INSTITUTO TECNOLÓGICO DE HERMOSILLO

REPORTE FINAL DEL PERIODO SABATICO

Periodo del 13 de agosto del 2018 al 12 de agosto del 2019.

ELABORACIÓN DE MATERIAL DIDACTICO

LIBRO DE TEXTO
PROBABILIDAD Y ESTADISTICA DESCRIPTIVA

PRESENTADO POR:

ING. JOSE ANTONIO VARGAS SANTOS

DICTAMEN:

AS-2-113/2018

HERMOSILLO, SONORA. AGOSTO 2019.

Dedicatoria

A dios por todo lo que me ha dado, gracias señor.

A la memoria de mis padres José Ramón Vargas Uribe que desde donde quiera que se encuentre en el cielo, sé que me sigue cuidando y guiando, a mi madre María Carmen Santos Flores por su apoyo en todo momento, por los valores que me inculcaron y por haberme dado la oportunidad de tener la mejor educación a su alcance en el transcurso de su vida. Pero, sobre todo, por tener un excelente ejemplo de vida a seguir.

En las siguientes líneas deseo expresar mi más profundo agradecimiento a las siguientes personas e institución, los cuales hicieron posible que cumpliera con otra de mis metas personales.

En primera instancia, quiero agradecer a mi familia por la paciencia y cariño compartidos durante todos estos años de trabajo, los cuales han permitido que me desarrollara profesionalmente y con esto haber conseguido cumplir una de mis metas profesionales y personales.

Finalmente me queda agradecer al Instituto Tecnológico de Hermosillo por brindarme la oportunidad de trabajar y desarrollar mi potencial académico en esta institución.

CONTENIDO

4. Muestreo

Introducción

El presente documento muestra una recopilación de temas que corresponden a la asignatura de Probabilidad y Estadística Descriptiva, para el programa de la carrera de ingeniería en gestión empresarial en el Instituto Tecnológico de Hermosillo, que tiene como objetivo elaborar un protocolo de investigación con el que se propongan soluciones científico - tecnológicas a problemáticas relacionadas con el contexto de la ingeniería, además buscar generar habilidades que le permitan averiguar y obtener conocimientos sobre determinados temas.

Los temas que se presentan están divididos en cuatro unidades las cuales están estructuradas de forma ordenada según el temario de la asignatura, con la finalidad que el estudiante cuente con una herramienta a la que tenga acceso de forma rápida y confiable.

Estos temas fueron tomados de diferentes fuentes bibliográficas, de forma sensata, buscando proporcional un texto sencillo y comprensible que fortalezca el conocimiento adquirido en el salón de clases.

Es importante mencionar que la antología busca el autoestudio por parte de los estudiantes, sin embargo, es solo un apoyo para documentarse ya que de forma directa deberá resolver sus dudas con el facilitador de la asignatura, para que se logre el objetivo de la materia.

Los temas están ordenados de la siguiente manera: en la primera unidad se observa las características de las distribuciones de frecuencia que permitirá diferenciarlas, así como la importancia de estas para la sociedad e innovación de las mismas.

En la segunda unidad, se presenta la introducción a la probabilidad y valor esperado, explicando los elementos que lo contienen, de igual forma que características deben de poseer para que sean válidos y concisos.

En la tercera unidad se presentan algunos elementos que permitirán integrar el protocolo, cuidando la redacción del mismo, así mismo se presentan algunos puntos que deben tomar en cuenta para la presentación escrita del libro.

Y, por último, en la cuarta unidad, se presenta algunos elementos que permitirán integrar el protocolo.

Unidad 1. Distribuciones de Frecuencias

1.1 Breve historia de la estadística.
La Estadística se ocupa de la obtención, organización y análisis de la información numérica que tiene cada vez un papel más importante en el mundo sumamente complejo de nuestros días. Todos los ciudadanos, sufren de tanta información de datos, que pueden verse incapaces de tomar decisiones inteligentes. Por tal razón se requieren de un análisis, para poder dar una respuesta o solución.

Se puede afirmar que la historia de la estadística comienza alrededor de 1749, aunque con el tiempo, ha habido cambios en la interpretación de la palabra “estadística”. En un principio, el significado estaba restringido a la información acerca de los estados. Este fue extendido posteriormente para incluir toda colección de información de cualquier tipo y más tarde fue extendido para incluir el análisis e interpretación de los datos. En términos modernos, “estadística” significa tantos conjuntos de información recopilada, por ejemplo: Registro de temperatura, contabilidad nacional, como trabajo analítico que requiera inferencia estadística.

Un gran número de conceptos de la estadística han tenido un importante impacto en un amplio rango de ciencias. Estos incluyen la secuencia en el desarrollo de las ideas que subyacen en la estadística moderna.

En el siglo XVIII, el término "estadística" designaba la colección sistemática de datos demográficos y económicos por los estados. A principios diseño de experimentos y enfoques a la inferencia estadística como la inferencia bayesiana, se puede considerar que tiene su propia del siglo XIX, el significado de "estadística" fue ampliado para incluir la disciplina ocupada de recolectar, resumir y analizar los datos. Hoy la estadística es ampliamente usada en el gobierno, los negocios y todas las ciencias. Las computadoras electrónicas han acelerado la estadística computacional y ha permitido a los estadísticos el desarrollo de métodos que usan recursos informáticos intensivamente.

El término "estadística matemática" designa las teorías matemáticas de la probabilidad e inferencia estadística, las cuales son usadas en la estadística aplicada. La relación entre estadística y probabilidades se fue desarrollando con el tiempo. En el siglo XIX, las estadísticas usaron de forma gradual la teoría de probabilidades, cuyos resultados iniciales fueron encontrados en los siglos XVII y XVIII, particularmente en el análisis de los juegos de azar (apuestas). Para 1800, la astronomía usaba modelos probabilísticos y teorías estadísticas, particularmente el método de los mínimos cuadrados, el cual fue inventado por Legendre y Gauss. La incipiente teoría de las probabilidades y estadísticas fue sistematizada y extendida por Laplace; después de este, las probabilidades y estadísticas han experimentado un continuo desarrollo. En el siglo XIX, el razonamiento estadístico y los modelos probabilísticos fueron usados por las ciencias sociales para el avance las nuevas ciencias de psicología experimental y sociología, y por las ciencias físicas en termodinámica y mecánica estadística. El desarrollo del razonamiento estadístico estuvo fuertemente relacionado con el desarrollo de la lógica inductiva y el método científico.

La estadística puede ser considerada no como una rama de las matemáticas, sino como una ciencia matemática autónoma, como las ciencias de la computación y la investigación de operaciones. A diferencia de las matemáticas, la estadística tuvo sus orígenes en la administración pública. Fue usada en la demografía y la economía. Con el énfasis en el aprendizaje de los datos y en la elaboración de las predicciones más acertadas, la estadística se ha solapado con la teoría de la decisión y la microeconomía. Con el enfoque de los datos, la estadística se ha solapado con la ciencia de la información y las ciencias de la computación.

1.2 CONCEPTOS DE ESTADISTICA Y SU CLASIFICACIÓN

Estadística
La estadística es la ciencia que se ocupa de la ordenación y análisis de datos procedentes de muestra y de la realización de inferencias sobre las poblaciones de las que estas proceden.

Generalmente se pueden distinguir dos fases en la realización de cualquier experimento o estudio científico. La primera consiste en la observación y análisis de los hechos que acontecen (entrega de información y colecciones de datos), y la segunda de la interpretación y obtención de conclusiones.

La estadística es una ciencia matemática que trata de la recolección, clasificación y presentación de los hechos sujetos a una apreciación numérica y se utiliza para describir, analizar e interpretar ciertas características de un fenómeno o conjunto de individuos llamado población.

La estadística es la rama de las matemáticas que se ocupa de reunir, organizar, analizar los datos numéricos para la deducción de soluciones y la toma de decisiones razonables.

La estadística es el conjunto de medios aplicados para recopilar, procesar y analizar información que se obtiene de una muestra, para inferir las características o parámetros de una población o de un problema determinado. Esta información puede ser usada para conocer a fondo problemáticas basándose en pequeños datos.

Clasificación de la estadística
El estudio de la estadística se ha concretado primordialmente en el análisis de datos y su aplicación en la toma de decisiones, lo que ha permitido dividir a la estadística por su aplicación en:

Estadística Descriptiva
La Estadística descriptiva se emplea para describir las características básicas de los datos en estudio. Por medio de tablas, representaciones gráficas y valores numéricos (media, mediana, rango, desviación estándar, etc.), se analizan y se dan resúmenes de los datos de que se dispone. Por lo tanto, suele decirse que la estadística descriptiva es una serie de métodos para organizar, resumir, presentar y analizar datos de manera eficiente e informativa.

La estadística descriptiva es la técnica matemática que obtiene, organiza, presenta y describe un conjunto de datos con el propósito de facilitar su uso generalmente con el apoyo de tablas, medidas numéricas o gráficas.

Clasificación de la estadística
El estudio de la estadística se ha concretado primordialmente en el análisis de datos y su aplicación en la toma de decisiones, lo que ha permitido dividir a la estadística por su aplicación en:

Además, calcular los parámetros estadísticos como las medidas de centralización y de dispersión que describen el conjunto estudiado. (por ejemplo: las edades de una población, la altura de los estudiantes, la temperatura en los meses de verano, etc.), con el fin de describir apropiadamente las diversas características de ese conjunto.

Es la ciencia encargada de analizar y describir la totalidad de individuos de una población.

Esta incluye las técnicas que se relacionan con el resumen y descripción de datos numéricos, estos pueden ser gráficos o pueden incluir análisis mediante cálculos.

La Estadística descriptiva es la herramienta para el manejo de los datos y proporciona métodos para resumirlos y organizarlos. Describe, analiza y representa un grupo de datos utilizando métodos numéricos y gráficos que resumen y presentan la información contenida en ellos. Es la primera fase de toda investigación.

Estadística Inferencial
La Estadística Inferencial es una parte de la estadística que comprende los métodos y procedimientos que por medio de la inducción determina propiedades de una población estadística, a partir de una pequeña parte de esta. Su objetivo es obtener conclusiones útiles para hacer deducciones sobre una totalidad, basándose en la información numérica.

Esquema de una Situación estadística
La estadística es una rama de las matemáticas que contribuye a otras ciencias con métodos que hacen posibles conclusiones correctas con una precisión deseada a partir de datos numéricos que presentan variación y que se asocian a una o varias

variables. Por tanto, la naturaleza y la calidad de los datos hacen factible incrementar la certidumbre en las conclusiones estadísticas.

La estadística es la ciencia de:
1. La recopilación
2. La organización
3. El análisis
4. La extracción de conclusiones a partir de datos.

1.3 RECOPILACIÓN DE DATOS

La Recopilación de datos se refiere al uso de una gran diversidad de técnicas y herramientas que pueden ser utilizadas por el analista para desarrollar los sistemas de información, los cuales pueden ser las entrevistas, las encuestas, el cuestionario, la observación, el diagrama de flujo y el diccionario de datos.

Todos estos instrumentos se aplicarán en un momento en particular, con la finalidad de buscar información que será útil a una investigación en común. En la presente investigación trata con detalle los pasos que debe seguir en el proceso de recopilación de datos con las técnicas ya antes nombrados.

Los analistas utilizan una variedad de métodos a fin de recopilar los datos sobre una situación existente, como entrevistas, inspección de registros (revisión en el sitio) y observación.

La noción de “recopilación” refiere al proceso y el resultado de recolectar (reunir, recoger o cosechar algo). Un “dato” por su parte, una información que permite generar un cierto conocimiento.

Esto quiere decir que la recopilación de datos es la actividad que consiste en la recopilación de información dentro de un cierto contexto.

Dentro de la recopilación de datos se pueden escoger diversas técnicas: Las encuestas, la observación, la toma de muestras y las entrevistas, entre otras permiten realizar la tarea. De acuerdo al tipo de datos, la persona utilizara distintos instrumentos (grabadora de audio, cámara de fotos, etc.)

1.4 DISTRIBUCIÓN DE FRECUENCIAS

En estadística, se le llama distribución de frecuencias a la agrupación de datos en categorías que indican el número de observaciones en cada categoría. Esto proporciona un valor añadido a la agrupación de datos. La distribución de frecuencias presenta las observaciones clasificadas de modo que se pueda ver el número existente en cada clase.

Se pueda ver el número existente en cada clase.

Se dispone de varios diagramas para representar gráficamente una distribución de frecuencias: histogramas de frecuencias, polígonos de frecuencias y polígonos de frecuencias acumuladas u ojivas. Dada la naturaleza de las tablas de distribución de frecuencias, los datos que se manejan corresponden a la escalas de nivel de intervalo y de razón.

Las distribuciones de frecuencias son tablas en que se dispone las modalidades de la variable por filas. En las columnas se dispone el número de ocurrencias por cada valor, porcentajes, etc. La finalidad de las agrupaciones en frecuencias es facilitar la obtención de la información que contienen los datos.

Ejemplo: Quieren conocer si un grupo de individuos está a favor o en contra de la exhibición de imágenes violentas por televisión, para lo cual han recogido los siguientes datos:

X: 2,1,5,3,3,2,3,1,4,2,4,2,3,2,3,4,3,3,1,2

(Regla de codificación:
1= En contra.
2= Bastante en contra.
3= Indiferente.
4= Bastante a favor.
5= A favor.)

La inspección de los datos originales no permite responder fácilmente a cuestiones como cuál es la actitud mayoritaria del grupo, y resulta bastante más difícil determinar la magnitud de la diferencia de actitud entre hombres y mujeres.

Podemos hacernos mejor idea si disponemos en una tabla los valores de la variable acompañados del número de veces (la frecuencia) que aparece cada valor:

X	f
1	3
2	6
3	7
4	3
5	1
Total	20

X: Símbolo genérico de la variable.

f: Frecuencia (también se simboliza como ni).

La distribución de frecuencias de los datos del ejemplo muestra que la actitud mayoritaria de los individuos del grupo estudiado es indiferente.

La interpretación de los datos ha sido facilitada porque se ha reducido el número de números a examinar (en vez de los 20 datos originales, la tabla contiene 5 valores de la variable y 5 frecuencias).

Generalmente las tablas incluyen varías columnas con las frecuencias relativas (son el número de ocurrencias dividido por el total de datos, y se simbolizan "fr" o "pi"), frecuencias acumuladas (la frecuencia acumulada es el total de frecuencias de los valores iguales o inferiores al de referencia, y se simbolizan "fa" o "na". No obstante, la frecuencia acumulada también es definida incluyendo al valor de referencia), frecuencias acumuladas relativas (la frecuencia acumulada relativa es el total de frecuencias relativas de los valores iguales o inferiores al de referencia, y se simbolizan "fr" o "pa")

Ejemplo: Consideremos el siguiente grupo de datos:

18,35,22,41,35,68,30,30,30,46,42,32,30,16,28,35,35,35,44,44,44,39,44,61,55,32,32,28, 28,29,25,25,28,54,53,35,60,35,35,35,64,35,35,34,22,44,17,16,46,46,27,25,46,47,46,35, 39,59,59,32,32,28,35,27,31,30,32,61,35,54,57,35,56,44,58,41,42,44,30,40,46,46,50,49, 50,36,41,29.

La distribución de frecuencias es:

X	f
16	2
17	1
18	1
22	2
25	3
27	2
28	5
29	2
30	7
31	1
32	7
33	1
34	1
35	16
36	2
39	2
40	1
41	5
42	2
44	7
45	1
46	7
47	1
49	1
50	2
53	1
54	3

La reducción de datos mediante el agrupamiento en frecuencias no facilita su interpretación: La tabla es demasiado grande. Para reducir el tamaño de la tabla agrupamos los valores en intervalos, y las frecuencias son las de los conjuntos de valores incluidos en los intervalos:

Distribución de frecuencias agrupadas en intervalos

X_i	f_i	fr_i	f_a	fr_a
64-69	2	0.02	100	1.00
58-63	8	0.08	98	0.98
52-57	7	0.07	90	0.90
46-51	11	0.11	83	0.83
40-45	16	0.16	72	0.72
34-39	22	0.22	56	0.56
28-33	21	0.21	34	0.34
22-27	9	0.09	13	0.13
16-21	4	0.04	4	0.04
	100	1.00		

Ahora es más sencillo interpretar los datos. Por ejemplo, podemos apreciar inmediatamente que el intervalo con mayor número de datos es el 34-39, o que el 75% de los datos tiene valor inferior a 46.

Este tipo de tabla es denominado "tabla de datos agrupados en intervalos".

Elementos básicos de las tablas de intervalos:

Intervalo: Cada uno de los grupos de valores de la variable que ocupan una fila en una distribución de frecuencias

Límites aparentes: Valores mayor y menor del intervalo que son observados en la tabla. Dependen de la precisión del instrumento de medida. En el ejemplo, los límites aparentes del intervalo con mayor número de frecuencias son 34 y 39.

Límites exactos: Valores máximo y mínimo del intervalo que podrían medirse si se contara con un instrumento de precisión perfecta. En el intervalo 34-39, estos límites son 33.5 y 39.5.

Punto medio del intervalo (Marca de clase): Es la Suma de los límites dividido por dos.

Amplitud del intervalo: Diferencia entre el límite exacto superior y el límite exacto inferior. En el ejemplo es igual a 6.

Tipos de Frecuencias

1. Frecuencia completa
2. Frecuencia relativa
3. Frecuencia acumulada
4. Frecuencia relativa acumulada
5. Distribución de frecuencias agrupadas

Frecuencia Completa

La frecuencia completa por su denominación es el número de veces que aparece un determinado valor en un valor estadístico. Se representa por fila.

La suma de la frecuencia completa es igual al número total de datos, que se representa por N.

$$f1 + f2 + f3 + \ldots\ldots + fn = N$$

Para indicar resumidamente estas sumas se utiliza la letra griega Σ (que se lee sumatoria).

$$\sum_{i=1}^{n} f_i = N$$

Frecuencia Relativa

Se dice que la frecuencia relativa es el cociente entre la frecuencia absoluta de un determinado valor y el número total de datos. Se puede expresar en tantos por ciento y se representa por ni.

La suma de las frecuencias relativas es igual a 1.

La Frecuencia Relativa (ni) es el cociente entre la frecuencia absoluta y el tamaño de la muestra (N). Es decir:

$$n_i = \frac{f_i}{N}$$

Siendo el fi para todo el conjunto i. Se presenta en una tabla o nube de puntos en una distribución de frecuencias.

Si multiplicamos la frecuencia relativa por 100 obtendremos el porcentaje o tanto por ciento.

La Frecuencia Relativa Acumulada. Es el cociente entre la frecuencia acumulada de un determinado valor y el número total de datos. Se puede expresar en tantos

por ciento. Ejemplo: Durante el mes de julio, en una ciudad se han registrado las siguientes temperaturas máximas:
32, 31, 28, 29, 33, 32, 31, 30, 31, 31, 27, 28, 29, 30, 32, 31, 31, 30, 30, 29, 29, 30, 30, 31, 30, 31, 34, 33, 33, 29, 29

Frecuencia Acumulada

La frecuencia acumulada es el resultado de sumar sucesivamente las frecuencias absolutas o relativas, desde el menor al mayor de sus valores.

Para calcular la frecuencia acumulada hay que ordenar los datos de menor a mayor. Para un cálculo más sencillo y una imagen más visual, estos se colocan en una tabla. Tras tener los datos ordenados y tabulados, la frecuencia acumulada se obtiene simplemente de ir sumando una clase o grupo de la muestra con la anterior (primer grupo + segundo grupo, primer grupo + segundo grupo + tercer grupo y así sucesivamente hasta llegar a acumular del primer grupo al último).

Tipos de frecuencias acumuladas.
Existen dos tipos de frecuencia acumulada, la absoluta y la relativa:

Frecuencia Absoluta Acumulada

La frecuencia absoluta nos da información acerca de la cantidad de veces que se repite un suceso al realizar un número determinado de experimentos aleatorios. Para hallar la frecuencia absoluta acumulada, no tendríamos más que acumular las frecuencias absolutas. Esta se denomina con las letras Fi.

Supongamos que las notas de 20 alumnos de primer curso de economía son las siguientes:

1,2,8,5,8,3,8,5,6,10,5,7,9,4,10,2,7,6,5,10.

Para hallar la frecuencia absoluta acumulada, en primer lugar, se ordenan los datos de menor a mayor, se tabulan y posteriormente se acumulan.

Frecuencia Relativa Acumulada

La frecuencia relativa se calcula como el cociente de la frecuencia absoluta de algún valor de la población/muestra (fi) entre el total de valores que componen la población/muestra (N). Para hallar la frecuencia relativa acumulada, no tendríamos más que ir acumulando las frecuencias relativas. Esta se denomina con las letras Hi. Supongamos que las notas de 20 alumnos de primer curso de economía son las siguientes:
1,2,8,5,8,3,8,5,6,10,5,7,9,4,10,2,7,6,5,10.
Por tanto, tenemos:
Xi = Variable aleatoria estadística, nota del examen de primer curso de economía.
N = 20
fi = Número de veces que se repite el suceso (en este caso, la nota del examen).
Hi = Proporción que representa el valor i-ésimo en la muestra.

A resaltar es que el total de acumular las frecuencias absolutas, tiene que coincidir con el total de la muestra. Esta es una buena forma de comprobar que se ha calculado correctamente.

Xi	Fi	Fi
1	1	1
2	2	3
3	1	4
4	1	5
5	4	9
6	2	11
7	2	13
8	3	16
9	1	17
10	3	20
∑	20	

	Fi	Hi	Hi
1	1	5%	5%
2	2	10%	15%
3	1	5%	20%
4	1	5%	25%
5	4	20%	45%
6	2	10%	55%
7	2	10%	65%
8	3	15%	80%
9	1	5%	85%
10	3	15%	100%

La frecuencia acumulada es la suma de las frecuencias absolutas de todos los valores inferiores o iguales al valor considerado.

Se representa por fi.

Distribución de Frecuencias Agrupadas

La distribución de frecuencias agrupadas o tabla con datos agrupados se emplea si las variables toman un número grande de valores o la variable es continua. Se agrupan los valores en intervalos que tengan la misma amplitud denominados clases. A cada clase se le asigna su frecuencia correspondiente. Límites de la clase. Cada clase está delimitada por el límite inferior de la clase y el límite superior de la clase.

La amplitud de la clase es la diferencia entre el límite superior e inferior de la clase. La marca de clase es el punto medio de cada intervalo y es el valor que representa a todo el intervalo para el cálculo de algunos parámetros. En caso de que el primer intervalo sea de la forma (-∞, k], o bien [k,+∞) donde k es un número cualquiera, en el caso de (-∞, k], para calcular la marca de clase se tomará la amplitud del intervalo adyacente al (ai+1), y la marca de clase será ((k-ai+1) +k)/2. En el caso del intervalo [k,+∞) también se tomará la amplitud del intervalo adyacente al (ai-1) siendo la marca de clase ((k+ai-1)+k)/2.

Construcción de una tabla de datos agrupados:
3, 15, 24, 28, 33, 35, 38, 42, 43, 38, 36, 34, 29, 25, 17, 7, 34, 36, 39, 44, 31, 26, 20, 11, 13, 22, 27, 47, 39, 37, 34, 32, 35, 28, 38, 41, 48, 15, 32, 13.

Se localizan los valores menor y mayor de la distribución. En este caso son 3 y 48.

Se restan y se busca un número entero un poco mayor que la diferencia y que sea divisible por el número de intervalos que queramos establecer.

Es conveniente que el número de intervalos oscile entre 6 y 15.

En este caso, 48 - 3 = 45, incrementamos el número hasta 50: 5 = 10 intervalos.

Se forman los intervalos teniendo presente que el límite inferior de una clase pertenece al intervalo, pero el límite superior no pertenece al intervalo, se cuenta en el siguiente intervalo.

Intervalo	c_i	n_i	N_i	f_i	F_i
[0, 5)	2.5	1	1	0.025	0.025
[5, 10)	7.5	1	2	0.025	0.050
[10,15)	12.5	3	5	0.075	0.125
[15, 20)	17.5	3	8	0.075	0.200
[20, 25)	22.5	3	11	0.075	0.275
[25, 30)	27.5	6	17	0.150	0.425
[30, 35)	32.5	7	24	0.175	0.600
[35, 40)	37.5	10	34	0.250	0.850
[40, 45)	42.5	4	38	0.100	0.950
[45, 50)	47.5	2	40	0.050	1
Total:		40		1	

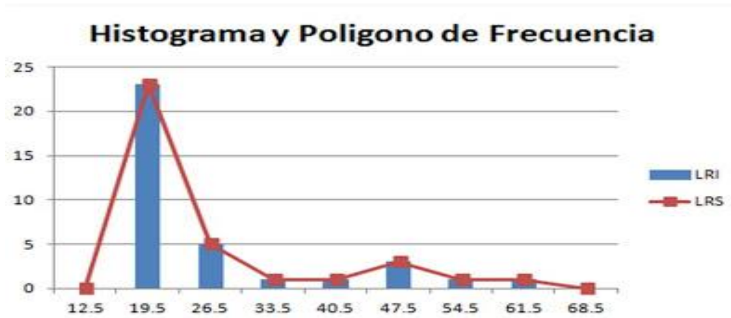
1.4.1 POLÍGONOS DE FRECUENCIA, HISTOGRAMA Y OJIVAS

Polígono de Frecuencia

El polígono de una variable cuantitativa corresponde al diagrama de líneas. Se construye a partir del histograma de frecuencias.

