsas de café asegura que la varianza en el peso neto de sus bolsas de café de 500 gramos es igual a 10 gramos cuadrados. Se toma una muestra aleatoria de 30 bolsas y se encuentra que la varianza muestral es de 11 gramos cuadrados. ¿Existe evidencia suficiente para rechazar la afirmación de la empresa a un nivel de significancia del 5%?

$$H_0 \sigma^2 = 10 \quad H_1 \sigma^2 \neq 10, \quad \chi_{calc.}^2 = 31.9, \quad \text{No Rechazar } H_0$$

Ejercicio 4. Una máquina embotelladora de refrescos está diseñada para llenar botellas con una varianza en el volumen de líquido no mayor a 10 ml cuadrados. Si la varianza es mayor, la máquina necesita ser recalibrada. Se toma una muestra de 28 botellas y se encuentra una varianza muestral de 12 ml cuadrados. Utilizando un nivel de significancia del 1%, ¿hay evidencia suficiente para indicar que la máquina necesita recalibración?

$$H_0 \sigma^2 \leq 10 \quad H_1 \sigma^2 > 10, \quad \chi_{calc.}^2 = 32.4, \quad \text{No Rechazar } H_0 \quad \text{Al parecer la maquina no requiere ser recalibrada.}$$

Ejercicio 5. Un proveedor de harina para panaderías afirma que la varianza en el contenido de gluten de su harina es de 2%. Un cliente realiza un análisis de 35 muestras de harina y encuentra una varianza muestral de 2.5%. Si el cliente utiliza un nivel de significancia del 10%, ¿debería dudar de la afirmación del proveedor?

$$H_0 \sigma^2 = 2 \quad H_1 \sigma^2 \neq 2, \quad \chi_{calc.}^2 = 42.5, \quad \text{No Rechazar } H_0 \quad \text{Al parecer la afirmación del proveedor es correcta.}$$

Caso 10. Prueba de hipótesis para la razón de varianzas.

Para este caso la fórmula es.

$$F = \frac{s_1^2}{s_2^2}$$

$$gl = n - 1$$

De lo anterior se puede observar que el estadístico de prueba es la distribución F de Fisher, lo cual implica el uso de la siguiente tabla.

Tabla A.6 Valores críticos de la distribución F

741

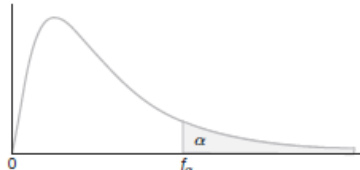


Tabla A.6 Valores críticos de la distribución F

$f_{\alpha,05}(v_1, v_2)$									
v_2	v_1								
	1	2	3	4	5	6	7	8	9
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88

Reproducida de la tabla 18 de *Biometrika Tables for Statisticians*, Vol. I, con autorización de E.S. Pearson y Biometrika Trustees.

Tabla A.6 (continuación) Valores críticos de la distribución F

v_2	$f_{0.05}(v_1, v_2)$									
	v_1									
	10	12	15	20	24	30	40	60	120	∞
1	241.88	243.91	245.95	248.01	249.05	250.10	251.14	252.20	253.25	254.31
2	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
3	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36
6	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
29	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
∞	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

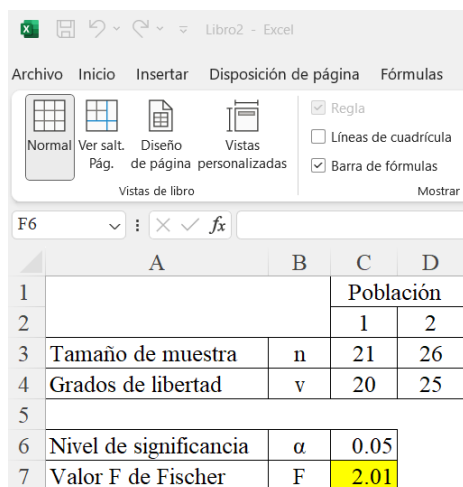
Ejemplo de uso. Cual es el valor F de Fisher para un nivel de significancia de 0.05, para dos poblaciones de las cuales se extrajeron muestras de 21 y 26 elementos, respectivamente.

Se consulta la tabla correspondiente a un $\alpha = 0.05$, con grados de libertad $v_1 = 20$ y $v_2 = 25$, obteniéndose un valor de $F = 2.01$, tal como se puede apreciar en la siguiente imagen.