Para esto, se unen los puntos medios de cada una de las barras con un segmento recta. Para que la gráfica sea un polígono, es necesario construir un segmento de recta. Para que la gráfica sea un polígono, es necesario construir un segmento de recta que inicie en el eje horizontal y termine en el punto medio de barra. De forma similar, se debe hacer en la última barra que la figura que se obtenga sea cerrada.

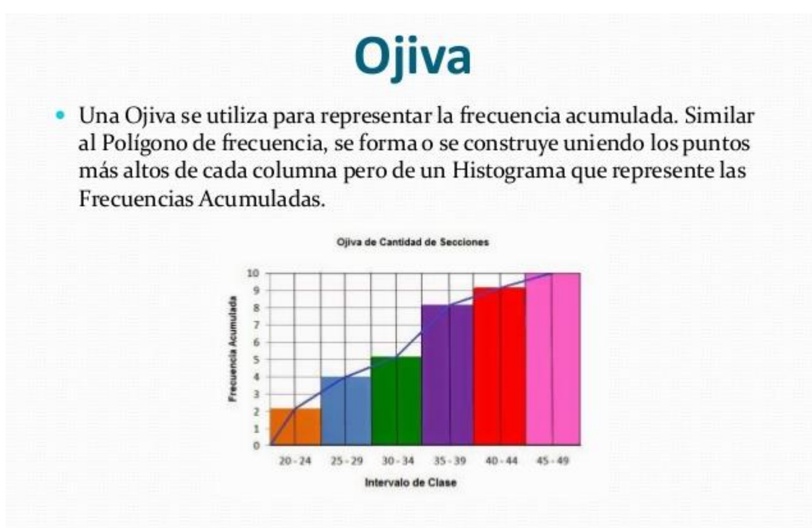
Para construir el polígono de frecuencia se toma la marca de clase que coincide con el punto medio de cada rectángulo de un histograma.

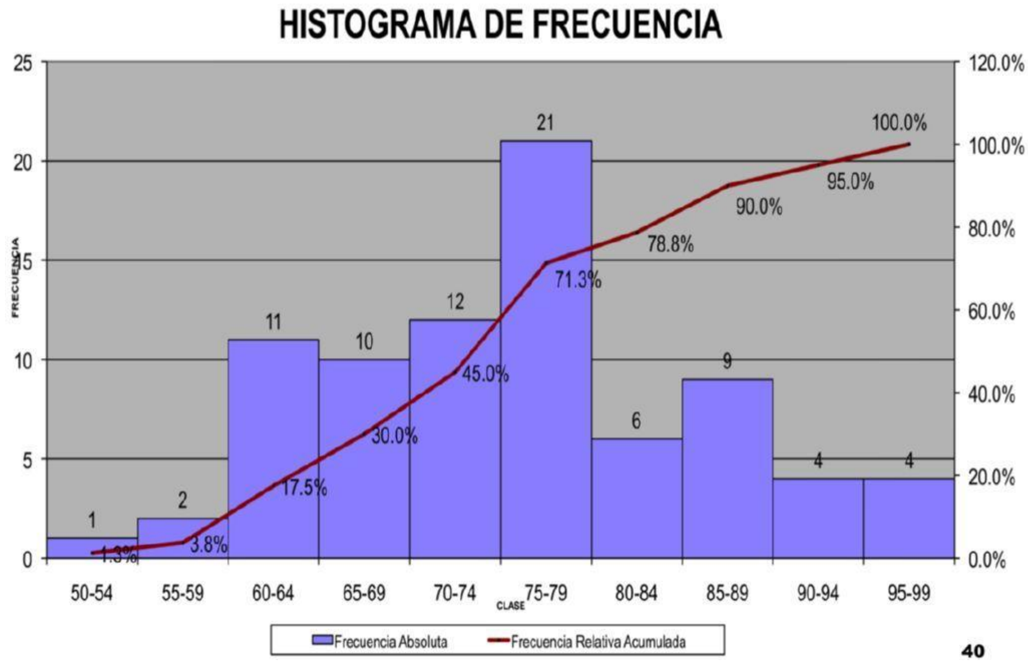


HISTOGRAMA

Un Histograma es un tipo especial de gráfica de barras que despliega la variabilidad dentro de un proceso, también toma datos variables (tales como alturas, pesos, densidades, tiempo, temperaturas, etc.) y despliega su distribución. Un histograma es una representación gráfica de una variable en forma de barras, donde la superficie de cada barra es proporcional a la frecuencia de los valores representados. En el eje vertical se representan las frecuencias y en el eje horizontal los valores de las variables, normalmente señalando las marcas de clase, es decir, la mitad del intervalo en el que están agrupados los datos.

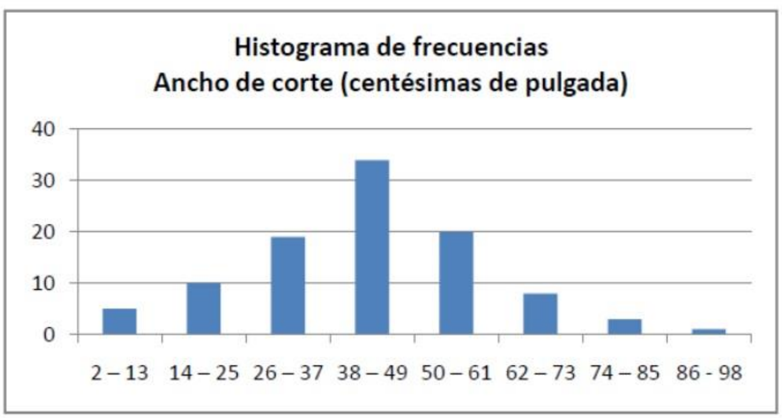
Un Histograma es un tipo especial de gráfica de barras que despliega la variabilidad dentro de un proceso, también toma datos variables (tales como alturas, pesos, densidades, tiempo, temperaturas, etc.) y despliega su distribución. Un histograma es una representación gráfica de una variable en forma de barras, donde la superficie de cada barra es proporcional a la frecuencia de los valores representados. En el eje vertical se representan las frecuencias, y en el eje horizontal los valores de las variables, normalmente señalando las marcas de clase, es decir, la mitad del intervalo en el que están agrupados los datos.



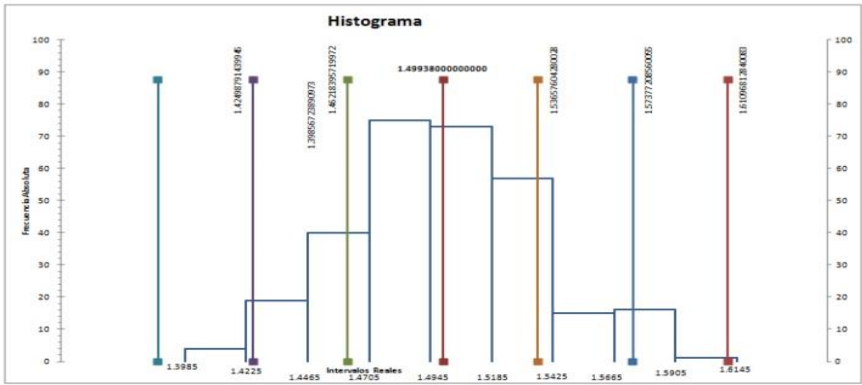

HISTOGRAMA DE FRECUENCIA
FRECUENCIA
CLASE
50-54
55-59
60-64
65-69
70-74
75-79
80-84
85-89
90-94
95-99
1
2
11
10
12
21
6
9
4
4
1.3%
3.8%
17.5%
30.0%
45.0%
71.3%
78.8%
90.0%
95.0%
100.0%
Frecuencia Absoluta
Frecuencia Relativa Acumulada

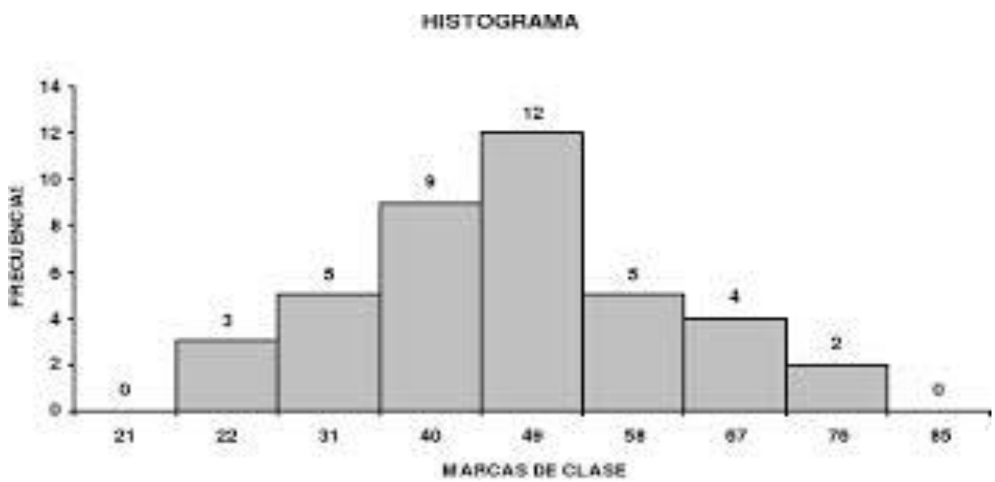
40

[redacted]


Histograma de frecuencias
Ancho de corte (centésimas de pulgada)
2 – 13
14 – 25
26 – 37
38 – 49
50 – 61
62 – 73
74 – 85
86 - 98

Histograma



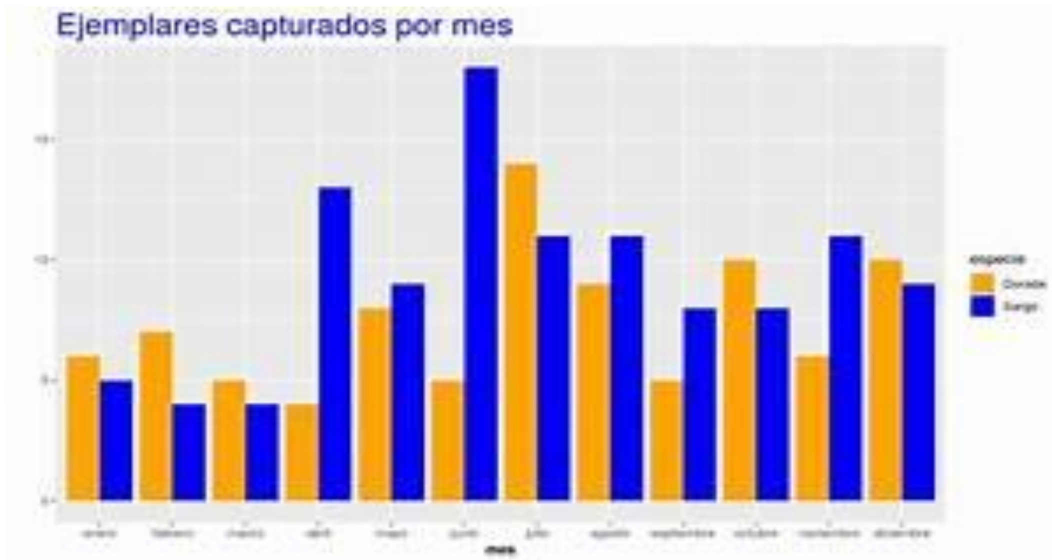


<u>Tipos de Histogramas</u>
Diagrama de Barras Simples

Representa la frecuencia simple (absoluta o relativa) mediante la altura de la barra la cual es proporcional a la frecuencia simple de la categoría que representa.

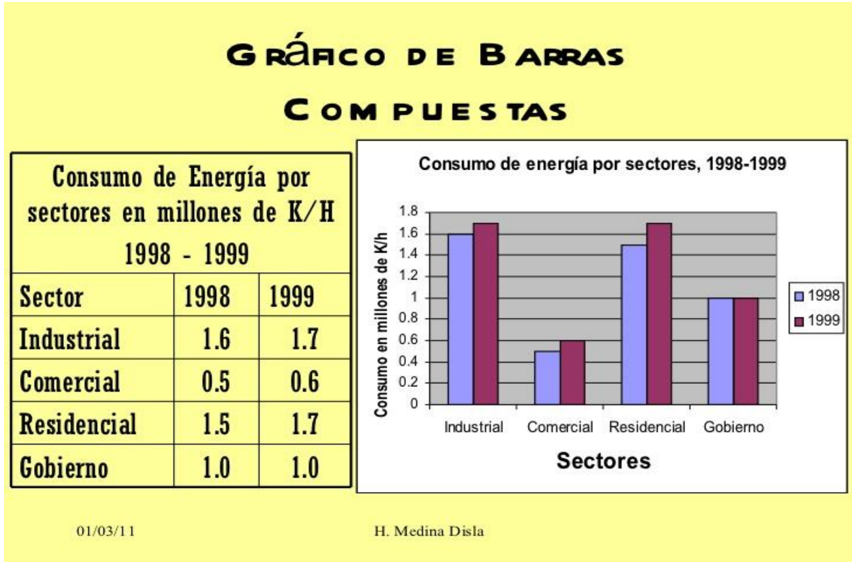
Un diagrama de barras, también conocido como gráfico de barras o diagrama de columnas, es una forma de representar gráficamente un conjunto de datos o valores, y está conformado por barras rectangulares de longitudes proporcionales a los valores representados. Los gráficos de barras son usados para comparar dos o más valores. Las barras pueden orientarse horizontal o verticalmente.

Existen evidencias de este tipo de diagramas desde hace más de 300 años.



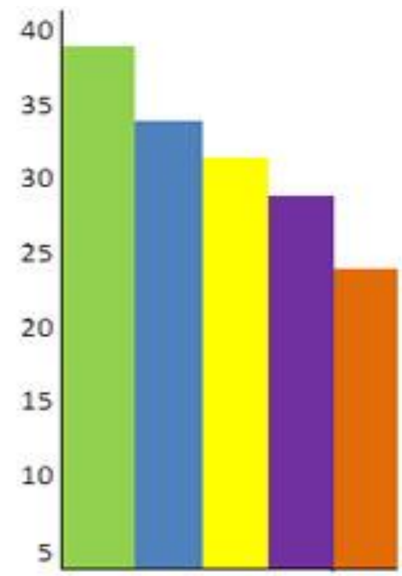
<u>Diagrama de Barras Compuestas</u>

Se usa para representar la información de una tabla de doble entrada o sea a partir de dos variables, las cuales se representan así; la altura de la barra representa la frecuencia simple de las modalidades o categorías de la variable y esta altura es proporcional a la frecuencia simple de cada modalidad.



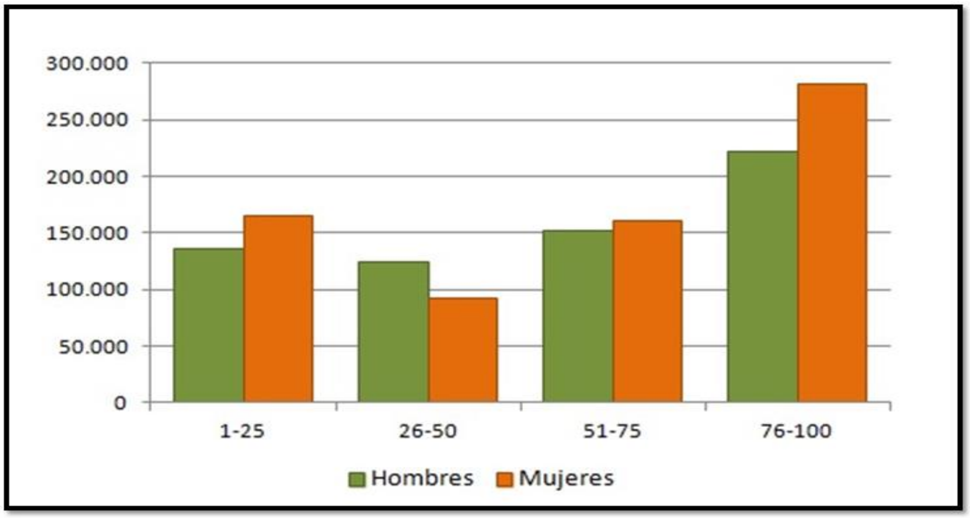
Consumo de Energía por sectores en millones de K/H 1998 - 1999		
Sector	1998	1999
Industrial	1.6	1.7
Comercial	0.5	0.6
Residencial	1.5	1.7
Gobierno	1.0	1.0

CURSOS	F
segundo ´´A´´	38
octavo ´´D´´	34
primero ´´B´´	31
tercero ´´B´´	29
noveno ´´A´´	25



<u>Diagrama de Barras Agrupadas</u>

Se usa para representar la información de una tabla de doble entrada o sea a partir de dos variables, el cual es representado mediante un conjunto de barras como se clasifican respecto a las diferentes modalidades.



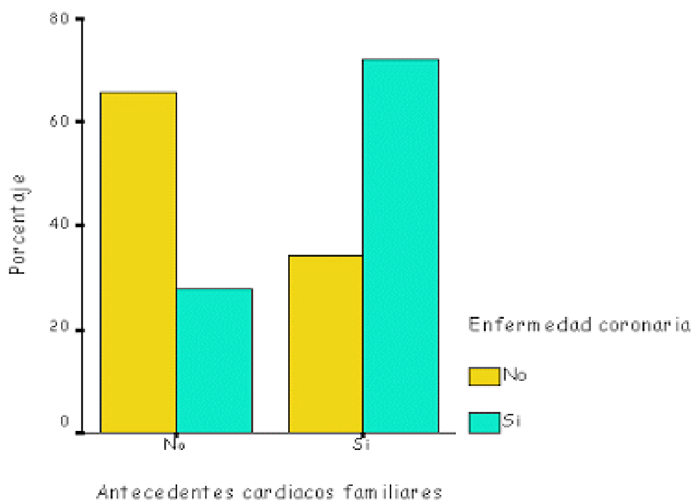


Diagrama de Barras Agrupadas o Apiladas

Al igual que el diagrama de barras compuestas, un diagrama de barras agrupadas se utiliza para presentar la información contenida en una tabla de doble entrada; es decir, se utiliza para representar la información obtenida a partir de la medición de dos variables.

Este tipo de grafico se muestra, mediante un conjunto de barras, como se clasifican los sujetos respecto a las distintas modalidades o categorías de una variable dentro de cada modalidad o categoría de la otra variable.

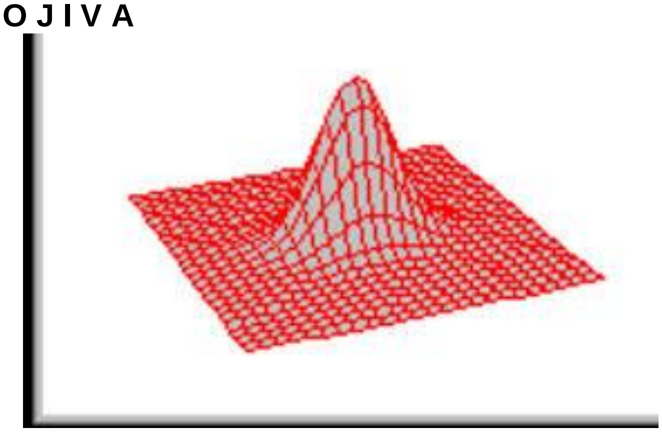
Ojivas

La representación gráfica de un cuadro de frecuencia acumulada son curvas llamadas ojivas. En la gráfica de ojiva, el último intervalo no se une con el eje horizontal.

La ojiva apropiada para información que presente frecuencias mayores que el dato que se está comparando tendrá una pendiente negativa (hacia abajo y a la derecha)

y en cambio la que se asigna a valores menores, tendrá una pendiente positiva. Una gráfica similar al polígono de frecuencias es la ojiva, pero ésta se obtiene de aplicar parcialmente la misma técnica a una distribución acumulativa y de igual manera que éstas, existen las ojivas mayores que y las ojivas menores que.
En estadística, la ojiva es un polígono frecuencial acumulado, es decir, que permite ver cuántas observaciones se encuentran por encima o debajo de ciertos valores, en lugar de solo exhibir los números asignados a cada intervalo.
La ojiva apropiada para información que presente frecuencias mayores que el dato que se está comparando tendrá va y de igual manera que éstas, existen las ojivas "mayor que" y las ojivas "menor que".
Existen dos diferencias fundamentales entre las ojivas y los polígonos de frecuencias (y por esto la aplicación de la técnica es parcial):
Un extremo de la ojiva no se toca al eje horizontal, para la ojiva "mayor que" sucede con el extremo izquierdo; para la ojiva "menor que", con el derecho.
En el eje horizontal, en lugar de colocar las marcas de clase, se colocan las fronteras de clase. Para el caso de la ojiva "mayor que" es la frontera menor; para la ojiva menor que, la mayor.
La ojiva "mayor que" se le denomina de esta manera porque viendo el aspire que está sobre el límite superior se ven las frecuencias que tienen por encima de ese límite superior. De forma análoga, en la ojiva "menor que" la frecuencia que se representa en cada frontera de clase son el número de observaciones menores que la frontera señalada (en caso de tiempos sería el número de observaciones antes de la hora que señala la frontera).[1]

O J I V A



1.5 MEDIDAS DE TENDENCIA CENTRAL PARA UN CONJUNTO DE DATOS NO AGRUPADOS Y DATOS AGRUPADOS

La mayor parte de los conjuntos de datos muestra una tendencia bien determinada a agruparse o aglomerarse alrededor de cierto punto central. Así para cualquier conjunto especifico de datos, casi siempre se puede seleccionar algún valor típico o promedio, para describir todo el conjunto. Este valor típico descriptivo es una medida de tendencia central o ubicación

Un promedio es un valor típico o representativo de un conjunto de datos. Como tales valores pueden situarse hacia el centro del conjunto de datos ordenados por magnitud, los promedios se conocen como medidas de tendencia central.

Las medidas de tendencia central son valores numéricos que señalan un tipo de centro de un conjunto de datos, centro que se emplea para representar el conjunto. Las más importantes son: la media aritmética, la mediana, la moda, la media ponderada y la media geométrica. El término genérico para ellas es el promedio. Cada una tiene sus ventajas y desventajas, según sus datos y el objetivo perseguido.

La tendencia central se refiere al punto medio de una distribución.
Las medidas estadísticas se agrupan en dos categorías: Las de tendencia central, las de dispersión y asimetría.

Las medidas de tendencia central señalan en donde están la mayoría de los datos; mientras que las medidas de dispersión, indican que tan esparcidos se encuentran estos datos con respecto a los valores centrales.

Quizá la medida numérica de tendencia central que se usa más ampliamente sea la media aritmética, a la cual se le llamara simplemente media. Así, la estadística descriptiva y la inferencial se sustentan básicamente en la media; por lo tanto, comenzaremos con las medidas de tendencia central, las cuales permiten caracterizar y distinguir la información acumulada.

1.5.1 MEDIA ARITMÉTICA PARA DATOS NO AGRUPADOS.
Equivale al cálculo del promedio simple de un conjunto de datos. Para diferenciar datos muéstrales de datos poblacionales, la media aritmética se representa con un símbolo para cada uno de ellos: Si los datos corresponden a una muestra, la media se representa por X, (se lee “x testada”); si trabajamos con una población, se utiliza la letra griega μ (se pronuncia “mu”).

Una razón del uso tan amplio de la media aritmética es que su cálculo es muy sencillo. Pero son sus cualidades matemáticas y estadísticas las que la convierten en el promedio más adecuado para la estimación o inferencia de parámetros a partir de una muestra. La media aritmética (también llamada media) es el promedio o medida de tendencia central que se utiliza con mayor frecuencia. Se calcula con la suma de todas las observaciones en un conjunto de datos, dividida entre el número de elementos involucrados. Así, para una muestra que contiene “n” observaciones, X1, X2, X3,…………Xn, la media aritmética se define por la fórmula:

$$X = \frac{X_1 + X_2 + X_3 + \ldots\ldots + X_n}{n}$$

Otra notación muy convencional para expresar el procedimiento de la media es utilizando el símbolo de sumatoria (Σ). La expresión que aparece en la mayoría de los libros de estadística para el cálculo de la media es:

$$\bar{X} = \sum_{i=1}^{n} \frac{X_i}{n}$$

donde:

X es la media aritmética de la muestra

n es el tamaño de la muestra o el número de elementos de la muestra

X_i es la i-esima observación de la variable x de la muestra

$i=1$ es el valor numérico 1, 2, 3,………….n

ΣX_i es la suma de todos los valores xi de la muestra

Ejemplos:

1. Las calificaciones de un estudiante en sus exámenes fueron: 84, 91, 72, 68, 87 y 78, encontrar la media aritmética?

Solución:

$$X = \frac{X_1 + X_2 + X_3 + \ldots\ldots + X_n}{n} = \sum_{i=1}^{n} \frac{X_i}{n}$$

$$X = \frac{84+91+72+68+87+78}{6} = \frac{480}{6} = 80$$

2. Las medidas del diámetro de un cilindro fueron: 3.88, 4.09, 3.92, 3.97, 4.02, 3.95, 4.03, 3.92, 3.98 y 4.06 centímetros, encontrar la media aritmética?

   Solución:

$$X = \frac{X_1 + X_2 + X_3 + \ldots\ldots + X_n}{n} = \sum_{i=1}^{n} \frac{X_i}{n}$$

$$X = \frac{3.88+4.09+3.92+3.97+4.02+3.95+4.03+3.92+3.98+4.06}{10} = \frac{39.82}{10} = 3.98$$

3. Los datos siguientes describen los días que los generadores de una planta de energía, se encuentra fuera de servicio debido a mantenimiento normal o falla. Encontrar la media aritmética de los generadores del 1 al 10, de los días fuera de servicio: 7, 23, 4, 8, 2, 12, 6, 13, 9 y 4.

Solución:

$$X = \frac{X_1 + X_2 + X_3 + \ldots\ldots + X_n}{n} = \sum_{i=1}^{n} \frac{X_i}{n}$$

$$X = \frac{7+23+4+8+2+12+6+13+9+4}{10} = \frac{88}{10} = 8.8 \text{ días}$$

4. Se observan las edades en años de 6 voluntarios en la cruz roja del poblado: 19, 23, 25, 43, 35 y 23. Encontrar la media aritmética para datos no agrupados.

Solución:

$$X = \frac{X_1 + X_2 + X_3 + \ldots\ldots + X_n}{n} = \sum_{i=1}^{n} \frac{X_i}{n}$$

$$X = \frac{19+23+25+43+35+23}{6} = \frac{168}{6} = 28 \text{ años}$$

5. Los siguientes datos representan las edades de 10 mujeres que se presentaron en una clínica hacerse un estudio por primera vez, 29, 30, 38, 33, 40, 46, y 41, determinar la edad promedio de este grupo de mujeres.

Solución:

$$X = \frac{X_1 + X_2 + X_3 + \ldots\ldots + X_n}{n} = \sum_{i=1}^{n} \frac{X_i}{n}$$

$$X = \frac{29+30+38+33+40+37+42+40+46+41}{10} = \frac{376}{10} = 37.6 \text{ años}$$

6. Calcular la media de los rendimientos porcentuales totales a un año de los fondos de acciones generales nacionales presentados con los siguientes datos: 32.2, 29.5, 29.9, 32.4, 30.5, 30.1, 32.1, 35.2, 10.0, 20.6, 28.6, 30.5, 38.0, 33.0, 29.4, 37.1, 28.6

Solución:

$$X = \frac{X_1 + X_2 + X_3 + \ldots\ldots + X_n}{n} = \sum_{i=1}^{n} \frac{X_i}{n}$$

$$X = \frac{32.2+29.5+29.9+\ldots\ldots+28.6}{17} = \frac{507.7}{17} = 29.86$$

7. Hemos tomado una muestra aleatoria de 35 cajas y registrado el número de jeringas contenidas en cada caja. En la siguiente tabla ilustramos nuestros resultados obtenidos a partir de una muestra. Encontrar la media de las jeringas.

101	103	112	102	98	97	93
105	100	97	107	93	94	97
97	100	110	106	110	103	99
93	98	106	100	112	105	100
114	97	110	102	98	112	99

Solución:

$$X = \frac{X_1 + X_2 + X_3 + \ldots\ldots + X_n}{n} = \sum_{i=1}^{n} \frac{X_i}{n}$$

$$X = \frac{101+105+97+93+114+\ldots\ldots+99}{35} = \frac{3{,}570}{35} = 102 \text{ jeringas}$$

8. Hay que considerar que se tienen 5 calificaciones de María en un examen de física: 7, 10, 6, 10 y 7. La calificación final para María empleando la media seria?

Solución:

n

$$X = \frac{X_1 + X_2 + X_3 + \ldots\ldots.. + X_n}{n} = \Sigma_{i=1} \frac{X_i}{n}$$

$$X = \frac{7+10+6+10+7}{5} = \frac{40}{5} = 8$$

TAREA: Resolver los siguientes problemas para encontrar la media aritmética.

1. Un estudiante obtuvo las calificaciones 85, 76, 93, 82 y 96 en 5 materias. Encuentre la media aritmética de las calificaciones.
2. Un psicólogo midió los tiempos de reacción de un individuo a ciertos estímulos, siendo estos 53, 46, 50, 49, 52, 53, 44 y 55 segundos, en ese orden. Determine la media del tiempo de reacción del individuo a los estímulos.
3. Encuentre la media aritmética de los números 5, 3, 6, 5, 2, 8, 6, 5, 4, 8, 3, 4, 5, 4, 8, 2, 5, y 4.
4. El siguiente resultado con Minitab muestra el tiempo por semana que pasaron en línea 30 usuarios de internet. ¿Diría usted que este promedio es típico de los 30 promedios?

3	4	4	5	5	5	5	5	5	6
6	6	6	7	7	7	7	7	8	8
9	10	10	10	10	10	10	10	10	10

5. Un empresario, dueño de una gasolinera “A”, desea comprar las ventas diarias en litros de su gasolinera con las de un competidor “B”. Ambas gasolineras son muy parecidas en cantidad de operarios, capacidad y ubicación en la ciudad de Aguascalientes. Los datos correspondientes para 40 días del año 2,010, tomados al azar, se muestran en las siguientes tablas:

Ventas en cientos de litros por dia en la gasolinera “A”

28.7,	42.9,	39.4,	35.6,	37.0,	37.9,	37.6,	29.0
36.6,	38.5,	37.5,	35.4,	40.8,	33.7,	38.5,	29.2
33.0,	35.9,	25.3,	32.7,	31.4,	34.9,	29.3,	31.2,
37.6,	30.6,	31.1,	33.1	35.8,	37.2,	31.1,	37.7
39.5,	33.2,	32.4,	35.5,	33.4,	32.8,	33.0,	35.9

Ventas en cientos de litros por día en la gasolinera “B”

39.7,	30.6,	36.5,	41.5,	37.7,	28.5,	34.6,	32.8
39.4,	44.6,	37.1,	31.4,	43.9,	38.4,	40.4,	37.9
42.4,	40.0,	34.0,	35.8,	36.1,	34.8,	35.2,	30.5
28.0,	48.9,	47.6,	43.0,	44.2,	40.6,	41.5,	52.7
36.0,	33.0,	44.4,	46.3,	34.1,	52.1,	27.7,	47.9

¿Cuál gasolinera podría decirse que es la más productiva?
¿Calcular la media aritmética de las gasolineras?

La Media Aritmética para una Población

De la misma forma, para una población con “N” elementos y por lo tanto con “N” observaciones x_1, x_2, x_3,……………….x_N, la expresión algebraica que define la media es:

$$\mu = \frac{X_1 + X_2 + X_3 + \ldots\ldots\ldots\ldots + X_N}{N} = \sum_{i=1}^{N} \frac{X_i}{N}$$

μ es la media aritmética de la población
N es el tamaño de la población o número de elementos de la población
X es la i-esima observación de la variable x de la población
i=1 es el valor numérico 1, 2, 3,…………….n
ΣX_i es la suma de todos los valores x_i de la población

Ejemplos:

1. El censo de una población de los años anteriores del 2010 al 2017, fueron: 35,000; 42,000; 49,000; 54,000; 63,000; 67,000; 76,000; 81,000; 91,000; encontrar la media aritmética de la población.

Solución:

$$\mu = \frac{X_1 + X_2 + X_3 + \ldots\ldots\ldots + X_N}{N} = \sum_{i=1}^{N} \frac{X_i}{N}$$

$$\mu = \frac{35{,}000+42.000+49{,}000+\ldots\ldots+91.000}{9} = \frac{558{,}000}{9} = 62{,}000$$

Nota: Aplicando la fórmula de la media aritmética para una muestra y para una Población es el mismo procedimiento, lo único que cambia es la simbología, "X" y "n" para una muestra, "μ" y "N" para una Población.

La Media Aritmética para Datos Agrupados (con frecuencias)

Cuando los datos se agrupan en tablas, la media aritmética es igual a la división de la sumatoria del producto de las clases por la frecuencia sobre el número de datos. Cuando los datos se agrupan en tablas de frecuencias, el cálculo de la media varia un poco, ya que existe una pérdida de información en el momento en el que se trabaja con intervalos de frecuencia y no con los datos directamente (los datos se agrupan por intervalo, desconociendo el valor exacto de cada uno de ellos).
Se debe de multiplicar la variable por la frecuencia respectiva, luego se halla la suma de todos estos productos y a este valor se le divide entre el número de elementos.
.
La media aritmética para datos agrupados es la que se indica, si los valores x_1, x_2, x_3,…x_n, ocurren f_1, f_2, f_3,..……f_n, veces respectivamente o sea con frecuencias, se aplica la siguiente formula:

$$X_f = \frac{f_1 x_1 + f_2 x_2 + f_3 x_3 + \ldots\ldots + f_n x_n}{f_1 + f_2 + f_3 + \ldots\ldots + f_n} = \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{\Sigma f x}{N}$$

Ejemplos:

1. Cuatro grupos de estudiantes consisten en: 15, 20,10, y 18, individuos, dieron pesos medios de: 162, 148, 153 y 140 libras, respectivamente. Encontrar el peso medio de todos esos estudiantes.

Solución:

$$X_f = \frac{f_1 x_1 + f_2 x_2 + f_3 x_3 + \ldots\ldots + f_n X_n}{f_1 + f_2 + f_3 + \ldots\ldots + f_n} = \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{\Sigma f x}{N}$$

Frecuencias: 15, 20, 10, 18

Valores de x: 162, 148, 153, 140 Libras

$$X_f = \frac{15(162)+20(148)+10(153)+18(140)}{15+20+10+18} = \frac{2,430+2960+1,530+2,520}{63} = \frac{9,440}{63} =$$

$X_f = 149.84$ Libras

2. De entre 100 números, 20 son cuatros, 40 son cinco, 30 son seises y los restantes sietes. Encontrar su media aritmética.

Solución:

$$X_f = \frac{f_1 x_1 + f_2 x_2 + f_3 x_3 + \ldots\ldots + f_n X_n}{f_1 + f_2 + f_3 + \ldots\ldots + f_n} = \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{\Sigma f x}{N}$$

Frecuencias: 20, 40, 30, 10

Valores de x: 4, 5, 6, 7

$$X_f = \frac{20(4)+40(5)+30(6)+10(7)}{20+40+30+10} = \frac{80+200+180+70}{100} = \frac{530}{100} = 5.30$$

3. Hallar la media aritmética de los números: 5, 3, 6, 5, 4, 5, 2, 8, 6, 5, 4, 8, 3, 4, 5, 4, 8, 2, 5 y 4. Para datos agrupados.

Solución:

$$X_f = \frac{f_1 x_1 + f_2 x_2 + f_3 x_3 + \ldots\ldots + f_n X_n}{f_1 + f_2 + f_3 + \ldots.+ f_n} = \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{\Sigma f x}{N}$$

Frecuencias: 6, 2, 2, 5, 2 y 3

Valores de X: 5, 3, 6, 4, 2 y 8

$$X_f = \frac{6(5)+2(3)+2(6)+5(4)+2(2)+3(8)}{6+2+2+5+2+3} = \frac{30+6+12+20+4+24}{20} = \frac{96}{20} = 4.8$$

4. Las calificaciones finales de un estudiante en cuatro asignaturas fueron 82, 86, 90 y 70. Si los respectivos créditos otorgados a esos cursos son 3, 5, 3 y 1. Determinar una calificación media apropiada.

Solución:

$$X_f = \frac{f_1 x_1 + f_2 x_2 + f_3 x_3 + \ldots\ldots + f_n X_n}{f_1 + f_2 + f_3 + \ldots\ldots + f_n} = \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{\Sigma f x}{N}$$

Frecuencias: 3, 5, 3 y 1

Valores de x: 82, 86, 90 y 70

$$X_f = \frac{3(82)+5(86)+3(90)+1(70)}{3+5+3+1} = \frac{246+430+270+70}{12} = \frac{1,016}{12} = 84.66$$

5. Tres profesores de Economía dieron notas medias en sus cursos con 32, 25, y 17 estudiantes, de 79, 74 y 82 puntos, respectivamente. Hallar la puntuación media de los tres cursos.

Solución:

$$X_f = \frac{f_1 x_1 + f_2 x_2 + f_3 x_3 + \ldots\ldots + f_n X_n}{f_1 + f_2 + f_3 + \ldots\ldots + f_n} = \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{\Sigma f x}{N}$$

Frecuencias: 32, 25, 17

Valores de x: 79, 74, 82

$$X_f = \frac{32(79)+25(74)+17(82)}{32+25+17} = \frac{2,528+1,850+1,394}{74} = \frac{5,772}{74} = 78$$

6. Si el conjunto de números 5, 8, 6 y 2 de la media, ocurren con frecuencias 3, 2, 4 y 1, respectivamente. Encontrar la media aritmética

Solución:

$$X_f = \frac{f_1 x_1 + f_2 x_2 + f_3 x_3 + \ldots\ldots + f_n X_n}{f_1 + f_2 + f_3 + \ldots.. + f_n} = \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{\Sigma fx}{N}$$

Frecuencias: 3, 2, 4 y 1

Valores de x: 5, 8, 6 y 2

$$X_f = \frac{3(5)+2(8)+4(6)+1(2)}{3+2+4+1} = \frac{15+16+24+2}{10} = \frac{57}{10} = 5.7$$

7. Usar la distribución de frecuencias de alturas para datos agrupados, para hallar la altura media de esos 100 estudiantes, aplicando la siguiente tabla:

ALTURA	**MARCA DE CLASE "x"**	**FRECUENCIA "f"**	**" f x"**
60-62	61	5	305
63-65	64	18	1,152
66-68	67	42	2,814
69-71	70	27	1,890
72-74	73	8	584
		100	6,745

Solución:

$$X_f = \frac{\Sigma fx}{\Sigma f} = \frac{6,745}{100} = 67.45$$

8. Calcular el salario semanal medio de los 65 empleados de la empresa P&R a partir de la distribución de frecuencias de la tabla:

SALARIO "x"	**FRECUENCIA "f"**	**"fx"**
$255.00	8	$ 2,040.00
$265.00	10	$ 2,650.00
$275.00	16	$ 4,400.00
$285.00	14	$ 3,990.00
$295.00	10	$ 2,950.00
$305.00	5	$ 1,525.00
315.00	2	$ 630.00
	65	$ 18,185.00

Solución:

$$X_f = \frac{\Sigma fx}{\Sigma f} = \frac{18,185.00}{65} = 279.77$$

9. Una muestra de 600 clientes de cierto negocio refleja los datos siguientes: Encontrar la media para datos agrupados.

CLASE ($)	**MARCA DE CLASE "x"**	**FRECUENCIA "f"**	**"fx"**
0 - 49.99	25	78	1,950
50.00 – 99.99	75	123	9,225
100.00 –149.00	125	187	23,375
150.00 –199.99	175	82	14,350
200.00 –249.99	225	51	11,475
250.00 –299.99	275	47	12,925
300.00 –349.99	325	13	4,225
350.00 –399.99	375	9	3,375

400.00 –449.99	425	6	2,250
450.00 –499.99	475	4	1,900
		600	85,350

Solución:

$$X_f = \frac{\Sigma fx}{\Sigma f} = \frac{85,350}{600} = 142.25$$

10. La siguiente tabla muestra la distribución de los diámetros de los remaches salidos de una fábrica. Calcular el diámetro medio.

DIAMETRO (cms)	FRECUENCIA "f"	MARCA DE CLASE "x"	"fx"
0.7247-0.7249	2	2	4
0.7250-0.7252	6	2	12
0.7253-0.7255	8	2	16
0.7256-0.7258	15	2	30
0.7259-0.7261	42	2	84
0.7262-0.7264	68	2	136
0.7265-0.7267	49	2	98
0.7268-0.7270	25	2	50
0.7271-0.7273	18	2	36
0.7274-0.7276	12	2	24
0.7277-0.7279	4	2	8
0.7280-0.7282	1	2	2
	250		500

Solución:

$$X_f = \frac{\Sigma fx}{\Sigma f} = \frac{500}{250} = 2$$

TAREA: Resolver los siguientes problemas para encontrar la media aritmética para datos agrupados.

1. Calcule el salario semanal medio de los 70 empleados de la empresa P&R, a partir de la distribución de frecuencias para datos agrupados, de la siguiente tabla:

FRECUENCIAS "f"	Marca de clase ($) "x"	"fx"
9	$215.00	
10	$225.00	
16	$235.00	
14	$245.00	
10	$255.00	
7	$265.00	
4	$275.00	

2. Cinco grupos de estudiantes, consistentes de 20, 25, 15, 17 y 13 individuos, reportaron pesos medios de 158, 152, 154, 145 y 161 libras, respectivamente. Encuentre el peso medio de todos los estudiantes.

3. Una empresa tiene 80 empleados, 30 ganan $10.00 por hora, 20 ganan $8.00 por hora, 13 ganan $15.00 por hora, 12 ganan $12.00 por hora y 5 ganan $20.00 por hora. Determine la ganancia media por hora.
4. De un total de 80 números, 15 eran cuatros, 25 eran cincos, 20 eran seises, 12 eran sietes y 8 eran nueves. Obtenga la media de los números.
5. Utilice la distribución de frecuencias de estaturas de la siguiente tabla: Para encontrar la estatura media de los 150 estudiantes hombres de la universidad XYZ.

ESTATURA (pulg.)	MARCA DE CLASE “x”	FRECUENCIA “f”	“fx”
70 – 72	71	6	
73 – 75	74	17	
76 – 78	77	25	
79 – 81	80	21	
82 – 84	83	15	
85 – 87	86	13	
88 – 90	89	9	
91 – 93	92	3	

La Media Aritmética Ponderada

Se denomina media ponderada de un conjunto de datos al resultado de multiplicar cada uno de los datos por un valor particular para cada uno de ellos, llamado peso, obteniendo a continuación la suma de estos productos y dividiendo el resultado de esta suma de productos entre la suma de los pesos. Este peso depende de la importancia o significancia de cada uno de los datos. En otros términos, es un promedio en la que cada valor u observación se pondera con algún factor de acuerdo con su importancia.

Una variante de la media es la media aritmética ponderada, en la que cada dato tiene un peso o ponderación diferente, es decir, algunas actividades tienen más valor que otras.

Para un conjunto de datos de “n” elementos representados como: $x_1, x_2, x_3, \ldots x_n$, cuyos pesos correspondientes son $w_1, w_2, w_3, \ldots\ldots\ldots w_n$, la media ponderada se calcula con la siguiente formula:

$$Xp = \frac{W_1X_1 + W_2X_2 + W_3X_3 + \ldots\ldots\ldots + W_nX_n}{W_1 + W_2 + W_3 + \ldots\ldots\ldots + W_n} = \frac{\sum_{i=1}^{n} W_iX_i}{\sum W_i} = \frac{\sum WX}{\sum W}$$

donde:

Xp es la media ponderada

X_i es la i-esima observación de la variable x de la media ponderada

i=1 es el valor numérico 1, 2, 3,…………….n

W es el peso asignado a cada elemento (ponderación)

$\Sigma(WX)$ es la sumatoria del producto de la ponderación de cada elemento correspondiente

ΣW es la suma de todas las ponderaciones

Ejemplos

1. La empresa coca cola maneja tres tipos de salarios para la calidad del personal y de ahí se toma para la producción de sus refrescos (coca cola clásica y coca cola light).
   Nivel operativo: técnicos, supervisores, gerentes

Salario por hora x:	\$40	\$57	\$90
Producto 1 (w):	3	2	4
Producto 2 (w):	4	2	5

   Encontrar la media ponderada de los productos.

Solución:

$$Xp = \frac{W_1X_1 + W_2X_2 + W_3X_3 + \ldots\ldots + W_nX_n}{W_1 + W_2 + W_3 + \ldots\ldots + W_n} = \frac{\sum_{i=1}^{n} W_iX_i}{\sum W_i} = \frac{\sum WX}{\sum W}$$

$$Xp_1 = \frac{3(40)+2(57)+4(90)}{3+2+4} = \frac{120+114+360}{9} = \frac{594}{9} = 66 \quad \text{(coca cola clásica)}$$

$$Xp_2 = \frac{4(40)+2(57)+5(90)}{4+2+5} = \frac{160+114+450}{11} = \frac{724}{11} = 65.81 \quad \text{(coca cola light)}$$

2. **2.** Determinar el costo de trabajo de una empresa llamada Yaqui, cuyos datos son los siguientes: Mano de obra.
   Nivel operativo: Operadores, Empacadores, Supervisores y Gerentes
   Salario por hora (\$): 11.50, 12.50, 30.40, y 58.00
   Producto 1 (leche 2%): 6, 9, 7 y 10
   Producto 2 (leche): 6, 7, 10 y 10
   Producto 3 (crema): 8, 6, 9 y 7

Solución:

$$Xp = \frac{W_1X_1 + W_2X_2 + W_3X_3 + \ldots\ldots + W_nX_n}{W_1 + W_2 + W_3 + \ldots\ldots + W_n} = \frac{\sum_{i=1}^{n} W_iX_i}{\sum W_i} = \frac{\sum WX}{\sum W}$$

$$Xp_1 = \frac{6(11.50)+9(12.50)+7(30.40)+10(58.00)}{6+9+7+10} = \frac{69.00+112.50+212.80+580.00}{32}$$

$$Xp_1 = \frac{974.3}{32} = 30.44$$

$$Xp_2 = \frac{6(11.50)+7(12.50)+10(30.40)+10(58.00)}{6+7+10+10} = \frac{69.00+87.50+304.00+580.00}{33}$$

$$Xp_2 = \frac{1,040.50}{33} = 31.53$$

$$Xp_3 = \frac{8(11.50)+6(12.50)+9(30.40)+7(58.00)}{8+6+9+7} = \frac{92.00+72.00+273.60+406.00}{30}$$

$$Xp_3 = \frac{843.60}{30} = 28.22$$

3. Cierta empresa elaboradora de tortillas y tostadas, maneja tres niveles de operación para sus productos: Mano de obra
   Nivel operativo: No calificado, semi-calificado y calificado.
   Salario por hora (\$): 30, 40, y 55
   Producto 1 (tortillas): 3, 7, y 9
   Producto 2 (tostadas): 2, 3, y 6

Solución:

$$Xp = \frac{W_1X_1 + W_2X_2 + W_3X_3 + \ldots\ldots + }{W_1 + W_{2+} W_3 + \ldots\ldots + } \frac{\Sigma W_nX_n}{\Sigma W_n} = \frac{\sum^{n} W_iX_i}{\sum_{i=1} W_i} = \frac{\Sigma WX}{\Sigma W}$$

$$Xp_1 = \frac{3(30)+7(40)+9(55)}{3+7+9} = \frac{90+280+495}{19} = \frac{865}{19} = 45.5$$

$$Xp_2 = \frac{2(30)+3(40)+6(55)}{2+3+6} = \frac{60+120+330}{11} = \frac{510}{11} = 46.36$$

4. Una compañía cuyos datos presentamos a continuación, utiliza tres niveles operación, para cinco productos finales. Determinar el promedio costo por l producto.
   Mano de Obra:
   Nivel operativo: Chef, cocinero, mesero
   Salario por hora ($): 106, 84, 92, 75, y 78
   Producto A: 9, 6, 10, 7, 3
   Producto B: 5, 8, 11, 3, 1
   Producto C: 4, 8, 12, 7, 5
   Producto D: 7, 1, 13, 9, 2
   Producto E: 2, 5, 14, 6, 3

Solución:

$$Xp = \frac{W_1X_1 + W_2X_2 + W_3X_3 + \ldots\ldots + }{W_1 + W_{2+} W_3 + \ldots\ldots + } \frac{\Sigma W_nX_n}{\Sigma W_n} = \frac{\sum^{n} W_iX_i}{\sum_{i=1} W_i} = \frac{\Sigma WX}{\Sigma W}$$

$$Xp_A = \frac{9(106)+6(84)+10(92)+7(75)+3(78)}{9+6+10+7+3} = \frac{954+504+920+525+234}{26}$$

$$Xp_A = \frac{3,137}{26} = 120.65$$

$$Xp_B = \frac{5(106)+8(84)+11(92)+3(75)+1(78)}{5+8+11+3+1} = \frac{530+672+1012+225+78}{28}$$

$$Xp_B = \frac{2,517}{28} = 89.89$$

$$Xp_C = \frac{4(106)+8(84)+12(92)+7(75)+5(78)}{4+8+12+7+5} = \frac{424+672+1,104+525+390}{36}$$

$$Xp_C = \frac{3,115}{36} = 86.53$$

$$Xp_D = \frac{7(106)+1(84)+13(92)+9(75)+2(78)}{7+1+13+9+2} = \frac{742+84+1,196+855+156}{32}$$

$$Xp_D = \frac{3,033}{32} = 94.78$$

$$Xp_E = \frac{2(106)+5(84)+14(92)+6(75)+3(78)}{2+5+14+6+3} = \frac{212+420+1,288+450+234}{30}$$

$$Xp_E = \frac{2,604}{30} = 86.8$$

5. Calcular la media ponderada de la nota final de un curso escolar, en donde se asigna distinta importancia (W = peso) a los distintos exámenes que se realicen. Los dos primeros exámenes tienen un peso o valor del 30% y 20%

respectivamente y el ultimo del 50%; las calificaciones respectivas son de 64, 92 y 81, entonces la nota final corresponde a la siguiente media ponderada.

Valores de "x":	64,	62,	81
Pesos:	0.30,	0.20	0.50

Solución:

$$X_p = \frac{W_1X_1 + W_2X_2 + W_3X_3 + \ldots\ldots + \Sigma W_nX_n}{W_1 + W_{2+} W_3 + \ldots\ldots + \Sigma W_n} = \frac{\sum_{i=1}^{n} W_iX_i}{\Sigma W_i} = \frac{\Sigma WX}{\Sigma W}$$

$$X_p = \frac{030(64)+0.20(92)+0.50(81)}{0.30+0.20+0.50} = \frac{19.2+18.4+40.5}{1.00} = 78.1$$

6. Se tiene una información acerca de las utilidades por pan y cantidades vendidas de panes de tres tiendas. Calcular la media ponderada de la utilidad por pan.