Tabla A.6 (continuación) Valores críticos de la distribución F

v_2	$f_{0.05}(v_1, v_2)$									
	v_1									
	10	12	15	20	24	30	40	60	120	∞
1	241.88	243.91	245.95	248.01	249.05	250.10	251.14	252.20	253.25	254.31
2	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
3	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36
6	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
29	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
∞	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

La obtención de este variable mediante Excel implica usar la función INV.F.CD, tal como se muestra en la siguiente imagen.



	A	B	C	D
1			Población	
2			1	2
3	Tamaño de muestra	n	21	26
4	Grados de libertad	v	20	25
5				
6	Nivel de significancia	α	0.05	
7	Valor F de Fischer	F	2.01	

Esta prueba de hipótesis de razón de varianzas es de gran utilidad para determinar si las varianzas son iguales o diferentes, planteando las hipótesis siguientes:

$$H_o \frac{\sigma_1^2}{\sigma_2^2} = 1 \quad H_o \frac{\sigma_1^2}{\sigma_2^2} \neq 1$$

El hecho de rechazar la hipótesis nula, implica la aceptación de la hipótesis alternativa, lo que lleva a concluir que las varianzas son diferentes. En caso de no rechazar la hipótesis nula, se tiene que al parecer las varianzas son iguales.

Ejemplo 1. Dos fábricas de galletas producen un tipo similar de galleta. Se desea comparar la variabilidad en el diámetro de las galletas producidas por cada fábrica. Se toma una muestra aleatoria de 25 galletas de la fábrica A y 30 galletas de la fábrica B. La varianza muestral de la fábrica A es de 4 mm cuadrados, mientras que la varianza muestral de la fábrica B es de 6 mm cuadrados. ¿Existe evidencia suficiente a un nivel de significancia del 10% para concluir que la varianza en el diámetro de las galletas es diferente entre las dos fábricas?

1. Formulación de hipótesis.

$$H_0: \frac{\sigma_1^2}{\sigma_2^2} = 1$$

$$H_a: \frac{\sigma_1^2}{\sigma_2^2} \neq 1$$

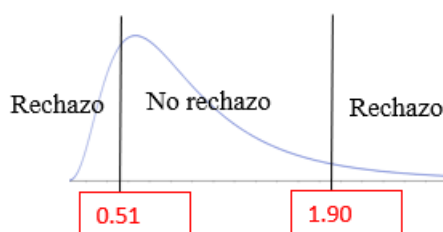
2. Nivel de significancia. $\alpha = 0.1$

3. Estadístico de prueba. $F = \frac{s_1^2}{s_2^2}$

4. Establecimiento de la región crítica.

	Fábrica	
	A	B
Tamaño de muestra (n)	25	30
grados de libertad (gl)	24	29
Varianza muestral (s)	4	6

Nivel de significancia (α)	0.1
F crítica por la izquierda	0.51
F crítica por la derecha	1.90



5. Cálculo del estadístico de prueba

F Calculada	$F_{Calc.}$	0.67
-------------	-------------	------

$$F = \frac{s_1^2}{s_2^2} \quad F = \frac{4}{6} = 0.667$$

6. Toma de decisión.

Como el valor de F calculada se encuentra entre los valores críticos, la decisión es no rechazar la hipótesis nula

7. Conclusión.

A un nivel de significancia de 0.10 al parecer las varianzas de las galletas de ambas fábricas son iguales

Ver video de solución en Excel dando clic en el siguiente link
<https://youtu.be/ZmEYqTWoiEM>

En el apartado de prueba de hipótesis para la diferencia de medias con desviaciones estándar poblacional desconocidas e iguales, en los ejercicios resueltos se puede verificar si el supuesto o consideración de que eran las desviaciones estándar poblacionales son iguales, es correcto o no. A continuación, se presenta un ejemplo de ello.

Ejemplo 1 del caso 4. Una empresa de alimentos está comparando dos métodos de producción de salsa de chile jalapeño. El método A es el método actual, mientras que el método B es un nuevo método que se espera que reduzca el tiempo de cocción. Se han tomado muestras aleatorias de los tiempos de cocción (en minutos) de ambos métodos:

Método A: 25, 28, 22, 24, 27, 26, 29, 23, 25, 28, 21, 24, 26, 27, 22

Método B: 24, 23, 23, 21, 22, 25, 24, 24, 23, 26, 22, 24

¿Existe evidencia suficiente para concluir que el método B reduce significativamente el tiempo de cocción en comparación con el método A? Utiliza un nivel de significancia del 5% y considera que las desviaciones estándar poblacional son desconocidas e iguales.

Verificar si efectivamente las varianzas son iguales.

1. Formulación de hipótesis.

$$H_0 \frac{\sigma_1^2}{\sigma_2^2} = 1$$

$$H_a \frac{\sigma_1^2}{\sigma_2^2} \neq 1$$

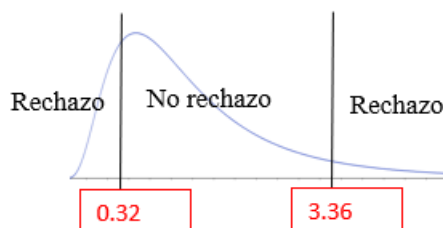
2. Nivel de significancia. $\alpha = 0.05$

3. Estadístico de prueba. $F = \frac{s_1^2}{s_2^2}$

4. Establecimiento de la región crítica.

	Método	
	A	B
Tamaño de muestra (n)	15	12
grados de libertad (gl)	14	11
Varianza muestral (s)	5.98	1.90

Nivel de significancia (α)	0.05
F crítica por la izquierda	0.32
F crítica por la derecha	3.36



5. Cálculo del estadístico de prueba

F Calculada	$F_{Calc.}$	3.15
-------------	-------------	------

$$F = \frac{s_1^2}{s_2^2} \quad F = \frac{5.98}{1.9} = 3.15$$

6. Toma de decisión.

Como el valor de F calculada se encuentra entre los valores críticos, la decisión es no rechazar la hipótesis nula

7. Conclusión.

A un nivel de significancia de 0.05 al parecer las varianzas ambos métodos de cocción de la salsa de chile jalapeño son iguales

La consideración de varianzas iguales fue correcta.

Ejercicios propuestos.

Ejercicio 1. Una empresa de producción de aceite de oliva está comparando el rendimiento de dos variedades de aceitunas. Se cosechan muestras de 15 árboles de cada variedad y se mide la cantidad de aceite producido por árbol. Si la media de producción de aceite muestral para la variedad A es de 12 kg y para la variedad B es de 10 kg, y las desviaciones estándar muestrales son de 2 kg y 1.5 kg respectivamente, construye un intervalo de confianza del 99% para la diferencia de medias poblacionales. Considerando que las desviaciones estándar poblacionales son desconocidas e iguales. Verificar si efectivamente las varianzas son iguales. Respuesta. $F_{calc.} = 1.778$ Efectivamente la consideración es correcta.

Ejercicio 2. Un productor de vino está evaluando el efecto de dos tipos de levadura en el proceso de fermentación. Se miden los niveles de alcohol en 13 muestras de vino elaboradas con cada tipo de levadura. Si la media del contenido de alcohol muestral para la levadura 1 es 13.5% y para la levadura 2 es 14%, y las desviaciones estándar muestrales son de 0.5% y 0.4% respectivamente, construye un intervalo de confianza del 90% para la diferencia de medias poblacionales. Considerando que las desviaciones estándar poblacionales son desconocidas e iguales. Verificar si efectivamente las varianzas son iguales.

Respuesta. $F_{calc.} = 1.563$ Efectivamente la consideración es correcta.

Ejercicio 3. Una empresa de alimentos congelados está comparando dos métodos de congelación. Se mide la pérdida de peso de los alimentos después de un cierto período de tiempo. Si la media de pérdida de peso de 14 muestras del método A es de 5% y para 9 muestras del método B es de 3%, y las desviaciones estándar muestrales son de 1% y 0.8% respectivamente, construye un intervalo de confianza del 95% para la diferencia de medias poblacionales. Considerando que las desviaciones estándar poblacionales son desconocidas e iguales. Verificar si efectivamente las varianzas son iguales.

Respuesta. $F_{calc.} = 1.563$ Efectivamente la consideración es correcta.

Ejercicio 4. Una empresa que produce alimentos congelados quiere comparar la cantidad de tiempo que tardan en descongelarse dos tipos de pescado: el salmón y el bacalao. Se toman muestras aleatorias de 14 filetes de salmón y 15 filetes de bacalao, y se mide el tiempo de descongelación en minutos. Si el tiempo promedio de descongelación del salmón es de 28 minutos con una desviación estándar muestral de 5 minutos, y el tiempo promedio de descongelación del bacalao es de 32 minutos con una desviación estándar muestral de 6 minutos, construye un intervalo de confianza del 95% para la diferencia de medias poblacionales. Considerando que las desviaciones estándar poblacionales son desconocidas e iguales. Verificar si efectivamente las varianzas son iguales.