PAN	UTILIDAD/PAN "X"	CANTIDAD VENDIDA "W"
1	1	2,000
2	0.8	1,800
3	0.9	2,100

Solución:

$$X_p = \frac{W_1X_1 + W_2X_2 + W_3X_3 + \ldots\ldots + \Sigma W_nX_n}{W_1 + W_{2+} W_3 + \ldots\ldots + \Sigma W_n} = \frac{\sum_{i=1}^{n} W_iX_i}{\Sigma W_i} = \frac{\Sigma WX}{\Sigma W}$$

$$X_p = \frac{2,000(1)+1,800(0.8)+2,100(0.9)}{2,000+1,800+2,100} = \frac{2,000+1,440+1,890}{5,900} = \frac{5,330}{5,900} = 0.90$$

7. En una materia dada se asignan pesos de importancia, de la siguiente forma: Unidad I (20% curso), Unidad II (25% del curso), Unidad III (20% del curso), Unidad IV (15% del curso), Unidad V (de la calificación).
Si las calificaciones de un alumno son: 8 en la primera unidad, 5 en la segunda unidad, 8 en la tercera unidad, 10 en la cuarta unidad y 8 en la última unidad, es decir, se tiene la siguiente tabla:

Solución:

UNIDAD	PONDERACION (W)	DATOS (X)
I	20% = 0.20	8
II	25% = 0.25	5
III	20% = 0.20	8
IV	15% = 0.15	10
V	20% = 0.20	8

Solución:

$$X_p = \frac{W_1X_1 + W_2X_2 + W_3X_3 + \ldots\ldots + \Sigma W_nX_n}{W_1 + W_{2+} W_3 + \ldots\ldots + \Sigma W_n} = \frac{\sum_{i=1}^{n} W_iX_i}{\Sigma W_i} = \frac{\Sigma WX}{\Sigma W}$$

$$X_p = \frac{0.2(8)+0.25(5)+0.20(8)+0.15(10)+0.20(8)}{0.20+0.25+0.20+0.15+0.20} = \frac{1.6+1.25+1.6+1.5+1.6}{1.00}$$

$$Xp = \frac{7.55}{1.00} = 7.55$$

8. La nota final de una asignatura es una media ponderada de las notas que han obtenido los alumnos en los cuatro elementos evaluables que determina el profesor. El responsable de la asignatura otorga un peso de 3 al examen inicial, 1 al trabajo entregado, 2 al trabajo final y 4 al examen final. Las notas de un alumno han sido las siguientes:

NOTAS DEL ALUMNO

ELEMENTOS EVALUADOS	NOTA (x)	PESO (w)
Examen inicial	5.2	3
Trabajo entregado	8.2	1
Trabajo final	7.4	2
Examen final	5.7	4

Solución:

$$Xp = \frac{W_1X_1 + W_2X_2 + W_3X_3 + \ldots\ldots + }{W_1 + W_2 + W_3 + \ldots\ldots + } \quad \frac{\Sigma W_nX_n}{\Sigma W_n} = \frac{\sum_{i=1}^{n} W_iX_i}{\Sigma W_i} = \frac{\Sigma WX}{\Sigma W}$$

$$Xp = \frac{3\,(5.2) + 1\,(8.2) + 2\,(7.4) + 4\,(5.7)}{3+1+2+4} = \frac{15.6+8.2+14.8+22.8}{10} = \frac{61.4}{10} = 6.14$$

TAREA: Resolver los siguientes problemas para encontrar la media ponderada.

1. Calcula la media aritmética ponderada hallando la nota promedio de cinco asignaturas que tienen valores diferentes. Química tiene un valor de 3 créditos, Física vale 4 créditos, matemáticas 5 créditos, biología 2 créditos, Estadística 3 créditos. Si Pedro ha sacado un 8 en la primera, un 7 en la segunda, un 9 en la tercera, un 10 en la cuarta y un 8 en la quinta.
2. Para calcular la nota final del curso de literatura en donde cada apartado ha tenido distinta importancia. Los dos primeros trabajos tienen valor de 20% y 20% respectivamente, y el examen de 60%; las calificaciones respectivas son de 6.4, 9.2 y 8.1
3. Un grupo de 6 amigas tienen distintas edades. Son las siguientes: 2 de ellas tienen 28 años y otras 2 tienen 32 años, el resto tienen 29 y 30 años respectivamente. Calcula la media aritmética del grupo.
4. Encontrar la media ponderada de los productos de la siguiente tabla:

		Mano de Obra		
NIVEL OPERATIVO:	SALARIO POR HORA $	P_1	P_2	P_3
1	12.50	8	9	5
2	13.50	4	2	6
3	35.20	9	6	8
4	60.50	12	10	11

5. Se tiene una información acerca de las utilidades por la venta de refrescos y cantidades vendidas de refrescos de tres tiendas. Calcular la media ponderada de la utilidad por la venta de los refrescos.

COCA COLA	UTILIDAD/ (x)	CANTIDAD VENDIDA
1	3	2,000

2	2	1,800
3	5	2,100

La Media Geométrica

En matemáticas y estadística, la media geométrica de una cantidad arbitraria de números (“n” números) es la raíz n-esima del producto de todos los números, es recomendada para datos de progresión geométrica, para promediar para promediar razones, interés compuesto y números índices.

La media geométrica se utiliza en el cálculo de tasas de crecimiento, ya sea en los negocios o en el crecimiento poblacional.

La media geométrica de un conjunto de datos de “N” números enteros positivos X_1, X_2, X_3,…..X_n, es la raíz N-esima del producto de los n valores. La expresión algebraica es:

$$MG= \sqrt[N]{X1X2X3X4 \ldots .Xn}$$

Si un valor $x_i = 0$, entonces la media geométrica se anula o no queda definida. Solo es relevante la media geométrica si todos los números son positivos.

Ejemplos:

1. Encontrar la media geométrica de los siguientes conjuntos de datos:

a). 3.2; 5.2; 4.8; 6.1 y 4.2

$$MG= \sqrt[5]{(3.2)(5.2)(4.8)(6.1)(4.2)} = \sqrt[5]{2{,}046.32} = 4.59$$

b). 6, 3, 5, 12, 7, 10 y 6.

$$MG= \sqrt[7]{(6)(3)(5)(12)(7)(10)(6)} = \sqrt[7]{453{,}600} = 6.42$$

c), 27, 38, 32 y 43

$$MG= \sqrt[4]{(27)(38)(32)(43)} = \sqrt[4]{1'411{,}776} = 34.47$$

d). 8, 2, 6, 3, 10 y 4

$$MG= \sqrt[6]{(8)(2)(6)(3)(10)(4)} = \sqrt[6]{11{,}520} = 3.22$$

e). 12, 17, 13, 9, 1, 4, 10, 6 y 15

$$MG= \sqrt[9]{(12)(17)(13)(9)(1)(4)(10)(15)} = \sqrt[9]{14'320{,}800} = 3{,}784.28$$

f). 8, 3, 10, 4, 12, 19, 11 y 15

$$MG= \sqrt[8]{(8)(3)(10)(4)(12)(19)(11)(15)} = \sqrt[8]{36'115,200} = 6,009.59$$

g). 3, 2, 7 y 11

$$MG= \sqrt[3]{(3)(2)(7)(11)} = \sqrt[3]{462} = 21.49$$

h). 14 y 31

$$MG= \sqrt{(14)(31)} = \sqrt{434} = 20.83$$

TAREA: Resolver los siguientes problemas para encontrar la media geométrica.

1. 5, 4, 8, 3, 7, 2 y 9
2. 71, 89, 78, 80 y 70
3. 32, 25, 17, 21, 30 y 19
4. 12, 5 , 17, 6, 12, 3 y 14
5. 98, 75, 56, 42, 30, 21, 15 y 11

Propiedades de la Media Geométrica

- Está basada en todos los elementos de un conjunto de datos. El valor de cada elemento afecta el valor de la media geométrica. 3
- Si uno de los datos es cero, su valor es cero. MG= 0x18x54 = 0
- Si uno de los datos es negativo y el número de datos es par, el valor de la media geométrica es un número complejo y carece de interpretación.
- MG= - 6x18= + 10.39i
- Si uno de los datos es negativo y el número de datos es impar, la media geométrica es negativa y puede ser no representativa.

  3
  MG= -6x18x54 = - 18
- La media geométrica es afectada por valores extremos en menor cantidad que la media aritmética.
- La media geométrica resulta conveniente cuando se manejan proporciones.

1.5.2 LA MEDIANA

Para una distribución de datos ordenados en magnitud, la mediana es el valor central o la media de los dos valores centrales, es decir, el dato que divide a la distribución en dos partes iguales; el 50% de los datos se encuentran a la izquierda de ese valor y el otro 50% a la derecha. La mediana se representa por (md).
El procedimiento para calcular la mediana es:
Los datos deben ordenarse ya sea de menor a mayor o mayor a menor.
Si el número de datos es impar, la mediana es el valor que está en el centro de la distribución. Si el número de datos es par, la mediana debe de tomarse como la media de estos dos valores centrales (tomamos los dos de en medio y sacamos su promedio).

La mediana no se calcula mediante operaciones aritméticas en las que intervengan todos los valores, como se hace con la media aritmética.
Se puede también encontrar el valor de la mediana por medio de su posición utilizando la siguiente formula:

Posición de la mediana $md_p = \frac{n-1}{2}$; donde "n" es el número de observaciones

Ejemplos:

1. El tiempo de los integrantes de un equipo de atletismo, en donde su participación realizó los siguientes tiempos: 4.8, 4.3, 9.1, 5.0, 4.9, 4.2 y 5.1, encontrar la mediana y la posición de la mediana.

Solución:

Primero lo ponemos en orden de menor a mayor, es un número impar, por lo tanto, sale el número que se encuentra en el centro de la distribución de datos.

4.2, 4.3, 4.8, 4.9, 5.0, 5.1, 9.1 La mediana es 4.9

La posición de la mediana es aplicando la fórmula:

$md_p = \frac{n+1}{2}$

$md_p = \frac{7+1}{2} = \frac{8}{2} = 4$ La mediana se encuentra en la posición 4

2. Los accidentes tratados en el municipio de Miguel Alemán, durante los días consecutivos fueron: 11, 86, 30, 52, 31, 49, 35 y 43; Encontrar la mediana y su posición de la mediana.

Solución:

Primero lo ponemos en orden de menor a mayor, es un numero par, salen los dos números centrales se suman y se divide entre dos, así obtenemos la mediana.

11, 30, 31, 35, 43, 49, 52, 86

$md = \frac{35+43}{2} = \frac{78}{2} = 39$ La mediana es el 39

La posición de la mediana es aplicando la fórmula:

$md_p = \frac{n+1}{2}$

$md_p = \frac{8+1}{2} = \frac{9}{2} = 4.5$ La mediana se encuentra en la posición 4.5

3. Se registró el número de transacciones de ATM por día en 15 lugares de una gran ciudad. Los datos fueron: 35, 49, 225, 50, 30, 65, 40, 55, 52, 76, 48, 325, 47, 32 y 60. Encuentre la mediana y la posición de la mediana.

Solución:

Primero lo ponemos en orden de menor a mayor, es un número impar, por lo tanto sale el número que se encuentra en el centro de la distribución de datos.

30, 32, 35, 40, 47, 48, 49, 50, 52, 55, 60, 65, 76, 225 y 325

La mediana es 50.

La posición de la mediana es aplicando la fórmula:

$md_p = \frac{n+1}{2}$

$md_p = \frac{15+1}{2} = \frac{16}{2} = 8$ La mediana se encuentra en la posición 8

4. Encuentre la mediana de los pesos de los 40 estudiantes de la universidad estatal; utilizando la siguiente distribución:
   119, 163, 145, 125, 164, 145,126, 165, 146, 128, 168, 146, 132, 150, 173, 147, 135,176, 147, 135, 148, 136, 150, 135, 149, 138, 161, 138, 152, 140,158, 153, 140, 154, 144, 156, 142, 168 y 149,

Solución:

Primero lo ponemos en orden de menor a mayor, es un numero par, salen los dos números centrales se suman y se divide entre dos, así obtenemos la mediana.

119, 125, 126, 128, 132, 135, 135, 135, 136, 138, 138, 140, 140, 142, 144, 145, 145, 146, 146, 147, 147, 148, 149, 149, 150, 150, 152, 153, 154, 156,157, 158, 161, 163, 164, 165, 168, 168, 173, y 176

$Md = \frac{147+147}{2} = \frac{294}{2} = 147$ La mediana

La posición de la mediana es aplicando la fórmula:

$md_p = \frac{n+1}{2}$

$md_p = \frac{40+1}{2} = \frac{41}{2} = 20$ La mediana se encuentra en la posición 8

5. Encuentre la mediana del siguiente conjunto de números:
   5, 14, 8, 3, 17, 2, 9, 13, 20, 15 y 19. Encuentre la mediana y la posición de la mediana.

Solución:

Primero lo ponemos en orden de menor a mayor, es un número impar, por lo tanto sale el número que se encuentra en el centro de la distribución de datos.

2, 3, 5, 8, 9, 13, 14, 15, 17, 19, 20. La mediana es el 13

La posición de la mediana es aplicando la fórmula:

$md_p = \frac{n+1}{2} = \frac{11}{2} = 5.5$ La mediana se encuentra en la posición 5.5

6. Encuentre la mediana de la distribución de los números y la posición de la mediana: 18, 20,19, 22, 20, 18, 19 y 20

Solución:

Primero lo ponemos en orden de menor a mayor, es un numero par, salen los dos números centrales se suman y se divide entre dos, así obtenemos la mediana.

18, 18, 19, 19, 20, 20, 20, 22

$Md = \frac{19+20}{2} = \frac{39}{2} = 19.5$ La mediana 19.5

La posición de la mediana es aplicando la fórmula:

$md_p = \frac{n+1}{2} = \frac{8+1}{2} = 4.5$ La mediana se encuentra en la posición 4.5

7. ¿Cuál es la mediana y la posición de la mediana de las calificaciones de un estudiante de ingeniería del sexto semestre? 90, 80, 75, 90, 85, 100, 95.

Solución:

Primero lo ponemos en orden de menor a mayor, es un número impar, por lo tanto sale el número que se encuentra en el centro de la distribución de datos:

75, 80, 85, 90, 90, 95, 100 La mediana es el 90

La posición de la mediana es aplicando la fórmula:

$md_p = \frac{n+1}{2} = \frac{7+1}{2} = \frac{8}{2} = 4$ La mediana se encuentra en la posición 4

8. ¿Encuentre la mediana y la posición de la mediana del conjunto de números? 5, 4, 8, 3, 7, 2, 9, 6, 12, 15, 11, 18, 10, 17, 20 y 1

Solución:

Primero lo ponemos en orden de menor a mayor, es un numero par, salen los dos números centrales se suman y se divide entre dos, así obtenemos la mediana.

1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 15, 17, 18, 20

$Md = \frac{8+9}{2} = \frac{17}{2} = 8.5$ La mediana es 8.5

La posición de la mediana es aplicando la fórmula:

$md_p = \frac{n+1}{2} = \frac{16+1}{2} = \frac{17}{2} = 8.5$ La mediana se encuentra en la posición 4.5

TAREA: Resolver los siguientes problemas para encontrar la mediana y la posición de la mediana.

1. Un estudiante obtuvo las calificaciones 75, 86, 73, 96, 100, 80 y 90. Encuentre la mediana y la posición de la mediana.
2. Las calificaciones obtenidas por un estudiante en laboratorio, teoría y práctica de un curso de física son: 95, 80, 100, 85 y 70. Encontrar la mediana y la posición de la mediana.
3. Calcule la mediana y posición de la mediana del conjunto de números: 7, 4, 10, 9, 15, 12, 17, 7, 5, 8, 3 y 19.
4. Calcule la mediana y la posición de la mediana del tiempo de reacción de los siguientes datos: 8, 11, 3, 2, 5, 10, 4, 1, 10,12, 6 y 7.
5. Determine la mediana y posición de la mediana para el diámetro de los remaches 6, 2, 8, 4, 10, 5, 11, 1, 14 y 17

1.5.3 LA MODA

Es una medida de tendencia central diferente de la media, pero un tanto parecida a la mediana, en realidad no se calcula mediante un proceso aritmético ordinario.
La moda de un conjunto de números es el valor que ocurre con mayor frecuencia.
La moda puede no existir e incluso no ser única en caso de existir
La moda tiene algunas desventajas: Una desventaja de esta medida estadística es que a veces no se puede definir porque ningún valor se repite. Otra desventaja es que en ocasiones se obtienen varias modas.
Cuando un conjunto de datos tiene una moda se dice que es “unimodal”, si tiene dos modas bimodal y más de tres modas multimodal.

Ejemplos:

1. Encontrar la moda y la categoría de la moda del conjunto de datos de una población: 115, 110, 130, 160, 50, 80, 120, 40, 20, 60, 90, 170, 210, 190, 10, 30, 310, 280, 250 y 70.

Solución:

Mo. = No hay moda no se repite ningún número.

2. Encontrar la moda y la categoría de la moda del conjunto de datos:
   7, 5, 4, 9, 7, 10, 4, 2, 8, 5, 1, 4, 3, 7 y 5

Solución:

La moda es; 4, 5 y 7;   La categoría es bimodal

3. Encontrar la moda y la categoría de la moda de los elementos de una fábrica de rines: 2, 5, 9, 10, 18, 2, 19, 7, 9, 10, 11, 9 y 12

Solución:

La moda es: 9 y la categoría es Unimodal.

4. Encontrar la moda y la categoría de la moda de los siguientes pares de tenis: 4, 15. 9, 7, 5, 2, 6, 4, 8, 10, 15, 13, 5, 7, 10, 2, 13, 8, 16, 13, 4, 17, 15, 10 y 14

Solución:

La moda es: 4, 10, 13, 15 y la categoría es multimodal

5. Determine la moda y su categoría del conjunto de números: 1, 3, 5, 7, 9, 3, 10, 2, 15, 17, 12, 3, 9, 10, 4, 9, 6 y 1

Solución:

La moda es: 3 y 9   y la categoría es bimodal

6. Busque la moda de las cargas máximas para los cables 16, 17, 11, 9, 15, 20, 13, 17, 19, 7, 9, 6, 8, 3, 1, 17, 13 y 5

Solución:

La moda es: 17 y la categoría es unimodal

7. Encuentre la moda y la categoría de la distribución de los números: 1, 3, 2, 5, 7, 9, 6, 13, 15, 8, 17, 19, 10, 21, 14, 25, 12, 27, 16, 23, 18, 22, 20, 29, 24

Solución:

La moda es: no hay moda

8. Determine cuál es la moda y la categoría del conjunto de datos: 21, 19, 1, 23, 17, 25, 15, 27, 13, 29, 11, 31, 13, 9, 33, 7, 35, 5, 37, 3, 17, 39, 1, 40, 13, 20, 10 y 30.

Solución:

La moda es: 1, 9, 13 y 17 y la categoría multimodal.

TAREA: Resolver los siguientes problemas para encontrar la moda y la categoría de los siguientes conjuntos de números:

1. 31, 14, 86, 93, 89, 90, 58, 95, 13, 41, 68, 39, 98, 9, 85 y 59
2. 10, 7, 3, 11, 5, 12, 1, 13, 10, 17, 21, 13, 18, 4, 9, 23, 30 y 25
3. 23, 10, 8, 2, 19, 6, 14, 3, 23, 12, 1, 5, 11, 17, 15, 23, 6, 20 y 30
4. 2, 6, 10, 14, 18, 22, 1, 5, 9. 13, 17, 21, 2, 14, 9, 1, 17, 20 y 25
5. 28,16, 24, 5, 2, 14, 31, 28, 4, 1, 16, 8, 12, 10, 11, 58, 60, 52 y 54

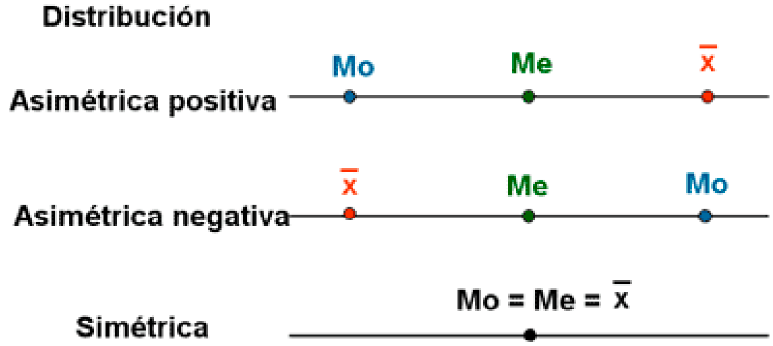
1.5.4 RELACIÓN ENTRE MEDIA, MEDIANA Y MODA

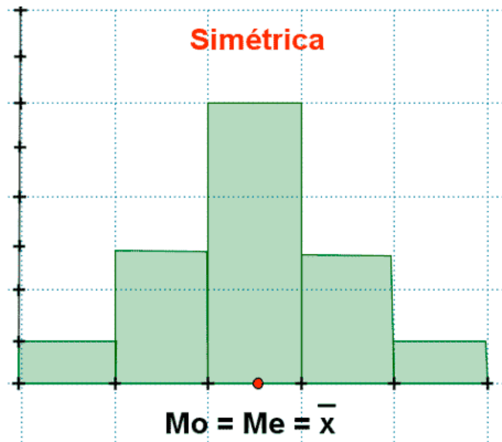
En el caso de distribuciones uni-modales, la mediana esta con frecuencia comprendida entre la media y la moda (incluso más cerca de la media).
En distribuciones que presentan cierta inclinación, es más aconsejable el uso de la mediana. Sin embargo, en estudios relacionados con propósitos estadísticos y de inferencia suele ser más apta la media.
Cuando una distribución de frecuencia es simétrica, la media, mediana y moda coinciden en su valor (X = Me = Mo). En el caso de una distribución binomial simétrica, es necesario calcular el promedio de las modas.
En una distribución sesgada a la izquierda, la moda es menor a la mediana, y esta a su vez menor que la media (X < Mo < Me)
En una distribución sesgada a la derecha la relación se invierte, la moda es mayor a la mediana, y esta a su vez mayor que la media (Mo > Me >).

Posiciones relativas de la media, mediana y moda
En el perfil idealizado de un histograma, se tienen líneas continuas suaves. En distribuciones simétricas, con sesgo a la izquierda y a la derecha.

En una distribución con una sola moda:
Si la distribución es asimétrica positiva, la media es superior a la mediana y ésta superior a la moda.

Si la distribución es asimétrica negativa. La media es inferior a la mediana y ésta inferior a la moda. Si la distribución es simétrica, la media, la mediana y la moda coinciden.





Ejemplos:

1. **Un médico atendió en 200 días las siguientes urgencias: 1, 3, 1, 1, 0, 1, 0, 2, 2, 0, 0, 1, 1, 2, 0, 6, 3, 1, 4, 0**
   a). Resumir los datos en una tabla que muestre frecuencias absolutas y porcentajes, y dibujar el correspondiente diagrama de barras.
   b). Calcular la media, la mediana y la moda del conjunto de datos. ¿Es simétrica la distribución anterior?

a).

x_i	f_i	h_i	p_i
0	6	0,30	30
1	7	0,35	35
2	3	0,15	15
3	2	0,10	10
4	1	0,05	5
6	1	0,05	5
Sumatorio:	20	1	100

b) Mo = 1
x = 6 x 0

$$X = \frac{6 \times 0+7 \times 1+3 \times 2+2 \times 3+1 \times 4+1 \times 6}{20} = \frac{29}{20} = 1.45$$

Como tenemos 20 valores (número par) de una variable discreta, la mediana será la semisuma de los valores que ocuparan los lugares 10 y 11

$$Me = \frac{1+1}{} = \frac{2}{} = 1$$

2 2

La distribución NO es simétrica puesto que Me = Mo =/ X / Algo que ya habíamos observado en el diagrama de barras.

2. Una oficina bancaria ha tabulado las cantidades de dinero que retirarán 100 clientes en un determinado día en un cajero automático:

Euros (€)	nº de clientes
[0, 20)	33
[20, 40)	27
[40, 60)	19
[60, 80)	14
[80, 100)	7

Hallar:
a) Cantidad media de dinero retirado por cada cliente y desviación típica.

b) La mediana. Interpretar el resultado.

a).

Intervalos	x_i	f_i	$x_i \cdot f_i$	x_i^2	$f_i \cdot x_i^2$	F_i
[0, 20)	10	33	330	100	3300	33
[20, 40)	30	27	810	900	24300	60
[40, 60)	50	19	950	2500	47500	79
[60, 80)	70	14	980	4900	68600	93
[80, 100)	90	7	630	8100	56700	100
Sumatorio:		100	3700		200400	

$$\bar{x} = \frac{3700}{100} = 37$$ 37€ es la cantidad de dinero media retirada por cada cliente

$$\sigma = \sqrt{\frac{200400}{100} - 37^2} = \sqrt{635} = 25{,}20$$

b).
Frecuencia acumulada de [0, 20]

Frecuencia acumulada de [20, 40]

N = 100 = 50

2 2

$$\frac{N}{2} = \frac{100}{2} = 50$$

$$\begin{cases} \text{Frecuencia acumulada de } [0,\ 20) \Rightarrow 33 \\ \text{Frecuencia acumulada de } [20,\ 40) \Rightarrow 60 \end{cases}$$

La mediana estará en el primer intervalo cuyo $N_i \geq 50$

es decir, el intervalo mediano [20, 40)

$$Me = 20 + 20 \cdot \frac{\frac{100}{2} - 33}{27} = 32{,}59$$

Al ser la distribución unimodal y la media mayor que la mediana, la distribución será asimétrica positiva.

1.6 MEDIDAS DE DISPERSIÓN PARA UN CONJUNTO DE DATOS NO AGRUPADOS Y DATOS AGRUPADOS

La dispersión de los datos intenta dar una idea de cuan esparcidos se encuentran estos.

Hay varias medidas de tal dispersión, siendo las más comunes el rango, la desviación media, la varianza, la desviación estándar, la desviación típica por el método corto.

Son indicadores estadísticos que muestran la distancia promedio que existe entre los datos y la media aritmética.

1.6.1 RANGO

Un inconveniente del rango es que emplea dos valores el mayor y el menor. No toma en consideración todos los demás valores.

El Rango de un conjunto de números es la diferencia entre el número mayor y el número menor de los datos.

El Rango de un conjunto de datos es una medida de la dispersión de los mismos. Sin embargo, el rango depende únicamente de los puntos extremos del conjunto de datos y no proporciona información respecto a la dispersión de estos entre esos dos puntos.

De forma simbólica:

R = valor del dato mayor menos valor del dato menor

Ejemplo:

I. Encuentre el Rango de los siguientes conjuntos de números:

1. 5, 2, 3, 8, 5, 12, 8, 3, 10, 5, 1, 4, 6 y 8

Solución:

Rango = 12 – 1 = 11

2. 3, 5, 7, 9, 11, 2, 4, 6, 8, 10, 14, 18, 13, 5, 9, 2, 7, 3, 4, 6, 3, 13, 11 y 7

Solución:

Rango = 18 – 2 = 16

3. 14, 12, 18, 11, 16, 20, 9, 13, 10, 17, 22, 15, 21, 19, 13, 22, 18, 14, 10 y 17

Solución:

Rango = 22 – 9 = 13

4. 61, 58, 65, 60, 63, 67, 62, 70, 58, 63, 45, 49, 66, 50, 61, 72, 64, 55, 52 y 57

Solución:

Rango = 72 – 45 = 27

5. 45, 78, 54, 61, 32, 68, 63, 84, 39, 64, 72, 79, 68, 80, 34, 38, 44 y 58

Solución:

Rango = 84 – 32 = 52

6. Una panificadora importante desea tomar la decisión sobre tres marcas de leche. Seleccionar aleatoriamente muestras de las tres marcas. El volumen de los envases corresponde a 1,000 ml, y por lo tanto, el valor medio tomando varios envases debería ser cercano a 1,000 ml. Los volúmenes de la muestra fueron:

Proveedor

A	1,009;	950;	990;	1,005;	1,023;	970;	1,010
B	987;	1,002;	1,004;	975;	1,009;	1,015;	1,000
C	1,006;	1,012;	970;	982;	997;	980;	990

Encontrar el Rango de las tres marcas de leche.

Solución:

R_A = 1,023 – 950 = 73

R_B = 1,015 – 975 = 40

R_C = 1,012 – 970 = 42

TAREA: Resolver los siguientes problemas para encontrar el Rango.

1. 12, 17, 19, 13, 24, 26, 21, 23, 29, 27, 31, 33, 39, 37, 35, 32, 20
2. 41, 46, 42, 48, 47, 43, 57, 51, 53, 59, 52, 58, 56, 54, 63, 68, 65
3. 25, 36, 14, 74, 85, 96, 52, 41, 63, 62, 42, 75, 95, 35, 15, 37, 46
4. 64, 31, 97, 46, 13, 79, 82, 71, 93, 37, 19, 28, 52, 58, 41, 47,, 63
5. 81, 89, 85, 82, 59, 62, 49, 73, 50, 90, 64, 78, 45, 56, 58, 54, 80

1.6.2 DESVIACIÓN MEDIA

Un análisis estadístico en el que se utiliza toda la información disponible permite obtener mejores conclusiones. Para describir la descripción de los datos, se prefiere utilizar una medida de tendencia central como referencia porque se calcula con todos los valores de las mediciones y su posición es central. Esto es lo que sucede con la desviación media., una medida de la dispersión de las mediciones alrededor de la media aritmética.

Para cualquier conjunto de datos es cierto que la suma de las desviaciones con respecto a la media no es útil para describir la dispersión de los datos, sin embargo, consideremos el valor absoluto de la desviación media. Ya que el valor absoluto puede ser positivo o cero. Para calcular el promedio de las desviaciones se utiliza el valor absoluto, evitando así que el resultado sea cero. A este promedio se le conoce como desviación media y se define como la media aritmética de los valores absolutos de las desviaciones de las variables respecto a la media.

La desviación respecto a la media es la diferencia en el valor absoluto entre cada valor de la variable estadística y la media aritmética.

$$D_i = /x_i – x/$$

La desviación media es la media aritmética de los valores absolutos de las desviaciones respecto a la media. La desviación media se representa por DM.

La desviación media o desviación promedio de un conjunto de N números x_1, x_2, x_3,….x_n, es abreviada DM y se define como:

$$DM = \frac{|X_1 – X| + |X_2 – X| + |X_3 – X| + \dots + |X_n – X|}{n}$$

$$DM = \frac{1}{n} \sum_{i=1}^{n} |X_i – X| \qquad DM = \sum_{i=1}^{n} \frac{|X_i – X|}{n}$$

donde:

DM es la desviación media

X_i es la i–esima observación de la variable “x” de la muestra

i=1 es el valor numérico 1, 2, 3,…………….n

X es la media aritmética de los números.

n es el tamaño de la muestra o número de elementos de la muestra

$|X_i$-X | es el valor absoluto de las diferencias entre el i-esimo valor numérico y la media aritmética.

Ejemplos:

1. Calcular la desviación media del conjunto de datos:8, 10, 6, 9, 7, 4, 11 y 13

Solución:

$$X = \frac{X_1+X_2+X_3+\dots+X_n}{n}$$

$$DM = \frac{|X_1– X| + |X_2 – X| + |X_3 – X| + \dots + |X_n – X|}{n}$$

$$X = \frac{8+10+6+9+7+4+11+13}{8} = \frac{68}{8} = 8.5$$

$$DM= \frac{|8-8.5|+|10-8.5|+|6-8.5|+|9-8.5|+|7-8.5|+|4-8.5|+|11-8.5|+|13-8.5|}{8} =$$

$$DM = \frac{|-0.5|+|1.5|+|-2.5|+|0.5|+|-1.5|+|-4.5|+|2.5|+|4.5|}{8} =$$

$$DM = \frac{0.5+1.5+2.5+0.5+1.5 +4.5+2.5+4.5}{8} = \frac{18}{8} = 2.25$$

2. Desde muy temprana edad, los zapatos para los niños tienden a ser más anchos que los de las niñas. ¿Esto se debe a que los niños tienen los pies más anchos o porque se supone que las niñas, incluso en la primaria, prefieren sacrificar el confort por la moda? Para evaluar lo primero, un estadístico mide los pies de los niños. Los resultados de un muestreo aleatorio fueron: 22.4, 23.4, 24.1, 21.6, 25.0, 20.5, 19.0

Solución:

$$X = \frac{X_1+X_2+X_3+\ldots\ldots\ldots\ldots+X_n}{n}$$

$$DM = \frac{|X_1- X| + |X_2 - X| + |X_3 - X| + \ldots\ldots\ldots + |X_n - X|}{n}$$

$$X = \frac{22.4+23.0+24.0+21.6+25.0+20.0+18.0}{7} = \frac{154}{7} = 22$$

$$DM = \frac{|22.4-22|+|23.0-22|+|24.0-22|+|21.6-22|+|25.0-22|+|20.0-22|+|18.0-22|}{7} =$$

$$DM = \frac{|0.4|+|1|+|2|+|-0.4|+|3|+|-2|+|-4|}{7} = \frac{0.4+1+2+0.4+3+2+4}{7} = \frac{12.8}{7} = 1.83$$

3. Hallar la desviación media del conjunto de datos: 2, 3, 6, 8 y 11

Solución:

$$X = \frac{X_1+X_2+X_3+\ldots\ldots\ldots\ldots+X_n}{n}$$

$$DM = \frac{|X_1- X| + |X_2 - X| + |X_3 - X| + \ldots\ldots\ldots + |X_n - X|}{n}$$

$$X = \frac{2+3+6+8+11}{5} = \frac{30}{5} = 6$$

$$DM = \frac{|2-6|+|3-6|+|6-6|+|8-6|+|11-6|}{5} = \frac{|-4|+|-3|+|0|+|2|+|5|}{5} = \frac{4+3+0+2+5}{5} = \frac{14}{5} = 2.8$$

4. Encontrar la desviación media de los conjuntos de números: 12, 6, 7, 3, 15, 10, 18 y 5

Solución:

$$X = \frac{X_1+X_2+X_3+\ldots\ldots\ldots\ldots+X_n}{n}$$

$$DM = \frac{|X_1- X| + |X_2 - X| + |X_3 - X| + \ldots\ldots\ldots + |X_n - X|}{n}$$

$$X = \frac{12+6+7+3+15+10+18+5}{8} = \frac{76}{8} = 9.5$$

$$DM = \frac{|12-9.5|+|6-9.5|+|7-9.5|+|3-9.5|+15-9.5|+|10-9.5|+|18-9.5|+|5-9.5|}{8} =$$

$$DM = \frac{|2.5|+|-3.51+|-2.5|+|-6.5|+|5.5|+|0.5|+|8.5|+|-4.5|}{8} =$$

$$DM = \frac{2.5+3.5+2.5+6.5+5.5+0.5+8.5+4.5}{8} = \frac{34}{8} = 4.25$$

5. Hallar la desviación media del conjunto de datos: 28, 16, 5, 2, 24 y 9

Solución:

$$\bar{X} = \frac{X_1+X_2+X_3+\ldots\ldots\ldots\ldots+X_n}{n}$$

$$DM = \frac{|X_1 - X| + |X_2 - X| + |X_3 - X| + \ldots\ldots\ldots + |X_n - X|}{n}$$

$$X = \frac{28+16+5+2+24+9}{6} = \frac{84}{6} = 14$$

$$DM = \frac{|28-14|+|16-14|+|5-14|+|2-14|+|24-14|+|9-14|}{6} = \frac{|14|+|2|+|-9|+|-12|+|10|+|-5|}{6} =$$

$$DM = \frac{14+2+9+12+10+5}{6} = \frac{52}{6} = 8.67$$

6. Hallar la desviación estándar del conjunto de números: 5, 8, 10, 12, 9, 4 y 7

Solución:

$$X = \frac{X_1+X_2+X_3+\ldots\ldots\ldots\ldots+X_n}{n}$$

$$DM = \frac{|X_1 - X| + |X_2 - X| + |X_3 - X| + \ldots\ldots\ldots + |X_n - X|}{n}$$

$$X = \frac{5+8+10+13+9+4+7}{7} = \frac{56}{7} = 8$$

$$DM = \frac{|5-8|+|8-8|+|10-8|+|13-8|+|9-8|+|4-8|+|7-8|}{7} = \frac{|-3|+|0|+|2|+|5|+|1|+|-4|+|-1|}{7} =$$

$$DM = \frac{3+0+2+5+1+4+1}{7} = \frac{16}{7} = 2.28$$

7. Hallar la desviación media de los números: 5, 28, 9, 14, 18 y 2

Solución:

$$X = \frac{X_1+X_2+X_3+\ldots\ldots\ldots\ldots+X_n}{n}$$

$$X = \frac{5+24+9+14+16+4}{6} = \frac{72}{6} = 12$$

$$DM = \frac{|5-12|+|24-12|+|9-12|+|14-12|+|16-12|+|4-12|}{6} = \frac{|-7|+|12|+|-3|+|2|+|4|+|-8|}{6} =$$

$$DM = \frac{7+12+3+2+4+8}{6} = \frac{36}{6} = 6$$

8. Hallar la desviación media de los pacientes que atendieron en el hospital mes de septiembre: 86, 95, 89, 93, 90, 91, 58, 61

Solución:

$$X = \frac{X_1+X_2+X_3+\ldots\ldots\ldots\ldots+Xn}{n}$$

$$DM = \frac{|X_1 - X| + |X_2 - X| + |X_3 - X| + \ldots\ldots\ldots + |X_n - X|}{N}$$

$$X = \frac{76+85+86+93+90+91+58+61}{8} = \frac{663}{8} = 80$$

$$DM = \frac{|76-80|+|85-80|+|86-80|+|93-80|+|90-80|+|91-80|+|58-80|+|61-80|}{8} =$$

$$DM = \frac{|-4|+|5|+|6|+|13|+|10|+|9|+|-22|+|19|}{8} = \frac{|4|+|5|+|6|+|13|+|10|+|9|+|22|+|19|}{8} = \frac{88}{8} = 11$$

TAREA:

I. Encontrar la desviación media de los conjuntos de datos:
   1. 12, 15, 9, 6 y 13
   2. 15, 7, 10, 4, 9, 13 y 5
   3. 16, 18, 14, 5, 25, 17, 15 y 10
   4. 6, 9, 3, 10, 4, 8, 2, 14 y 7
   5. 23, 27, 32, 19 y 34

<u>Desviación media para datos agrupados (con frecuencias).</u>

Si los datos vienen agrupados en una tabla de frecuencias, la expresión de la desviación media es: DM_f

$$DM_f = \frac{1}{N} \sum_{i=1}^{N} f_i |x_i - x| \qquad DM_f = \sum_{i=1}^{N} \frac{f_i|x_i - x|}{N} \qquad DM_f = \frac{f_i |x_i - x|}{N = \Sigma fi}$$

$$DM_f = \frac{f_1|x_1-x|+f_2|x_2-x|+x_3+x|+\ldots\ldots\ldots+f_N |x_N-x|}{f_1+f_2+f_3+\ldots\ldots\ldots+ f_N}$$

dónde:

DM_f es la desviación media con frecuencia

f_i es la frecuencia

N es el tamaño de la muestra o el número de elementos de frecuencias

$N= \Sigma_f$ es la suma de las frecuencias

$|X_i-X|$ es el valor absoluto de las diferencias entre el i-esimo valor numérico y la media aritmética.

Ejemplos:

1. Encontrar la desviación media para datos agrupados de la siguiente distribución de datos: 7, 4, 10, 3 y 6, ocurren con frecuencias 2, 1, 8, 5 y 9

Solución:

$$X = \frac{X_1+X_2+X_3+\ldots\ldots\ldots\ldots+X_n}{n}$$

$$DM_f = \frac{f_1|x_1-x|+f_2|x_2-x|+x_3+x|+\ldots\ldots\ldots+f_n |x_N-x|}{f_1+f_2+f_3+\ldots\ldots\ldots+ f_n}$$

$$X = \frac{7+4+10+3+6}{5} = \frac{30}{5} = 6$$

$$DM_f = \frac{2|7-6|+1|4-6|+8|10-6|+5|3-6|+9|6-6|}{2+1+8+5+9} = \frac{2|1|+1|-2|+8|4|+5|-3|+9|0|}{25} =$$

$$DMf = \frac{2(1)+1(2)+8(4)+5(3)+9(0)}{25} = \frac{2+2+32+15+0}{25} = \frac{51}{25} = 2.04$$

2. Encontrar la desviación media para datos agrupados de los siguientes números: 12, 10, 6, 8, 14 y 4, con frecuencias: 2, 1, 3, 1, 1 y 2

Solución:

$$X = \frac{X_1+X_2+X_3+\dots\dots\dots+X_n}{n}$$

$$DM_f = \frac{f_1|x_1-x|+f_2|x_2-x|+x_3+x|+\dots\dots\dots+f_n\,|x_N-x|}{f_1+f_2+f_3+\dots\dots\dots+f_n}$$

$$X = \frac{12+10+6+8+14+4}{6} = \underline{54} = 9$$

$$DM_f = \frac{2|12-9|+1|10-9|+3|6-9|+1|8-9|+1|14-9|+2|4-9|}{2+1+3+1+1+2} =$$

$$DM_f = \frac{2|3|+1|1|+3|-3|+1|-1|+1|5|+2|-5|}{10} = \frac{2(3)+1(1)+3(3)+1(1)+1(5)+2(5)}{10} =$$

$$DM_f = \frac{6+1+9+1+5+10}{10} = \frac{32}{10} = 3.2$$

3. Calcule la desviación media para datos agrupados, del conjuntos de números: 6, 2, 15, 7 y 10, que ocurren con frecuencias 3, 5, 1, 4 y 6

Solución:

$$X = \frac{X_1+X_2+X_3+\dots\dots\dots+X_n}{n}$$

$$DM_f = \frac{f_1|x_1-x|+f_2|x_2-x|+x_3+x|+\dots\dots\dots+f_n\,|x_N-x|}{f_1+f_2+f_3+\dots\dots\dots+\quad f_n}$$

$$X = \frac{6+2+15+7+10}{5} = \frac{40}{5} = 8$$

$$DM_f = \frac{3|6-8|+5|2-8|+1|15-8|+4|7-8|+6|10-8|}{3+5+1+4+6} = \frac{3|-2|+5|-6|+1|7|+4|-1|+6|2|}{19} =$$

$$DM_f = \frac{3|2|+5|6|+1|7|+4|1|+6|2|}{19} = \frac{6+30+7+4+12}{19} = \frac{59}{19} = 3.10$$

4. ¿Cuál es la desviación media para datos agrupados de las estaturas de 100 estudiantes de la universidad: 49, 53, 55, 60 y 63, teniendo en cuenta sus frecuencias 6, 17, 41, 28 y 8?

Solución:

$$X = \frac{X_1+X_2+X_3+\dots\dots\dots+X_n}{n}$$

$$DM_f = \frac{f_1|x_1-x|+f_2|x_2-x|+x_3+x|+\dots\dots\dots+f_n\,|x_N-x|}{f_1+f_2+f_3+\dots\dots\dots+\quad f_n}$$

$$X = \frac{49+53+55+60+63}{6+18+41+28+8} = \frac{280}{100} = 28$$

$$DM_f = \frac{6|49-28|+17|53-28|+41|55-28|+28|55-28|+8|63-28|}{6+18+41+28+8} =$$

$$DM_f = \frac{6|21|+17|25|+41|27|+28|32|+8|35|}{100} = \frac{126+425+1{,}107+896+280}{100} = \frac{2{,}834}{100} =$$

$$DM_f = 28.34$$

5. Roció invito a Juan a divertirse una tarde después de clases en una feria con juegos mecánicos y los demás locales de diversión, Roció y Juan lanzaron 20 dardos en el tiro al blanco (diana), lo cual esta enumerado del 5 al 10. El resultado después de los 20 lanzamientos de cada uno de ellos se registra

en las siguientes tablas: Encontrar la desviación media de cada uno de ellos, para encontrar al mejor tirador.

Puntuaciones de Rocío

"f"	**"x"**	**"fx"**
1	**10**	**10**
2	**9**	**18**
8	**8**	**64**
5	7	35
3	6	18
1	5	5
20		150

Puntuaciones de Juan

"f"	**"x"**	**"fx"**
2	10	20
4	9	36
5	8	40
3	7	21
3	6	18
3	5	15
20		150

$$X_f = \frac{\Sigma fx}{N} = \frac{150}{20} = 7.5 \qquad\qquad X_f = \frac{\Sigma fx}{N} = \frac{150}{20} = 7.5$$

$$DM_f = \frac{f_1|x_1-x|+f_2|x_2-x|+x_3+x|+\dots\dots\dots+f_n\,|x_N-x|}{f_1+f_2+f_3+\dots\dots\dots+\quad f_n}$$

$$DM_R = \frac{1|10-7.5|+2|9-7.5|+8|8-7.5|+5|7-7.5|+3|6-7.5|+1|5-7.5|}{1+2+8+5+3+1} =$$

$$DM_R = \frac{1|2.5|+2|1.5|+8|0.5|+5|-0.5|+3|-1.5|+1|-2.5|}{20} = \frac{19}{20} = 0.95$$

$$DM_J = \frac{2|10-7.5|+4|9-7.5|+5|8-7.5|+3|7-7.5|+3|6-7.5|+3|5-7.5|}{2+4+5+3+3+3} =$$

$$DM_J = \frac{2|2.5|+4|1.5|+5|0.5|+3|-0.5|+3|-1.5|+3|-2.5|}{20} = \frac{27}{20} = 1.35$$

Juan fue el mejor tirador

6. Encontrar la desviación media para datos agrupados de las estaturas de las mujeres de la siguiente tabla:

Estatura (pulg.)	**Marca de clase (X)**	**Frecuencia f**	**fx**
70-72	71	5	32.25
73-75	74	18	62.10
76-78	77	42	18.90
79- 81	80	27	68.85
82-84	83	8	44.40
		100	226.50

$$X_f = \frac{\Sigma fx}{N} = \frac{226.50}{100} = 2.26$$

$$DM_f = \frac{5|71-2.26|+18|74-2.26|+42|77-2.26|+27|80-2.26|+8|83-2.26|}{100} =$$

$$DM_f = \frac{5|68.74|+18|71.74|+42|74.74|+27|77.74|+8|80.74|}{100} =$$

$$DM_f = \frac{343.70+1,291.32+3,139.08+2,098.98+645.92}{100} = \frac{7,519}{100} = 75.19$$

1.6.3 VARIANZA PARA UNA MUESTRA PARA DATOS NO AGRUPADOS

Es un concepto estadístico sumamente importante, ya que muchas de las pruebas cuantitativas se fundamentan en él. Diversos métodos parten de la descomposición de la varianza. Sin embargo, para fines descriptivos se utiliza preferentemente la desviación estándar.

La varianza es la suma de los cuadrados de las diferencias entre cada dato y la media aritmética, dividido entre el tamaño de la muestra menos uno y se representa con S2, para encontrar la varianza de una muestra.

Al promedio de las desviaciones cuadráticas se le llama varianza.

La varianza y la desviación estándar también se basan en la suma de las desviaciones, pero elevándolas al cuadrado, ya que al multiplicar una desviación negativa por si misma da un numero positivo. La varianza y la desviación estándar son las medidas de variabilidad más ampliamente usadas.

La Varianza es la media aritmética de las desviaciones de la media elevadas al cuadrado.

La Varianza y la desviación estándar poblacional y muestral difieren en sus cálculos, por lo que se abordan independientemente.

La varianza es más utilizada en estadística inferencial entre otras aplicaciones, para calcular el tamaño de una muestra.

Distinguimos dos símbolos para identificar la varianza s2 para datos muéstrales y σ2 para datos poblacionales.

La fórmula para la varianza muestral presenta en su denominador al tamaño de la muestra menos uno, tendencia adoptada por los estadísticos para denotar una varianza más conservadora.

$$S^2 = \sum_{i=1}^{n} \frac{(X_i - X)^2}{n-1} \qquad S^2 = \frac{(X_1-X)^2+(X_2-X)^2+(X_3+X)^2+\ldots\ldots\ldots+(Xn-X)^2}{n-1}$$

donde :

S^2 es la varianza para una muestra

X_i es la i–esima observación de la variable “X” de la muestra

i=1 es el valor numérico 1, 2, 3,…………….n

X es la media aritmética de la muestra

n es el tamaño de la muestra o número de elementos de la muestra

$|X_i-X|$ es la diferencia entre el i-esimo valor numérico y la media aritmética, que se conoce como las desviaciones cuadráticas.

Ejemplos:

1. Una muestra tiene los siguientes datos: 3, 8, 7, 5, 9 y 4, calcular la varianza para una muestra:

Solución:

$$X = \frac{X_1+X_2+X_3+\ldots\ldots\ldots+Xn}{n}$$

$$S^2 = \frac{(X_1-X)^2+(X_2-X)^2+(X_3+X)^2+\ldots\ldots\ldots+(Xn-X)^2}{n-1}$$

$$X = \frac{3+8+7+5+9+4}{6} = \frac{36}{6} = 6$$

$$S^2 = \frac{(3-6)^2+(8-6)^2+(7-6)^2+(5-6)^2+(9-6)^2+(4-6)^2}{6-1} = \frac{(-3)^2+(2)^2+(1)^2+(-1)^2+(3)^2+(-2)^2}{5} =$$

$$S^2 = \frac{9+4+1+1+9+4}{5} = \frac{28}{5} = 5.6$$

2. Calcular la varianza de un conjunto de datos: 4, 1, 11, 13, 2 y 7

Solución:

$$X = \frac{X_1+X_2+X_3+\ldots\ldots\ldots\ldots+Xn}{n}$$

$$S^2 = \frac{(X_1-X)^2+(X_2-X)^2+(X_3+X)^2+\ldots\ldots\ldots+(Xn-X)^2}{n-1}$$

$$X = \frac{4+1+11+13+2+7}{6} = \frac{38}{6} = 6.33$$

$$S^2 = \frac{(4-6.33)^2+(1-6.33)^2+(11-6.33)^2+(13-6.33)^2+(2-6.33)^2+(7-6.33)^2}{6-1} =$$

$$S^2 = \frac{(-2.33)^2+(-5.33)^2+(4.67)^2+(6.67)^2+(-4.33)^2+(0.67)^2}{5} =$$

$$S^2 = \frac{5.43+28.41+21.81+44.49+18.75+0.45}{5} = \frac{119.34}{5} = 23.87$$

3. En una granja se mide el total de litros diarios de leche que producen 100 vacas en el establo. Los resultados diarios de la primera semana de mayo del 2016, fueron: 400, 435, 450, 420, 410 y 440. Encontrar la varianza, del promedio de litros diarios de leche producidos por las 100 vacas.