Respuesta. $F_{calc.} = 0.694$ Efectivamente la consideración es correcta.

Ejercicio 5. Una compañía que fabrica jugos está investigando dos métodos de extracción de jugo de naranja. Se toman muestras aleatorias de 20 naranjas para el método A y 18 naranjas para el método B, y se mide la cantidad de jugo extraído en mililitros. Si la cantidad promedio de jugo extraído con el método A es de 120 ml con una desviación estándar muestral de 10 ml, y la cantidad promedio de jugo extraído con el método B es de 130 ml con una desviación estándar muestral de 12 ml, construye un intervalo de confianza del 90% para la diferencia de medias poblacionales. Considerando que las desviaciones estándar poblacionales son desconocidas e iguales. Verificar si efectivamente las varianzas son iguales.

Respuesta. $F_{calc.} = 0.694$ Efectivamente la consideración es correcta.

Una evaluación al contenido de esta unidad puede ser realizada escaneando el siguiente código QR.



Bibliografía consultada

Anderson, D. R., Sweeney, D. J., & Williams, T. A. (2012). *Estadística para administración y economía*. Cengage Learning.

Ballesteros, A., & Hernández, J. (2017). *Fundamentos de probabilidad y estadística*. McGraw-Hill.

Barbosa-Cánovas, G. V., Tapia, M. S., & Cano, M. P. (2005). *Novel food processing technologies*. CRC press.

Benítez, J. M. (2018). *Estadística aplicada al control de calidad alimentaria*. Ciudad de México: Editorial Alfaomega.

Conde, G. (2015). La distribución normal una rápida revisión histórica. *Heurística*, 17, 59-65.

Daniel, W. (1995). *Bioestadística. Base para el análisis de las ciencias de la salud*. México: Noriega Editores.

Dean, A., & Voss, D. (2019). *Design and analysis of experiments*. Springer.

Evans, M., Hastings, N., & Peacock, B. (2000). *Statistical distributions*. John Wiley & Sons.

Fellows, P. J. (2017). *Food processing technology: principles and practice*. Woodhead Publishing.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. Sage.

García, P. L. (2021). *Muestreo estadístico en la industria alimentaria*. Madrid: Editorial Paraninfo.

Gómez, A., & Ortega, L. (2021). *Introducción a la teoría de conjuntos*. Ediciones Académicas.

González, M. (2012). *Matemáticas combinatorias: Una introducción práctica*. Editorial Paraninfo.

Howell, D. C. (2017). *Statistical methods for psychology*. Cengage Learning.

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: principles and practice* (2nd ed.). OTexts.

López, J. (2023). *Fundamentos matemáticos: La teoría de conjuntos en perspectiva*. Editorial Matemática.

Martínez, J. y Martínez, I. (2011). *Distribución normal*. Valencia: Universidad de Valencia.

- Mendenhall, W., Beaver, R. J., & Beaver, B. M. (2010). *Introducción a la probabilidad y estadística*. Cengage Learning Editores.
- Montgomery, D. C. (2004). *Control estadístico de la calidad*. Limusa Wiley.
- Montgomery, D. C. (2021). *Introducción al control estadístico de la calidad*. Wiley.
- Montgomery, D. C., & Runger, G. C. (2016). *Probabilidad y estadística para ingenieros* (8ª ed.). Wiley.
- Montgomery, D. C., Runger, G. C., & Hubele, N. F. (2018). *Probabilidad y estadística para ingenieros* (9.ª ed.). McGraw-Hill.
- Montgomery, D. C., & Runger, G. C. (2010). *Applied statistics and probability for engineers*. John Wiley & Sons.
- López, R., & Torres, G. (2010). *Análisis combinatorio y probabilidad*. Pearson Educación.
- Moore, D. S., McCabe, G. P., & Craig, B. A. (2014). *Introduction to the practice of statistics*. W. H. Freeman.
- Moreno, P. (2022). *Conceptos básicos de teoría de conjuntos*. Editorial Universitaria.
- Ramírez, S. (2020). *Paradojas y fundamentos en la teoría de conjuntos*. Editorial Lógica.
- Walpole, R. E., Myers, R. H., Myers, S. L., & Ye, K. (2012). *Probabilidad y estadística para ingeniería y ciencias* (9ª ed.). Pearson Educación.
- Triola, M. F. (2019). *Estadística*. México: Pearson Educación.