Solución:

$$X = \frac{X_1+X_2+X_3+\ldots\ldots\ldots\ldots+Xn}{n}$$

$$S^2 = \frac{(X_1-X)^2+(X_2-X)^2+(X_3+X)^2+\ldots\ldots\ldots+(Xn-X)^2}{n-1}$$

$$X = \frac{40+50+46+42+52+}{5} = \frac{230}{5} = 46$$

$$S^2 = \frac{(40-46)^2+(50-46)^2+(46-46)^2+(42-46)^2+(52-46)^2}{5-1} = \frac{(-6)^2+(4)^2+(0)^2+(-4)^2+(6)^2}{4} =$$

$$S^2 = \frac{36+16+0+16+36}{4} = \frac{104}{4} = 26$$

4. Encontrar la varianza de la distribución de datos no agrupados:
   6, 8, 14, 18 y 24

Solución:

$$X = \frac{X_1+X_2+X_3+\ldots\ldots\ldots\ldots+Xn}{n}$$

$$S^2 = (X_1-X)^2+(X_2-X)^2+(X_3+X)^2+\ldots\ldots\ldots+(Xn-X)^2$$

n-1

$$X = \frac{6+8+14+18+24}{5} = \frac{70}{5} = 14$$

$$S^2 = \frac{(6-14)^2+(8-14)^2+(14-14)^2+(18-14)^2+(24-14)^2}{5-1} = \frac{(-8)^2+(-6)^2+(0)^2+(4)^2+(10)^2}{4} =$$

$$S^2 = \frac{64+36+0+16+100}{4} = 54$$

5. Calcular la varianza del siguiente conjunto de datos de una muestra: 6, 3, 8, 4, 15 y 12

Solución:

$$X = \frac{X_1+X_2+X_3+\ldots\ldots\ldots\ldots+Xn}{n}$$

$$S^2 = \frac{(X_1-X)^2+(X_2-X)^2+(X_3+X)^2+\ldots\ldots\ldots+(Xn-X)^2}{n-1}$$

$$X = \frac{6+3+8+4+15+12}{6} = \frac{36}{6} = 6$$

$$S^2 = \frac{(6-6)^2+(3-6)^2+(8-6)^2+(4-6)^2+(15-6)^2+(12-6)^2}{6-1} = \frac{(0)^2+(-3)^2+(2)^2+(-2)^2+(9)^2+(6)^2}{5} =$$

$$S^2 = \frac{0+9+4+4+81+36}{5} = \frac{134}{5} = 26.8$$

6. Los datos representan la edad de los miembros de un grupo de niños: 5, 8, 2, 7, 4, 6 y 10. Encontrar la varianza para datos no agrupados.

Solución:

$$X = \frac{X_1+X_2+X_3+\ldots\ldots\ldots\ldots+Xn}{n}$$

$$S^2 = \frac{(X_1-X)^2+(X_2-X)^2+(X_3+X)^2+\ldots\ldots\ldots+(Xn-X)^2}{n-1}$$

$$X = \frac{5+8+2+7+4+6+10}{7} = \frac{42}{7} = 6$$

$$S^2 = \frac{(5-6)^2+(8-6)^2+(2-6)^2+(7-6)^2+(4-6)^2+(6-6)^2+(10-6)^2}{7-1} =$$

$$S^2 = \frac{(-1)^2+(2)^2+(-4)^2+(1)^2+(-2)^2+(0)^2+(4)^2}{6} = \frac{1+4+16+1+4+0+16}{6} = \frac{42}{6} = 7$$

7. Una muestra tiene los siguientes datos: 8, 12, 6, 10 y 4. Encontrar la varianza de una muestra para datos no agrupados.

Solución:

$$X = \frac{X_1+X_2+X_3+\ldots\ldots\ldots\ldots+Xn}{n}$$

$$S^2 = \frac{(X_1-X)^2+(X_2-X)^2+(X_3+X)^2+\ldots\ldots\ldots+(Xn-X)^2}{n-1}$$

$$X = \frac{8+12+6+10+4}{5} = \frac{40}{5} = 8$$

$$S^2 = \frac{(8-8)^2+(12-8)^2+(6-8)^2+(10-8)^2+(4-8)^2}{5-1} = \frac{(0)^2+(4)^2+(-2)^2+(2)^2+(-4)^2}{4} = \frac{0+16+4+4+16}{4} =$$

$$S^2 = \underline{40} = 10$$

4

8. Los datos representan la edad de los miembros de un grupo de niños: 2, 9, 4, 7, 3 y 5. Encontrar la varianza para datos no agrupados.

Solución:

$$X = \frac{X_1+X_2+X_3+\ldots\ldots\ldots\ldots+X_n}{n}$$

$$S^2 = \frac{(X_1-X)^2+(X_2-X)^2+(X_3+X)^2+\ldots\ldots\ldots+(X_n-X)^2}{n-1}$$

$$X = \frac{2+9+4+7+3+5}{6} = \frac{30}{6} = 5$$

$$S^2 = \frac{(2-5)^2+(9-5)^2+(4-5)^2+(7-5)^2+(3-5)^2+(5-5)^2}{6-1} = \frac{(-3)^2+(4)^2+(-1)^2+(2)^2+(-2)^2+(0)^2}{5} =$$

$$S^2 = \frac{9+16+1+4+4+0}{5} = \frac{34}{5} = 6.8$$

TAREA: Encontrar la varianza de las siguientes distribuciones para una muestra:

1. 3, 10, 2, 8, 4, 1, 7 y 5
2. 12, 8, 5, 11 y 4
3. 4, 7, 5, 9, 3 y 8
4. 13, 10, 17, 9 y 6
5. 28, 16, 5, 2 y 24

La varianza poblacional para un número finito o de tamaño N de datos no agrupados es:

$$\sigma^2 = \sum_{i=1}^{N} \frac{(X_i-\mu)^2}{N} \qquad \sigma^2 = \frac{(X_1-\mu)^2+(X_2-\mu)^2+(X_3+\mu)^2+\ldots\ldots\ldots+(X_n-\mu)^2}{N}$$

donde:

σ^2 es la varianza para una población.
X_i es la i–esima observación de la variable "X" de la población.
i=1 es el valor numérico 1, 2, 3,…………….n
μ es la media aritmética de la población.
N es el tamaño de la muestra o número de elementos de la población.
$|X_i-X|$ es la diferencia entre el i-esimo valor numérico y la media aritmética, que se conoce como las desviaciones cuadráticas.

La letra σ es griega y se llama sigma (la media aritmética de una población se presenta con el símbolo "u" (mu).

La varianza es un concepto abstracto, siendo sus unidades cuadráticas (m^2, kg^2, etc.), para medir la dispersión de los datos en unidades "normales", se calcula la raíz cuadrada de la varianza, que se denomina desviación estándar.

Ejemplos:

1. Encontrar la varianza para cada uno de los niños con los siguientes datos, que estuvieron en espera de entrar a pre-escolar y primaria de una población: 4, 7, 3, 10, 5, 8 y 12

Solución:

$$\mu = \frac{X_1+X_2+X_3+\ldots\ldots\ldots\ldots+X_N}{N}$$

$$\sigma^2 = \frac{(X_1 - \mu)^2 + (X_2 - \mu)^2 + (X_3 + \mu)^2 + \dots\dots\dots + (Xn - \mu)^2}{N}$$

$$\mu = \frac{4+7+3+10+5+8+12}{7} = \frac{49}{7} = 7$$

$$\sigma^2 = \frac{(4-7)^2+(7-7)^2+(3-7)^2+(10-7)^2+(5-7)^2+(8-7)^2+(12-7)^2}{7} =$$

$$\sigma^2 = \frac{(-3)^2+(0)^2+(-4)^2+(3)^2+(-2)^2+(1)^2+(5)^2}{7} = \frac{9+0+16+9+4+1+25}{7} = \frac{64}{7} = 9.14$$

2. Busque la varianza para el conjunto de datos de una población, para datos no agrupados: 8, 11, 7, 10 y 9

Solución:

$$\mu = \frac{X_1+X_2+X_3+\dots\dots\dots\dots+X_N}{N}$$

$$\sigma^2 = \frac{(X_1 - \mu)^2 + (X_2 - \mu)^2 + (X_3 + \mu)^2 + \dots\dots\dots + (Xn - \mu)^2}{N}$$

$$\mu = \frac{8+11+7+10+9}{5} = \frac{45}{5} = 9$$

$$\sigma^2 = \frac{(8-9)^2+(11-9)^2+(7-9)^2+(10-9)^2+(9-9)^2}{5} = \frac{(-1)^2+(2)^2+(-2)^2+(1)^2+(0)^2}{5} = \frac{1+4+4+1+0}{5} =$$

$$\sigma^2 = \frac{10}{5} = 2$$

3. Hallar la varianza de la población con la siguiente distribución de datos: 20, 40, 10, 70 y 50

Solución:

$$\mu = \frac{X_1+X_2+X_3+\dots\dots\dots\dots+X_N}{N}$$

$$\sigma^2 = \frac{(X_1 - \mu)^2 + (X_2 - \mu)^2 + (X_3 + \mu)^2 + \dots\dots\dots + (Xn - \mu)^2}{N}$$

$$\mu = \frac{20+40+10+70+50}{5} = \frac{200}{5} = 40$$

$$\sigma^2 = \frac{(20-40)^2+(40-40)^2+(10-40)^2+(70-40)^2+(50-40)^2}{5} = \frac{(-20)^2+(0)^2+(-30)^2+(30)^2+(10)^2}{5} =$$

$$\sigma = \frac{400+0+900+900+100}{5} = \frac{2,300}{5} = 460$$

4. Con los siguientes datos de una población: 9, 6,1, 7, 3 y 10. Encontrar la varianza

Solución:

$$\mu = \frac{X_1+X_2+X_3+\dots\dots\dots\dots+X_N}{N}$$

$$\sigma^2 = \frac{(X_1 - \mu)^2 + (X_2 - \mu)^2 + (X_3 + \mu)^2 + \dots\dots\dots + (Xn - \mu)^2}{N}$$

$$\mu = \frac{9+6+1+7+3+10}{6} = \frac{36}{6} = 6$$

$$\sigma^2 = \frac{(9-6)^2+(6-6)^2+(1-6)^2+(7-6)^2+(3-6)^2+(10-6)^2}{6} = \frac{(3)^2+(0)^2+(-5)^2+(1)^2+(-3)^2+(4)^2}{6} =$$

$$\sigma^2 = \frac{9+0+25+1+9+16}{6} = \frac{60}{6} = 10$$

5. En el censo de una población se encontraron: 80, 30, 50 y 40 mujeres sin registrarse en los años 2015 y 2016, encontrar la desviación estándar

Solución:

$$\mu = \frac{X_1+X_2+X_3+\ldots\ldots\ldots\ldots+X_N}{N}$$

$$\sigma^2 = \frac{(X_1-\mu)^2+(X_2-\mu)^2+(X_3+\mu)^2+\ldots\ldots\ldots+(Xn-\mu)^2}{N}$$

$$\mu = \frac{80+30+50+40}{4}=\frac{200}{4}=50$$

$$\sigma^2=\frac{(80-50)^2+(30-50)^2+(50-50)^2+(40-50)^2}{4}=\frac{(30)^2+(-20)^2+(0)^2+(-10)^2}{4}=\frac{900+400+0+100}{4}$$

$$\sigma^2 = \frac{1,400}{4}=350$$

NOTA: La varianza muestral y la varianza poblacional es el mismo procedimiento, cambia nada más la simbología, cuando es una muestra **X** y $\mathbf{S^2}$, cuando es una población **μ(mu)** y $\boldsymbol{\sigma^2}$ **sigma.**

TAREA: Encontrar la varianza poblacional de los siguientes conjuntos de datos:

1. 6, 2, 5, 3, 8, 4 y 7
2. 7, 10, 3, 6 y 4
3. 11, 14, 17, 9, 1 y 8
4. 13, 17, 21 y 19
5. 10, 13, 9, 17 y 21

1.6.4 DESVIACIÓN ESTÁNDAR

La desviación estándar es la raíz cuadrada de la varianza s^2. Es decir, la raíz cuadrada de la media de los cuadrados de las puntuaciones de desviación.
La desviación estándar es una medida de la dispersión de un conjunto de datos alrededor de la media. Para obtener la desviación estándar se empieza por restar la media de cada uno de los puntos, con lo cual se llega a una serie de valores denominados puntos de desviación. Luego se elevan al cuadrado estos puntos de desviación, se suman los cuadrados y se divide la suma por el número de valores que integran la serie, con el fin de obtener la desviación cuadrática media. La desviación estándar es la raíz cuadrada de la desviación cuadrática media o varianza.

La fórmula para la desviación estándar para una muestra:

$$S = \frac{\sqrt{(x1-x)2+(x2-x)2+(x3-x)2+\cdots\ldots\ldots\ldots\ldots+(xn-x)2}}{n-1} = \frac{\sum_{i1}^{n}(x_i-x)^2}{n-1}$$

donde:

S es la desviación estándar para una muestra

X_i es la i–esima observación de la variable “X” de la muestra

i=1 es el valor numérico 1, 2, 3,…………….n
X es la media aritmética de la muestra
n es el tamaño de la muestra o número de elementos de la muestra
$|X_i-X|$ es la diferencia entre el i-esimo valor numérico y la media aritmética, que se conoce como las desviaciones cuadráticas.

Desviación Estándar

Esta medida nos permite determinar el promedio aritmético de fluctuación de los datos respecto a su punto central o media. La desviación estándar nos da como resultado un valor numérico que representa el promedio de diferencia que hay entre los datos y la media. Para calcular la desviación estándar basta con hallar la raíz cuadrada de la varianza, por lo tanto, su ecuación sería:

$S = \sqrt{S^2}$ Ecuación 5-8

Para comprender el concepto de las medidas de distribución vamos a suponer que el gerente de una empresa de alimentos desea saber que tanto varían los pesos de los empaques (en gramos), de uno de sus productos; por lo que opta por seleccionar al azar cinco unidades de ellos para pesarlos. Los productos tienen los siguientes pesos (490, 500, 510, 515 y 520) gramos respectivamente.

Por lo que su media es:

$$\overline{X} = \frac{490+500+510+515+520}{5} = \frac{2535}{5} = 507$$

La varianza sería:

$$S^2 = \frac{(490-507)^2+(500-507)^2+(510-507)^2+(515-507)^2+(520-507)^2}{(5-1)}$$

$$S^2 = \frac{(-17)^2+(-7)^2+(3)^2+(8)^2+(13)^2}{4} = \frac{289+49+9+64+169}{4} = \frac{580}{4} = 145$$

Por lo tanto la desviación estándar sería:

$$S = \sqrt{145} = 12.04 \cong 12$$

Con lo que concluiríamos que el peso promedio de los empaques es de 507 gramos, con una tendencia a variar por debajo o por encima de dicho peso en 12 gramos. Esta información le permite al gerente determinar cuánto es el promedio de perdidas causado por el exceso de peso en los empaques y le da las bases para tomar los correctivos necesarios en el proceso de empacado.

Ejemplos:

1. Calcular la desviación estándar del conjunto de datos no agrupados: 4, 7, 5, 3 y 6

Solución:

$X = \underline{X_1+X_2+X_3+\ldots\ldots\ldots\ldots+Xn}$

$$S = \frac{\sqrt{(X1-X)2+(X2-X)2+(X3-X)2+\cdots\ldots\ldots\ldots\ldots+(Xn-X)2}}{n-1} = \frac{\sum_{i1}^{n}(x_i-x)^2}{n-1}$$

$X = \frac{4+7+5+3+6}{5} = \frac{25}{5} = 5$

$S^2 = \frac{(4-5)^2+(7-5)^2+(5-5)^2+(3-5)^2+(6-5)^2}{5-1} = \frac{(-1)^2+(2)^2+(0)^2+(-3)^2+(1)^2}{4} = \frac{1+4+0+9+1}{4} = \frac{15}{4}$

$S^2 = \frac{\sqrt{15}}{4} = 3.75$

S = 3.75

2. Busque la desviación estándar de los datos: 3, 6, 2, 1, 7 y 5

Solución:

$X = \frac{X_1+X_2+X_3+\ldots\ldots\ldots\ldots+Xn}{n}$

$X = \frac{3+6+2+1+7+5}{6} = \frac{24}{6} = 4$

$S^2 = \frac{(3-4)^2+(6-4)^2+(2-4)^2+(1-4)^2+(7-4)^2+(5-4)}{6-1} = \frac{(-1)^2+(2)^2+(-2)^2+(-3)^2+(3)^2+(1)^2}{5} =$

$S^2 = \frac{1+4+4+9+9+1}{5} = \frac{\sqrt{28}}{5} = 1.06$

S= 1.06

3. Encontrar la desviación estándar del conjunto de datos: 8, 11, 7, 6, 12 y 10

Solución:

$X = \frac{X_1+X_2+X_3+\ldots\ldots\ldots\ldots+Xn}{n}$

$X = \frac{8+11+7+6+12+10}{6} = \frac{54}{6} = 9$

$S^2 = \frac{(8-9)^2+(11-9)^2+(7-9)^2+(6-9)^2+(12-9)^2+(10-9)^2}{6-1} = \frac{(-1)^2+(2)^2+(-2)^2+(-3)^2+(3)^2+(1)^2}{5} =$

$S^2 = \frac{1+4+4+9+9+1}{5} = \frac{\sqrt{28}}{5} = 1.06$

S = 1.06

Encuentre la desviación estándar para el conjunto de números: 3, 7, 9 y 5

Solución:

$X = \frac{X_1+X_2+X_3+\ldots\ldots\ldots\ldots+Xn}{n}$

n

$X = \frac{3+7+9+5}{4} = \frac{24}{4} = 6$

$S^2 = \frac{(3-6)^2+(7-6)^2+(9-6)^2+(5-6)^2}{4-1} = \frac{(-3)^2+(1)^2+(3)^2+(-1)^2}{3} = \frac{9+1+9+1}{3} = \frac{\sqrt{20}}{3} = 1.49$

S = 1.49

4. Busque la desviación estándar para la distribución de los datos no agrupados:
 4, 10, 7, 16, 13, 19

Solución:

$X = \frac{X_1+X_2+X_3+\ldots\ldots\ldots\ldots+Xn}{n}$

$X = \frac{2+10+7+16+13+18}{6} = \frac{66}{6} = 11$

$S^2 = \frac{(2-11)^2+(10-11)^2+(7-11)^2+(16-11)^2+(13-11)^2+(18-11)^2}{6-1} =$

$S^2 = \frac{(-9)^2+(-1)^2+(-4)^2+(5)^2+(2)^2+(7)^2}{5} = \frac{81+1+16+25+4+49}{5} = \frac{\sqrt{176}}{5} = 2.65$

S= 2.65

TAREA:

I. Calcular la desviación estándar para los conjuntos de números de una muestra:

1. 3, 7, 4, 12 y 9
2. 12, 6, 7, 3, 15, 10, 18 y 5
3. 10, 2, 5, 7, 1, 12 y 9
4. 4, 13, 8, 6, 9, 12, 10, 3 y 1
5. 17, 11, 8, 15, 12 y 9

Desviación Estándar para una Población para Datos no Agrupados:

$\mu = \frac{X_1+X_2+X_3+\ldots\ldots\ldots\ldots+X_N}{N}$

$$\sigma = \sqrt{\frac{(x1 - \mu)2 + (x2 - \mu)2 + (x3 - \mu)2 + \cdots + (xn - u)2}{N}} = \sum_{i1}^{N} \frac{(x_i-u)^2}{N}$$

La Desviación Estándar se calcula extrayendo la raíz cuadrada (positiva) de la varianza.

$S^2 = S$ para una muestra $\qquad \sigma^2 = \sigma$ para una población

1.6.5 MÉTODO CORTO PARA CALCULAR LA DESVIACIÓN TÍPICA

La desviación típica o desviación estándar (denotada con el símbolo **σ** o **s**, dependiendo de la procedencia del conjunto de datos) es una medida de dispersión para variables de razón (variables cuantitativas o cantidades racionales) y de intervalo. Se define como la raíz cuadrada de la varianza de la variable.

Para conocer con detalle un conjunto de datos, no solo basta con conocer las medidas de tendencia central, sino que necesitamos conocer también la desviación que presentan los datos en su distribución respecto de la media aritmética de dicha distribución, con objeto de tener una visión de los mismos más acorde con la realidad al momento de describirlos e interpretarlos para la toma de decisiones.

La desviación típica es una medida del grado de dispersión de los datos con respecto al valor promedio. Dicho de otra manera, la desviación típica es simplemente el "promedio" o variación esperada con respecto a la media aritmética.

La desviación típica es una medida de dispersión usada en estadística que nos dice cuánto tienden a alejarse los valores concretos del promedio en una distribución de datos. De hecho, específicamente, el cuadrado de la desviación típica es "el promedio del cuadrado de la distancia de cada punto respecto del promedio". σ {\displaystyle \sigma _{}^{}}

Esta medida es más estable que el recorrido y toma en consideración el valor de cada dato. Es decir, la raíz cuadrada de la media de los cuadrados de las puntuaciones de desviación.

Las ecuaciones pueden expresarse, respectivamente, en las formas equivalentes:
Para una muestra:

$$X = \frac{X_1 + X_2 + X_3 + \ldots\ldots X_n}{n} = \frac{\Sigma X}{n}$$ La Media para una muestra

$$X^2 = \frac{X_1 + X_2 + X_3 + \ldots\ldots X_N}{N} = \frac{\Sigma X^2}{N}$$ La Media de los cuadrados

$$S = \sqrt{\frac{\Sigma X2 - (\Sigma X)2}{N}} = \sqrt{X2 - (X)2}$$

Dónde: X^2 denota la media de los cuadrados de los distintos valores de X, mientras que X^2 denota el cuadrado de la media de los distintos valores de X.
Ejemplos:
Encontrar la desviación típica del conjunto de datos: 2, 6, 7, 3, 15, 10, 18 y 5
Solución:

$$X = \frac{\Sigma X}{n} = \frac{12+6+7+3+15+10+18+5}{8} = \frac{76}{8} = 9.5$$

$$X^2 = \frac{\Sigma X^2}{N} = \frac{(12)^2+(6)^2+(7)^2+(3)^2+(15)^2+(10)^2+(18)^2+(5)^2}{8} =$$

$$X^2 = \frac{144+36+49+9+225+100+324+25}{8} = \frac{912}{8} = 114$$

$$S = \sqrt{114-(9.5)2} = \sqrt{114-90.25} = \sqrt{23.75} = 4.87$$

2. Hallar la desviación típica del conjunto de números: 9, 3, 8, 8, 9, 8, 9 y 18

Solución:

$$X = \frac{\Sigma X}{n} = \frac{9+3+8+8+9+8+9+18}{8} = \frac{72}{8} = 9$$

$$X^2 = \frac{\Sigma X^2}{N} = \frac{(9)^2+(3)^2+(8)^2+(8)^2+(9)^2+(8)^2+(9)^2+(18)^2}{8} =$$

$$X^2 = \frac{81+9+64+64+81+64+81+324}{8} = \frac{768}{8} = 96$$

$$S = \sqrt{96-(9)2} = \sqrt{96-81} = \sqrt{15} = 3.87$$

3. Encontrar la desviación típica del conjunto de datos:11, 7, 8, 4, 14, 9, 17 y 4

Solución:

$$X = \frac{\Sigma X}{n} = \frac{11+7+8+4+14+9+17+4}{8} = \frac{74}{8} = 9.25$$

$$X^2 = \frac{\Sigma X^2}{N} = \frac{(11)^2+(7)^2+(8)^2+(4)^2+(14)^2+(9)^2+(17)^2+(4)^2}{8} =$$

$$X^2 = \frac{121+49+64+16+196+81+289+16}{8} = \frac{832}{8} = 104$$

$$S = \sqrt{104-(9.25)2} = \sqrt{104-85.56} = \sqrt{18.437} = 4.29$$

4. Encontrar la desviación típica del conjunto de números:
   10, 8, 9, 5, 13, 11, 19 y 6

Solución:

$$X = \frac{\Sigma X}{n} = \frac{10+8+9+5+13+11+19+6}{8} = \frac{81}{8} = 10.12$$

$$X^2 = \frac{\Sigma X^2}{N} = \frac{(10)^2+(8)^2+(9)^2+(5)^2+(13)^2+(11)^2+(19)^2+(6)^2}{8} =$$

$$X^2 = \frac{100+64+81+25+169+121+361+36}{8} = \frac{957}{8} = 119.62$$

$$S = \sqrt{119.62-(10.12)2} = \sqrt{119.62-102.41} = \sqrt{17.21} = 4.14$$

5. Dado el conjunto de números: 2, 5, 8, 11 y 14

Solución:

$$X = \frac{\Sigma X}{n} = \frac{2+5+8+11+14}{5} = \frac{40}{5} = 8$$

$X^2 = \frac{\Sigma X^2}{N} = \frac{(2)^2+(5)^2+(8)^2+(11)^2+(14)^2}{5} = \frac{4+25+64+121+196}{5} = \frac{410}{5} = 82$

$S = \sqrt{82 - (8)2} = \sqrt{82 - 64} = \sqrt{18} = 4.24$

6. Hallar la desviación típica del conjunto de números: 6, 3, 9, 7 y 5

Solución:

$X = \frac{\Sigma X}{n} = \frac{6+3+9+7+5}{5} = \frac{30}{5} = 6$

$X^2 = \frac{\Sigma X^2}{N} = \frac{(6)^2+(3)^2+(9)^2+(7)^2+(5)^2}{5} = \frac{36+9+81+49+25}{5} = \frac{200}{5} = 40$

$S = \sqrt{40 - (6)2} = \sqrt{40 - 36} = \sqrt{4} = 2$

7. Encontrar la desviación típica del conjunto de datos: 8, 2, 10, 3 y 7

Solución:

$X = \frac{\Sigma X}{n} = \frac{8+2+10+3+7}{5} = \frac{30}{5} = 6$

$X^2 = \frac{\Sigma X^2}{N} = \frac{(8)^2+(2)^2+(10)^2+(3)^2+(7)^2}{5} = \frac{64+4+100+9+49}{5} = \frac{226}{5} = 45.2$

$S = \sqrt{45.2 - (6)2} = \sqrt{45.2 - 36} = \sqrt{9.2} = 3.03$

8. Hallar la desviación típica de los números: 4, 7, 5, 9, 3, 8 y 6

$X = \frac{\Sigma X}{n} = \frac{4+7+5+9+3+8+6}{7} = \frac{42}{7} = 6$

$X^2 = \frac{\Sigma X^2}{N} = \frac{(4)^2+(7)^2+(5)^2+(9)^2+(3)^2+(8)^2+(6)^2}{7} = \frac{16+28+25+81+9+64+36}{7} =$

$X^2 = \frac{259}{7} = 37$

$S = \sqrt{37 - (6)2} = \sqrt{37 - 36} = \sqrt{1} = 1$

TAREA:

I. Encontrar la desviación típica por el método corto de los siguientes conjuntos de datos:

1. 2, 5, 8, 11, 14, 2, 8 y 14
2. 3, 6, 2, 1, 7 y 5
3. 4, 1, 3, 9, 6, 2, 5 y 10
4. 7, 4, 8, 6, 2, 5 y 3
5. 5, 9, 7, 3 y 1

2. INTRODUCCIÓN A LA PROBABILIDAD Y VALOR ESPERADO

La presente investigación tiene como finalidad conocer el tema de la unidad presente titulada Introducción a la Probabilidad y valor esperado, como primera instancia cabe mencionar que la Probabilidad es el Cálculo matemático de las posibilidades que existen de que una cosa se cumpla o suceda al azar. El primer tema en desarrollarse es la teoría de conjuntos que busca que conozcamos qué es un conjunto y los tipos, sus principales operaciones, así como sus propiedades. En segundo lugar, se encuentra el tema de tipos de variables aleatorias que nos dan una muestra de cómo podemos graficar posibles soluciones a los problemas planteados y con esta identificar el tipo de variable que se presenta. En el tema puntos de la muestra nos indica los elementos que la componen, así como las fórmulas que se aplican para cada caso que queramos determinar en cuánto cálculo de predicciones sobre una muestra determinada. Las reglas aditivas y multiplicativas ayudarán a la compresión de las operaciones con las fórmulas presentadas y realizarlas de la manera adecuada. La regla de Bayes nos da las posibilidades que ocurra un evento de forma porcentual y de esta forma poder determinar mediante su formula el resultado. La importancia de esta investigación es que con ella conocemos los elementos que determinan y tratan de predecir eventos, útiles para aplicarlos a través del aprendizaje significativo.

2.1 TEORÍA DE CONJUNTOS

La teoría de conjuntos es una rama de la matemática que estudia las propiedades y relaciones de los conjuntos. Los conjuntos son colecciones abstractas de objetos, consideradas como objetos en sí mismas y son una herramienta básica en la formulación de cualquier teoría matemática.

La teoría de los conjuntos es lo suficientemente rica como para construir el resto de objetos y estructuras de interés en matemáticas: números, funciones, figuras geométricas.

Los conjuntos son un agregado o colección de objetos de cualquier naturaleza con características bien definidas de manera que se puedan distinguir todos sus elementos, por ejemplo: el conjunto de días de la semana, el conjunto de las vocales, etc. A los objetos que lo componen se le llama elementos del conjunto.

Un conjunto es una colección de objetos llamados miembros o elementos del conjunto.
La Teoría de Conjuntos es una teoría matemática, que estudia básicamente a un cierto tipo de objetos llamados conjuntos y algunas veces, a otros objetos denominados no conjuntos, así como a los problemas relacionados con estos.

El concepto de conjunto es intuitivo y se podría definir como una "agrupación bien definida de objetos no repetidos y no ordenados"; así, se puede hablar de un conjunto de personas, ciudades, gafas, lapiceros o del conjunto de objetos que hay en un momento dado encima de una mesa. Un conjunto está bien definido si se sabe si un determinado elemento pertenece o no al conjunto. El conjunto de los bolígrafos azules está bien definido, porque a la vista de un bolígrafo se puede saber si es azul o no. El conjunto de las personas altas no está bien definido, porque a la vista de una persona, no siempre se podrá decir si es alta o no, o puede haber distintas personas, que opinen si esa persona es alta o no lo es. En el siglo XIX, según Frege, los elementos de un conjunto se definían sólo por tal o cual propiedad. Actualmente la teoría de conjuntos está bien definida por el sistema ZFC. Sin embargo, sigue siendo célebre la definición que publicó Cantor.

¿Qué es un Conjunto**?**

Un conjunto es la agrupación, clase, o colección de objetos o en su defecto de elementos que pertenecen y responden a la misma categoría o grupo de cosas, por eso se los puede agrupar en el mismo conjunto. Esta relación de pertenencia que se establece entre los objetos o elementos es absoluta y posiblemente discernible y observable por cualquier persona. Entre los objetos o elementos susceptibles de integrar o conformar un conjunto se cuentan por supuesto cosas físicas, como pueden ser las mesas, sillas y libros, pero también por entes abstractos como números o letras.

Clases de Conjuntos

Conjunto Finito: Es el conjunto al que se le puede determinar su cardinalidad o puede llegar a contar su último elemento.

Ejemplo:
M= {*/X es divisor de 24}
M= {1, 2, 3, 4, 6, 8, 12, 24}

Conjunto Infinito: Es el conjunto que, por tener muchísimos elementos, no se le puede llegar a contar su último elemento.

Ejemplo:
A= {*/X sea grano de sal}

Conjunto Vacío: Es el conjunto cuya cardinalidad es cero ya que carece de elementos. El símbolo del conjunto vacío ϕ.

Ejemplo:
C= {*/X sea habitantes del sol}

Conjunto Unitario: Es el conjunto que solo tiene un elemento. Su cardinalidad es uno (1).

Ejemplo:
D={*/x sea vocal de la palabra "pez"}

Operaciones con conjuntos**.**

Unión de conjuntos:

La unión de los conjuntos A y B es el conjunto formado por todos los elementos que pertenecen a A o a B o a ambos. Se denota: A U B. La unión de conjuntos se define como:

A U B = {x / x € A o x € B}

Ejemplos:

Intersección de conjuntos:
La intersección es el conjunto formado por los elementos que son comunes entre dos o más conjuntos dados. Se denota por A∩B, que se lee: A intersección B.

La intersección de A y B también se puede definir:
A∩ B = {X/X € A y x € B}

Ejemplos:

2.1.1 DEFINICIÓN PROPIEDADES Y OPERACIONES BÁSICAS CON CONJUNTOS

Existen varias operaciones básicas que pueden realizarse, partiendo de ciertos conjuntos dados, obtener nuevos conjuntos.

Unión: (símbolo U), la unión de dos conjuntos A y B, que se representa como AUB, es el conjunto de todos los elementos que pertenecen al menos a uno de los conjuntos A Y B.
A U B = + {x/x∈AUX∈B}

Que es una Unión
En la teoría de conjuntos, la unión de dos (o más) conjuntos es una operación que resulta en otro conjunto cuyos elementos son los elementos de los conjuntos iniciales.

Operaciones con Conjuntos
Unión: La unión de dos ***conjuntos*** A y B la denotaremos por A ∪ B y es el conjunto formado por los elementos que pertenecen al menos a uno de ellos ó a los dos. Lo que se denota por: A ∪ B = { X/X ∈ A ó X ∈ B }

Ejemplo: Sean los ***conjuntos*** A= {1, 3, 5, 7, 9} y B= {10, 11, 12}
A ∪ B = {1, 3, 5, 7, 9, 10, 11, 12}

Que es una Intersección

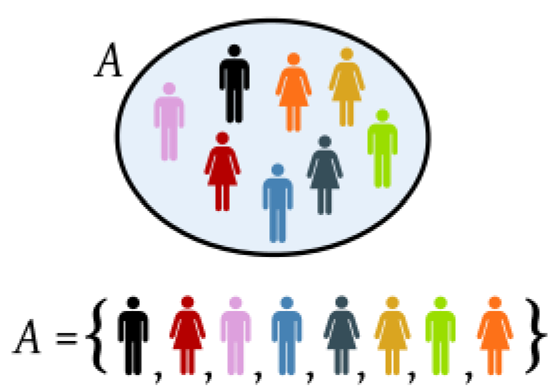
En teoría de conjuntos, la intersección de dos (o más) conjuntos es una operación que resulta en otro conjunto que contiene los elementos comunes a los conjuntos de partida.

Intersección: (símbolo ∩), la intersección de dos conjuntos A y B, es el conjunto $A \cap B$, de los elementos comunes A y B.
Ejemplo: Sean A= {1, 2, 3, 4, 5, 6, 8, 9} y B= {2, 4, 8, 12} Los elementos comunes a los dos ***conjuntos*** son: {2, 4, 8}. A este conjunto...
Diferencia: (símbolo /), la diferencia del conjunto A con B es el conjunto A / B, que resulta de eliminar de A cualquier elemento que este en B.

Complemento**:** El complemento de un conjunto A es el conjunto AC que contiene todos los elementos que no pertenecen a A, respecto a un conjunto U que lo contiene.
$AC = \{X/X \in U \cap X \notin A\}$

Operaciones Básicas de Conjuntos
Que es un conjunto
En matemáticas, un **conjunto** es una colección de objetos considerada como un objeto en sí. Los objetos de la colección pueden ser cualquier cosa: personas, números, colores, letras, figuras, etc. Cada uno de los objetos en la colección es un elemento o miembro del conjunto.
Un conjunto suele definirse mediante una propiedad que todos sus elementos poseen.

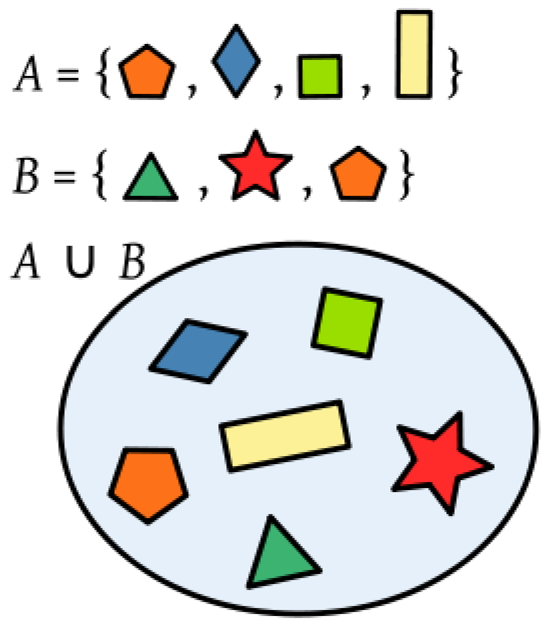
Un conjunto queda definido únicamente por sus miembros y por nada más. En particular el orden en el que se representen estos es irrelevantes. Además, cada elemento puede aparecer de manera idéntica una sola vez, esto es, no puede haber elementos totalmente idénticos repetidos.
Los conjuntos pueden ser finitos o infinitos. El conjunto de los números naturales es infinito, pero el conjunto de los planetas en el Sistema Solar es finito (tiene ocho elementos). Además, con los conjuntos pueden combinarse mediante operaciones, de manera similar a las operaciones con números.



<------ CONJUNTO DE PERSONAS

Por ejemplo, el conjunto de los números naturales es la unión del conjunto de los números pares positivos *P* y el conjunto de los números impares positivos *I*:
P = {2, 4, 6,...}
I = {1, 3, 5,...}
N = {1, 2, 3, 4, 5, 6,...}

La unión de conjuntos se denota por el símbolo ∪, de modo que por ejemplo, N = P U I.



A: PERTENECE A UN CONJUNTO
B: PERTENECE A OTRO CONJUNTO

EL CIRCULO DONDE APARECEN TODAS LAS FIGURAS JUNTAS DEL CONJUNTO A Y B SE LE LLAMA "UNIÓN"
Por ejemplo, dado el conjunto de los números pares *P* y el conjunto de los cuadrados *C* de números naturales, su intersección es el conjunto de los cuadrados pares *D*:
P = {2, 4, 6, 8, 10,...}
C = {1, 4, 9, 16, 25,...}
D = {4, 16, 36, 64,...}
La intersección de conjuntos se denota por el símbolo ∩ por lo que *D* = *P* ∩ *C*.

A: PERTENECE A UN CONJUNTO
B: PERTENECE A OTRO CONJUNTO

En el círculo se muestran las figuras que al unir los conjuntos A y B coinciden 2 figuras, las cuales se repiten en los 2 conjuntos a eso le llamamos intersección

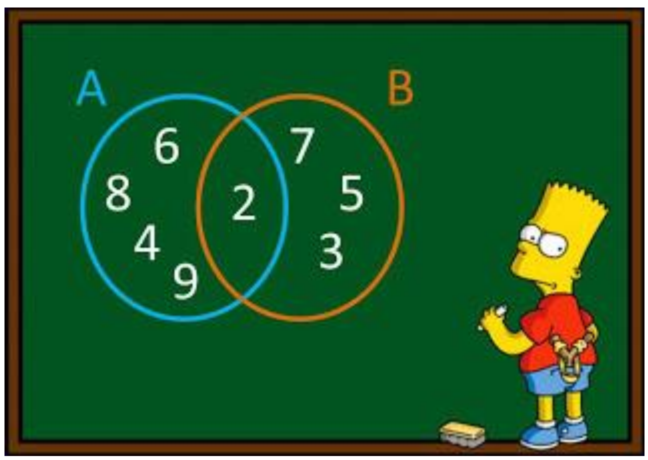
A continuación, les voy a mostrar una explicación más breve

En la imagen:
Nos muestra 2 círculos. Tomemos en cuenta solo el circulo azul que es el A.
A= (8, 6, **2**, 4, 9)   ese círculo y los números que están dentro de él, son el conjunto.
Ahora tomemos solo en cuenta el círculo anaranjado que es el B.
B= (**2**, 7, 5, 3) ese círculo y los números que están dentro de él, también son un conjunto.

Ahora si esos 2 círculos A y B los juntáramos, formaríamos una unión la cual estaría constituida por:

A y B = (8, 6, **2**, 4, 9, **2**, 7, 5, 3) juntando estos 2 círculos estaríamos uniendo todos los números.

Ahora si se dan cuenta al hacer una unión de A Y B notamos que 1 número se repite, el cual es el número **2**. Ese número (**2**) es la intersección de la unión de dos conjuntos.



Algo más fácil si no le entendiste:
Tenemos a una joven y un joven, cada uno es un conjunto, por decirlo así, si ellos 2 se casan, tenemos ahí una unión, ahora bien, si los 2 deciden tener un bebe, ese sería la intersección, ya que viene de esa unión de los jóvenes que ellos dos solos eran un conjunto.

Para conjuntos con intersección: P (A U B) = P (A) + P (B) – P (∩B).
Esto se debe a que sumamos la probabilidad de B, pero como ya habíamos sumado la intersección, entonces la restamos.

<u>Para conjuntos con mutuamente excluyentes</u>**:**
P(A U B) = P (A) + P (B)
En este caso, no hay ningún problema en sumar ambas probabilidades.
Ejemplo: Se lanza un dado, usted gana $3,000 pesos, si el resultado es par o divisible por tres. ¿Cuál es la probabilidad de ganar?

Lo primero que hacemos es definir los sucesos:
Sea A = resultado par: A = {2, 4, 6}.
Sea B = resultado divisible por 3: B = {3, 6}
Ambos sucesos tienen intersección?
$A \cap B = 3$ luego,
$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$$P(A \cup B) = \frac{3}{6} + \frac{2}{6} + \frac{1}{6} = \frac{4}{6} = \frac{2}{3}$$

2.1.2 TÉCNICAS DE CONTEO

Son aquellas técnicas que son usadas para enumerar eventos difíciles de cuantificar. Las técnicas de conteo permiten contar el número total de elementos que tiene un conjunto sin necesidad de hacerlo elemento por elemento, entre ellas tenemos, permutaciones, combinaciones y diagrama de árbol, hay que destacar que estas nos proporcionan la información de todas las maneras posibles en que ocurre un evento determinado, estas participaciones que se conocen en probabilidad como análisis combinatorio**.**

El principio fundamental en el proceso de contar ofrece un método general para contar el número de posibles arreglos de objetos dentro de un solo conjunto o entre varios conjuntos.

Las bases para entender el uso de las técnicas de conteo son el principio de multiplicativo y el principio aditivo, los que a continuación se definen y se hace uso de ellos.

Si un evento A puede ocurrir de n1 maneras y una vez que este ha ocurrido, otro evento B puede n2 maneras diferentes entonces, el número total de formas diferentes en que ambos eventos pueden ocurrir en el orden indicado, es igual a n1 x n2.

¿De cuántas maneras pueden repartirse 3 premios a un conjunto de 10 personas, suponiendo que cada persona no puede obtener más de un premio? Aplicando el principio fundamental del conteo, tenemos 10 personas que pueden recibir el primer premio. Una vez que éste ha sido entregado, restan 9 personas para recibir el segundo, y posteriormente quedarán 8 personas para el tercer premio. De ahí que el número de maneras distintas de repartir los tres premios. 10 x 9 x 8 = 720.

Principio fundamental del conteo**.**

Si un evento puede realizarse de n_1, maneras diferentes y si continuo con el procedimiento n_2, maneras diferentes y si después de efectuados estos n_3, otro procedimiento de maneras diferentes y así sucesivamente, entonces el número total (n_t), de formas o maneras en los que los eventos pueden realizarse en el orden indicado, es el producto de: $n_1 \times n_2 \times n_3 \times \ldots\ldots \times n_r = n_t$.

2.1.3 REGLAS DE ADICIÓN

La regla de la adición o regla de la suma establece que la probabilidad de ocurrencia de cualquier evento en particular es igual a la suma de las probabilidades individuales, si es que los eventos son mutuamente excluyentes, es decir, que dos no pueden ocurrir al mismo tiempo.

P(A y B)=P(A+B)= (B) = P(A)+P (B), SI A y B, son mutuamente excluyente

P(A o B)=P(A) + P (B) – P(A y B), si A y B, son no excluyentes.

Siendo: P(A)=probabilidad de ocurrencia del evento AC: P (B)=probabilidad de ocurrencia del evento B. P(A Y B)= probabilidad de ocurrencia simultanea de los eventos A y B.

Esta regla de la adición expresa la probabilidad de que ocurran dos o más sucesos a la vez, P (A U B).

Puede presentarse en dos formas: Para conjuntos con intersección y para conjuntos mutuamente excluyentes.

Establece que si dos eventos A y B son mutuamente excluyentes la probabilidad de que uno u otro evento ocurra es igual a la suma de sus probabilidades. De lo anterior se puede deducir que la probabilidad de que ocurra A más la probabilidad de que no ocurra A debe sumar 1. A esto se le llama la regla del complemento. Esta regla establece que para determinar la probabilidad de que ocurra un evento se puede restar de 1 la probabilidad de que no ocurra.

Ejemplos:

1. Una persona desea comprar una lavadora de ropa para lo cual ha pensado que puede seleccionar de entre las marcas whirpool, Easy y General electric, cuando acude hacer la compra se encuentra que la lavadora de la marca W se presenta en dos tipos de carga (8 y 11) kg, en tres tipos de carga, en cuatro colores diferentes y puede ser automática o semiautomática, mientras que la lavadora de la marca E se presenta en tres tipos de carga (8, 11 y 15) kg, en dos colores diferentes y puede ser automática o semiautomática y la lavadora de la marca GE se presenta en un solo tipo de carga, que es de 11kg, dos colores diferentes y solo hay semiautomática. ¿Cuántas maneras tiene esta persona de comprar una lavadora?

Solución:

M = Numero de maneras de seleccionar una lavadora marca Whirpool
N = Numero de maneras de seleccionar una lavadora marca Easy
W = Numero de maneras de seleccionar una lavadora marca General E.

M = 2x4x2 = 16 maneras
N = 3x2x2 = 12 maneras
W= 1x2x1 = 2 maneras
M+N+W = 16+12+2 = 30 maneras de seleccionar una lavadora

2. Una pareja que se tiene que casar, junta dinero para el enganche de su casa, en el fraccionamiento Lomas de la Presa, le ofrece un modelo económico o un condominio en el fraccionamiento Playas, le ofrecen un modelo económico como modelos, un residencial, un californiano y un provenzal. ¿Cuántas alternativas diferentes de vivienda le ofrecen a la pareja?

Solución:

Presa	Playas	
Económico	Residencial	M = 2
Condominios	California	N = 3
	Provenzal	M+N = 5

Principio de adición

Este principio establece que, si dos eventos m y n no pueden ocurrir al mismo tiempo, el número de formas como puede ocurrir el primer o segundo evento será la suma de m + n:

Número de formas = m + n… + x formas diferentes.

Ejemplo:

Antonio quiere hacer un viaje, pero no decide a cuál destino; en la Agencia de Turismo del Sur le ofrecen una promoción para viajar a New York o Las Vegas, mientras que la Agencia de Turismo del Este le recomienda viajar a Francia, Italia o España. ¿Cuántas alternativas de viajes diferentes le ofrecen Antonio?

Solución

Con la Agencia de Turismo del Sur Antonio tiene 2 alternativas (New York o Las Vegas), mientras que con la Agencia de Turismo del Este tiene 3 opciones (Francia, Italia o España). El número de alternativas diferentes es:

Número de alternativas= m + n = 2 + 3 = 5 alternativas

Principio fundamental del conteo**.**

Si un evento puede realizarse de n_1, maneras diferentes y si continuo con el procedimiento n_2, maneras diferentes y si después de efectuados estos n_3, otro procedimiento de maneras diferentes y así sucesivamente, entonces el número total (n_t), de formas o maneras en los que los eventos pueden realizarse en el orden indicado, es el producto de: $n_1 \times n_2 \times n_3 \times \ldots\ldots\times n_r = n_t$.

2.1.4 REGLA DE MULTIPLICACIÓN

La regla de la multiplicación establece que la probabilidad de ocurrencia de dos o más eventos estadísticamente independientes es igual al producto de sus probabilidades individuales.

P(A y B) = P (AB) = P(A) P (B), si A y B, son independientes.
P(A y B) = P (AB) = P(A) P (B/A) =, si A y B, son dependientes.

La regla de la multiplicación se refiere a la determinación de la probabilidad de la ocurrencia conjunta de A y b.

Existen dos opciones de esta regla:

Si los eventos son independientes:

P(A y B) = P(A o B) = P (A) P (B)

Si los eventos son dependientes:

Es la probabilidad de A multiplicada por la probabilidad condicional de B dado A.

P(A Y B) = P (A) P (B/A)

Si la posición de los dos eventos se invierte, se obtiene un valor equivalente.

P(A y B) = P (B/A) = P (B) P(A/B)

Ejemplo:

Si una moneda equilibrada se lanza dos veces, la probabilidad de que ambos lanzamientos den por resultado una "cara" es:

(1/2) x (1/2) = 1/4

El principio multiplicativo**.** Es una técnica que se utiliza para resolver problemas de conteo para hallar la solución sin que sea necesario enumerar sus elementos. Es conocido también como el principio fundamental del análisis combinatorio; se basa en la multiplicación sucesiva para determinar la forma en la que puede ocurrir un evento.

Este principio establece que, si una decisión (d_1) puede ser tomada de n maneras y otra decisión (d_2) puede tomarse de m maneras, el número total de maneras en las que pueden ser tomadas las decisiones d_1 y d_2 será igual a multiplicar de n * m. Según el principio, cada decisión se realiza una tras otra: número de maneras = $N_1 * N_2 \ldots * N_x$ maneras o formas.

El principio multiplicativo implica que cada uno de los pasos de la actividad deben ser llevados a efecto, uno tras otro. Si un evento E_1 puede suceder de n_1 maneras diferentes, el evento E_2 puede ocurrir de n_2 maneras diferentes y así sucesivamente hasta el evento E_p el cual puede ocurrir de n_p maneras diferentes. Entonces el total de maneras distintas en que puede suceder el evento "ocurren E_1 y E_2 es igual al producto".

Ejemplos:

1. Lanzamiento de dos dados. Como cada dado tiene 6 posibles resultados, el número total de posibles resultados es **6x6=36** Por supuesto, la regla de multiplicación puede extenderse a experimentos con más de dos partes.

2. Lanzamiento de 6 monedas. Como cada parte del experimento tiene 2 posibilidades (cara o cruz) tenemos entonces que el número total de posibles resultados es:

$$2x2x2x2x2x2 = 64$$

3. Una persona desea construir su casa, para lo cual considera que puede construir los cimientos de su casa de cualquiera de las dos maneras (concreto o block de cemento), mientras que en las paredes las puede hacer de adobe (adobe, adobón o ladrillo), el techo puede ser de concreto o lámina galvanizada) y por último los acabados los puede realizar de una sola manera, ¿cuantas maneras tiene esta persona de construir su casa?

Solución:

Considerando que r = 4 pasos
N_1 = maneras de hacer cimientos = 2
N_2 = maneras de construir paredes = 3
N_3 = maneras de hacer techos = 2
N_4 = maneras de hacer acabados = 1

N1 x N2 x N3 x N4 **= 2 x3 x x2 x 1 = 12 maneras de construir la casa.**

4. ¿Cuántas placas para automóvil pueden ser diseñadas si deben constar de tres letras seguidas de cuatro números, si las letras deben ser tomadas del abecedario y los números entre los dígitos del 0 al 9?
a). Si es posible repetir letras y números.
b). No es posible repetir letras y números.

Solución:

a). considerando 26 letras del abecedario y los dígitos del 0 al 9
AAA 1111
26x26x26x10x10x10x10 = 175*760,000 placas para automóvil
b). 26x25x24x10x9x8x7 = 78*624,000 placas para automóvil

5. Calcular cuántos números enteros diferentes de tres dígitos se pueden formar con los dígitos 2, 3, 4, 5, 6, 7 y 8, si los dígitos no pueden repetirse.

Solución:

7x6x5 = 210 números enteros

6. Calcular de cuantas maneras diferentes se pueden sentar tres niños en una banca de tres asientos.

Solución:

3x2x1 = 6 maneras diferentes

7. ¿de cuantas maneras pueden repetirse 3 premios a un conjunto de 10 personas, suponiendo que cada persona no puede obtener más de un premio?

Solución: Aplicando el principio fundamental del conteo, tenemos 10 personas que pueden recibir el 1er. Premio. Una vez que este ha sido entregado, restan 9 personas para recibir el 2do. Premio y posteriormente quedaran 8 personas para el 3er. premio. De ahí que el número de maneras distintas de repetir10x9x8 = 720

8. ¿Cuántas placas de automóvil se pueden hacer utilizando dos letras seguidas de tres cifras? no se admiten repeticiones.

Solución:

$$26x25x10x9x8 = 468,000$$

2.1.5 DIAGRAMA DE ÁRBOL

Este documento describe el proceso de construcción de un diagrama de árbol, mediante el cual se dispone de una metodología simple y sistemática para la identificación de las acciones que contribuyen a la consecución de un objetivo.
La muestra por medio de ejemplos, como la construcción de estos diagramas, es capaz de ofrecer una visión sencilla y concentrada del análisis de situaciones complejas.

Un diagrama de árbol es una herramienta que se utiliza para determinar todos los posibles resultados de un experimento aleatorio. En el cálculo de la probabilidad se requiere conocer el número de objetos que forman parte del espacio muestral, estos se pueden determinar con la construcción de un diagrama de árbol.

El diagrama de árbol es una representación gráfica de los posibles resultados del experimento, el cual consta una serie de pasos, donde cada uno de los pasos tiene un número finito de maneras de ser llevado a cabo. Se utiliza en los problemas de conteo y probabilidad.

Para la construcción de un diagrama de árbol se partirá poniendo una rama para cada una de las posibilidades, acompañada de su probabilidad. Cada una de esta rama se conoce como rama de primera generación.

En el final de cada rama de primera generación se construye a su vez, un nudo del cual parten nuevas ramas conocidas como ramas de segunda generación, según las posibilidades del siguiente paso, salvo si el nudo representa un posible final del experimento (nudo final).

Hay que tener en cuenta que la construcción de un diagrama de árbol no depende de tener el mismo número de ramas de segunda generación que salen de cada rama de primera generación y que la suma de probabilidades de las ramas de cada nudo ha de dar 1.

El diagrama de árbol se construye generalmente de izquierda a derecha. El número de las ramas en cada punto es el número de resultados posibles del experimento.

En un diagrama de árbol las líneas que unen los puntos de diagrama se denominan aristas y esos puntos nudos. Además de tener una raíz que es el nudo a donde no llega ninguna arista, un árbol tiene la propiedad de que ningún camino que parta de la raíz puede visitar dos veces el mismo nudo.

El Diagrama de Árbol es una herramienta que se utiliza para determinar todos los posibles resultados de un experimento aleatorio. En el cálculo de la probabilidad se con la construcción del Diagrama de Árbol. Requiere conocer el número de elementos que forman parte del espacio muestral, estos se pueden determinar

El Diagrama de Árbol es una representación gráfica de los posibles resultados del experimento, se construye horizontalmente y se colocan ramas que indican las alternativas del experimento.

La figura muestra unas ramas secas de un árbol, dando similitud a las ramificaciones que tiene el Diagrama que se utiliza en probabilidad. Cada rama representa un posible resultado de universo.

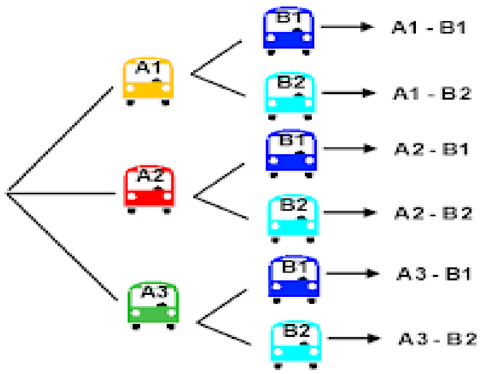
Un Diagrama de Árbol comienza en un punto de unión de donde parte dos o más ramas que representan cada una de las posibles combinaciones o resultados del experimento.

Recuerda que la probabilidad es el resultado del cociente entre “El número de elementos del evento” ENTRE “el número de elementos del espacio muestral”.

Ambos números de elementos se obtienen del Diagrama de Árbol.

En términos generales, el diagrama de árbol es una herramienta que permite analizar y evaluar cada una de las posibilidades que pueden existir entorno a la resolución de algún problema o los resultados posibles de algún experimento y las probabilidades de que cada posibilidad ocurra. Una cosa es conocer las posibilidades y otra es graficarlas o enumerarlas, de manera tal que nada se escape y que todo caso “posible” sea tomado en cuenta.

Para la planificación dentro de las organizaciones, el diagrama de árbol es una herramienta muy útil, ya que permite evaluar el camino a la meta, analizando las diferentes maneras de alcanzar el objetivo. Por otro lado, en una tormenta de ideas, por ejemplo, tiene como finalidad identificar aquellos elementos que pudiesen ser olvidados.



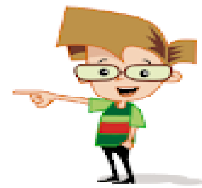
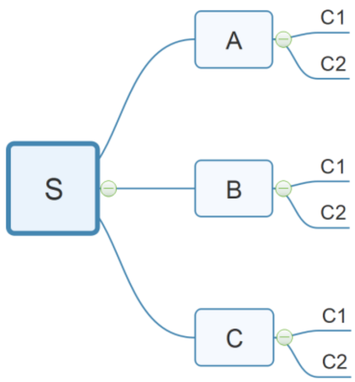


Diagrama de árbol

Ejemplos:

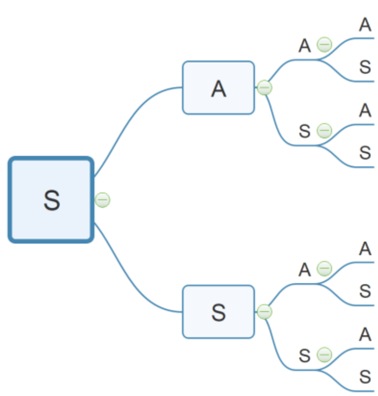
1. Si Juan tiene 3 pantalones y 2 camisas basta multiplicar 3x2=6 y son 6 posibilidades de que se pueda vestir. Resolverlo por el diagrama de árbol y el espacio muestral.



Espacio muestral S= {(A1) (A2) (B1) B2) (C1) (C2)}

Combinaciones 6 3X2 = 6

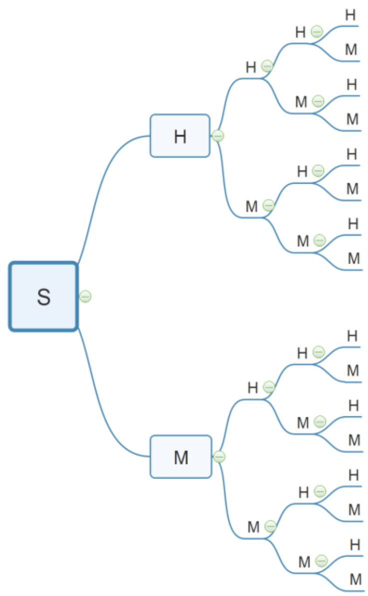
2. Encontrar las combinaciones del lanzamiento de tres monedas, por medio del diagrama del árbol y encontrar el espacio muestral.



Espacio muestral S= {(AAA)(AAS)(ASA)(ASS); (SAA)(SAS)(SSA)(SSS)}

Combinaciones 8 2X4 = 8

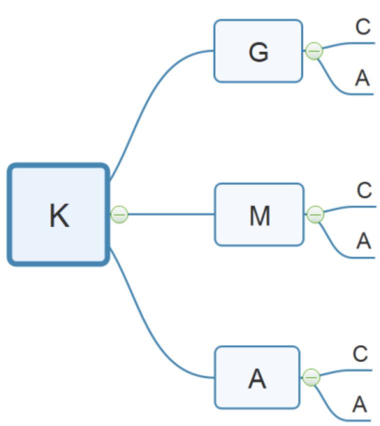
3. Una familia con cuatro hijos. Primero se tienen que determinar los resultados posibles que se puedan presentar en una familia. Para ello podemos utilizar un diagrama de árbol. Encontrar las posibles combinaciones y el espacio muestral.



Espacio muestral S= {(hhhh)(hhhm)(hhmh)(hhmm); (hmhhh)(hmhm)hmmh)(hmmm); (mhhh)(mhhm)(mhmh)(mhmm); (mmhh)(mmhm)(mmmh)(mmmm)}

Combinaciones 16 2x8 = 16

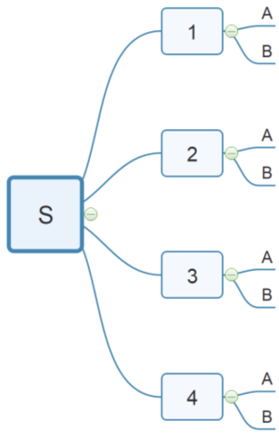
4. Elabora un diagrama de árbol del siguiente problema: Karen tiene tres lugares a donde ir de vacaciones, Guaymas, Mazatlán y Acapulco, ya sea en camión o en avión. Encontrar el espacio muestral y el total de las combinaciones.



Espacio muestral K = {(Gc)(Ga); (Mc)(Ma); (Ac)(Aa)}

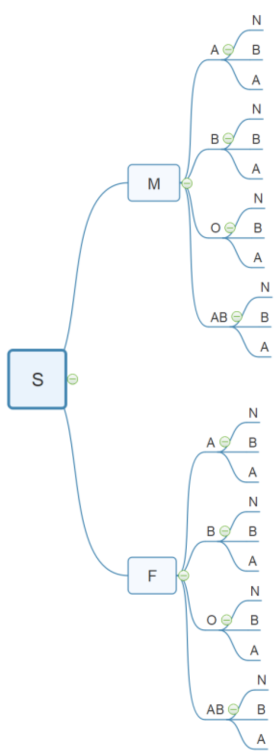
Combinaciones 6 3x2 = 6

5. Determinar cuántos números de dos cifras distintos se pueden formar con los números enteros 1, 2, 3 y 4, usando el diagrama de árbol. Encontrar el total de las combinaciones.



Espacio muestral S = {(1a)(1b); (2a)(2b); (3a)(3b); (4a)(4b)}
Combinaciones 8 4x2 = 8

6. Un médico general clasifica a sus pacientes de acuerdo a: su sexo (masculino o femenino), con los tipos de sangre (A, B, O, AB), y en cuanto a la presión sanguínea (normal, alta, baja), mediante un diagrama de árbol. Encuentre en cuantas clasificaciones pueden estar los pacientes de este médico?



Espacio muestral S= (man)(mab)(maa);(mbn)(mbb)(mba);(mon)(mob)(moa); (m,ab,n)(m,ab,b)(m,ab,a); (fan)(fab)(faa); (fbn)(fbb)(fba); (fon)(fob)(fba); (f,ab,n)(f,ab,b)(f,ab,a).

Combinaciones 24 2x4x3 = 24

TAREA:

1. Dos equipos denominados A y B, jugaran la final, aquel equipo que gane de un partido de baloncesto dos juegos seguidos o complete un total de tres juegos ganados será el que gane el torneo, mediante un diagrama de árbol. Encontrar de cuantas maneras puede ser ganado este torneo y resolver el espacio muestral.

2. Una universidad tiene de tres facultades:

- La primera con el 50% de estudiantes.
- La segunda con el 25% de estudiantes.
- La tercera con el 25% de estudiantes.

Las mujeres están repartidas uniformemente, siendo un 60% del total en cada facultad.

2.1.6 ANALISIS COMBINATORIO

Es la rama de la matemática que estudia los diversos arreglos o selecciones que podemos formar con los elementos de un conjunto dado, los cuales nos permite resolver muchos problemas prácticos. Por ejemplo, podemos averiguar cuántos números diferentes de teléfonos, placas o loterías se pueden formar utilizando un conjunto dado de letras y dígitos.

Además el estudio y comprensión del análisis combinatorio no va a servir de almacenaje para poder resolver y comprender problemas sobre probabilidades Principios fundamentales del Análisis Combinatorio: En la mayoría de los problemas de análisis combinatorio se observa que una operación o actividad aparece en forma repetitiva y es necesario conocer las formas o maneras que se puede realizar dicha operación.

Para dichos casos es útil conocer determinadas técnicas o estrategias de conteo que facilitarán el cálculo señalado.

El análisis combinatorio también se define como una manera práctica y abreviada de contar; las operaciones o actividades que se presentan son designadas como eventos o sucesos.

1) ¿Cuántos números de 2 cifras diferentes pueden formarse con los dígitos: 2 , 3 , 5 y 7?

Respuesta: 4 × 3 = 12

23 , 25 , 27 , 32 , 35 , 37 , 52 , 53 , 57 , 72 , 73 y 75

2) ¿De cuántas formas diferentes se pueden sentar 3 damas en 5 sillas linealmente dispuestas?

Respuesta: 5 × 4 × 3 = 60

3) Una empresa ferroviaria tiene 6 estaciones. ¿Cuántos tipos diferentes de boletos, donde se indique la estación de salida y de llegada, deben imprimirse?

Respuesta: 6 × 5 = 30

calcular las maneras diferentes de vestir de una persona, utilizando un número determinado de prendas de vestir.

2. Ordenar 5 artículos en 7 casilleros. pendiente

3. Designar 5 personas de un total 50 para integrar una comisión.

4. Escribir una palabra de 7 letras utilizando 4 consonantes y 3 vocales. Si se lanzan simultáneamente un dado con 6 caras numeradas del 1 al 6 y una moneda de cuantas maneras puede caer.

Suceso A (Dado): 6

Suceso B (Moneda): 2

6X2=12

Además el estudio y comprensión del análisis combinatorio no va a servir de almacenaje para poder resolver y comprender problemas sobre probabilidades Principios fundamentales del Análisis Combinatorio: En la mayoría de los problemas de análisis combinatorio se observa que una operación o actividad aparece en forma repetitiva y es necesario conocer las formas o maneras que se puede realizar dicha operación. Para dichos casos es útil conocer determinadas técnicas o estrategias de conteo que facilitarán el cálculo señalado.

El análisis combinatorio también se define como una manera práctica y abreviada de contar; las operaciones o actividades que se presentan son designadas como eventos o sucesos.

Ejemplo:
1. Señalar las maneras diferentes de vestir de una persona, utilizando un número determinado de prendas de vestir
2. Ordenar 5 artículos en 7 casilleros
3. Designar 5 personas de un total 50 para integrar una comisión
4. Escribir una palabra de 7 letras utilizando 4 consonantes y 3 vocales. l.

Si se lanzan simultáneamente un dado con 6 caras numeradas del 1 al 6 y una moneda de cuantas maneras puede caer.
Suceso A (Dado): 6
Suceso B(Moneda): 2
6X2=12

2.2 PERMUTACIONES Y COMBINACIONES

Para entender lo que son las permutaciones, es necesario definir lo que es una permutación y lo que es una combinación, para establecer su diferencia y de esta manera entender claramente, cuando es posible utilizar una permutación y cuando utilizar una combinación al momento de querer cuantificar los elementos de algún evento.

En la permutación a diferencia de la fórmula de la multiplicación se utiliza para determinar el número de posibles arreglos cuando solo hay un solo grupo de objetos.

Una permutación es un arreglo de todo o parte de un conjunto de objetos. Una permutación es una combinación en donde el orden es importante.

Permutación. Son arreglos o posición de “r” objetos seleccionados de un solo grupo de “n” objetos posibles.

n! = Es el producto donde la unidad hasta el valor que ostenta “n”.

La función factorial (símbolo!). Significa que se multiplican números descendentes.

Ejemplos:
4! = 4x3x2x1= 24
7! = 7x6x5x4x3x2x1 = 5,040
9! = 9x8x7x6x5x4x3x2x1 = 362,880

En matemáticas una permutación es la variación del orden o de la disposición de los elementos de un conjunto.

En matemáticas usamos un lenguaje más preciso:

- Si el orden no importa, es una combinación.

- Si el orden sí importa, es una permutación.
- Una permutación es una combinación ordenada.

Ejemplos:

1. ¿Cuántas permutaciones existen para las ocho letras a, b, c, d, e, f, g, h?
   $P_8 = 8! = 8x7x6x5x4x3x2x1 = 40,320$

2. ¿Cuántas de las permutaciones de (a), comienzan con la letra a?
   ab, ac, ad, ae, af, ag, ah.

   $P_7 = 7! = 7x6x5x4x3x2x1 = 5,040$

3. ¿De cuantas formas diferentes se pueden ordenar las letras de la palabra PERMUTACION?
   $P_{10} = 10! = 10x9x8x7x6x5x4x3x2x1 = 3'628,800$

4. ¿De cuantas formas diferentes se pueden ordenar las letras de la palabra COMBINACIONES?
   $P_{13} = 13! = 13x12x11x10x9x8x7x6x5x4x3x2x1 = 6''227'020,800$

5. ¿En el conjunto {1, 2, 3}, en cada ordenación posible de sus elementos, sin repetirlos, es una permutación?
   P (1, 2, 3) = 3! = 3x2x1 = 6

6. Si n = 5! = 5x4x3x2x1 = 120

7. Si n = 9! = 9x8x7x6x5x4x3x2x1 = 362,880

8. ¿Cuántos números de 5 cifras diferentes se puede formar con los dígitos 1, 2, 3, 4, 5, si entran todos los elementos, si importa el orden, no se repiten los elementos. El enunciado nos pide que las cifras sean diferentes?
   $P_5 = 5! = 5x4x3x2x1 = 120$

9. ¿calcular las permutaciones de 6 elementos?
   $P_6 = 6x5x4x3x2x1 = 720$

10. ¿De cuantas formas distintas pueden sentarse doce personas en una fila de butacas? Si entran todos los elementos que pueden sentarse las doce personas; si importa el orden, no se repiten los elementos, una persona no se puede repetir.
    $P_{12} = 12! = 12x11x10x9x8x7x6x5x4x3x2x1 = 479'001,600$

<u>Principio de Permutación</u>

Se trata de ordenar específicamente todos o algunos de los elementos que forman un conjunto, para facilitar el conteo de todos los posibles arreglos que pueden hacerse con los elementos.

El número de permutaciones de n elementos diferentes, tomados todos a la vez, se representa como:

$$\mathbf{{}_nP_n = n!}$$

Ejemplos:

1. Cuatro amigos quieren tomarse una fotografía y desean saber de cuántas formas diferentes pueden ordenarse.

Solución

Se quiere conocer el conjunto de todas las formas posibles en que las 4 personas se pueden colocar para tomarse la fotografía. Así, se tiene que:

$$\mathbf{{}_nP_n = n!}$$

${}_4P_4 = 4! = 4*3*2*1 = 24$ formas diferentes.

Si el número de permutaciones de n elementos disponibles es tomado por partes de un conjunto que está formado por r elementos, se representa como:

$$\mathbf{{}_nP_r = \frac{n!}{(n-r)!}}$$

Ejemplo

2. En una sala de aula se tienen 10 puestos. Si para la clase asisten 4 estudiantes, ¿de cuántas maneras distintas los estudiantes pueden ocupar los puestos?

Solución

Se tiene que el número total del conjunto de sillas es 10, y de estas solo serán usadas 4. Se aplica la fórmula dada para determinar el número de permutaciones:

$\mathbf{{}_nP_r = n! \div (n - r)!}$

${}_{10}P_4 = 10! \div (10 - 4)!$

${}_{10}P_4 = 10! \div 6!$

${}_{10}P_4 = 10* 9*8*7*6*5*4*3*2*1 \div 6*5*4*3*2*1 = 5040$ maneras de ocupar los puestos.

Existen casos en los que algunos de los elementos disponibles de un conjunto se repiten (son iguales). Para calcular el número de arreglos tomando todos los elementos a la vez se utiliza la siguiente fórmula:

$\mathbf{{}_nP_r = n! \div n_1!* n_2!\ldots n_r!}$

Ejemplo

¿Cuántas palabras diferentes de cuatro letras pueden ser formadas a partir de la palabra “lobo”?

Solución

En este caso se tienen 4 elementos (letras) de los cuales dos de ellos son exactamente iguales. Aplicando la fórmula dada, se sabe cuántas palabras diferentes resultan:

$\mathbf{{}_nP_r = n! \div n_1!* n_2!\ldots n_r!}$

${}_4P_{2,1,1} = 4! \div 2!*1!*1!$

${}_4P_{2,1,1} = (4*3*2*1) \div (2*1)*1*1$

${}_4P_{2,1,1} = 24 \div 2 = 12$ palabras diferentes.

Permutaciones de “n” Elementos Diferentes en Grupos de “r” Elementos

Toda ordenación de un conjunto de “n” elementos, se llama permutación del conjunto.

Una permutación de “n” objetos diferentes, tomados de “r” a la vez, es una ordenación de “r” objetos tomados de entre “n” objetos dados, con especial atención prestada al orden dela elección. El número de permutaciones de “n” objetos, tomando “r” a la vez, se denota por ${}_nP_r$, P(n, r) o $P_{n,r}$, está dada por

$$\mathbf{P(n,r) = \frac{n!}{(n-r)!}}$$

Donde:

P Es la permutación

n Es el número de elementos

r Es el número de elementos en grupos

n! n factorial

Para obtener el número de permutaciones de “n” objetos considerados en grupos de “r” objetos se hace r = n, de este modo se obtiene P(n, n) = n!

Por definición 0! = 1

Ejemplos:

1. Seis personas entran a un cuarto en el que hay diez sillas. ¿De cuantas maneras pueden ocupar las sillas? (haciendo uso del principio multiplicativo) y utilizando la fórmula de permutación.

Solución: puesto que únicamente se ocupan seis de las sillas, el número de diferentes maneras de ocupar las sillas es igual al número de permutaciones de diez objetos considerados en grupos de seis.

$$\mathbf{P(n, r) = \frac{n!}{(n-r)!}}$$

P_{10} = 10! = 10x9x8x7x6x5x4x3x2x1=151,200 maneras (Principio Multiplicativo)

$$P(10, 6) = \frac{10!}{(10-6)!} = \frac{10!}{4!} = \frac{10 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4!}{4!} = 10 \times 9 \times 8 \times 7 \times 6 \times 5 = 151{,}200$$

2. ¿Cuántas maneras hay de asignar los cuatro primeros lugares de un concurso de creatividad que se verifica en las instalaciones de nuestro instituto, si hay catorce participantes? (haciendo uso del principio multiplicativo) y utilizando la fórmula de permutación.

Solución:

Se debe de que al momento de asignar el primer lugar tenemos a 14 posibles candidatos, una vez asignado ese lugar nos quedan 13 posibles candidatos para el segundo lugar, luego tendríamos 12 candidatos posibles para el tercer lugar y por último tendríamos 11 candidatos posibles para el cuarto lugar.

$$P(n, r) = \frac{n!}{(n-r)!}$$

P_{14}=14!=14x13x12x11=24,024 maneras (Principio Multiplicativo)

$P(14, 4) = \frac{14!}{(14-4)!} = \frac{14!}{10!} = \frac{14x13x12x11x10!}{10!} = 14x13x12x11! = 24{,}024$

3. ¿Cuántas palabras diferentes de cuatro letras se pueden formar con las siguientes letras A, B, C, D, E, F, G, H, I, J, K, L, M? (haciendo uso del principio multiplicativo) y utilizando la fórmula de permutación.

Solución:

$$P(n, r) = \frac{n!}{(n-r)!}$$

P_{13} = 13! = 13x12x11x10 = 17,160 diferentes (Principio Multiplicativo)

$P(13,4) = \frac{13!}{(13-4)!} = \frac{13!}{9!} = \frac{13x12x11x10x9!}{9!} = 13x12x11x10 = 17{,}160$

4. ¿Cuántas representaciones diferentes serán posibles formar, si se desea que consten de presidente, secretario, tesorero, primer vocal y segundo vocal? Si esta representación puede ser formada entre 25 miembros del sindicato de una pequeña empresa. (haciendo uso del principio multiplicativo) y utilizando la fórmula de permutación.

Solución:

$$P(n, r) = \frac{n!}{(n-r)!}$$

P_{25} =25! = 25x24x23x22x21! = 6´375,600 maneras (Principio Multiplicativo)

$P(25,5) = \frac{25!}{(25-5)!} = \frac{25x24x23x22x21x20!}{20!} = 25x24x23x22x21! =$ 6´375,600 maneras

5. a) ¿Cuántas maneras diferentes hay que asignar las posiciones de salida de ocho autos que participan en una carrera de fórmula uno? (considere que las posiciones de salida de los autos participantes en la carrera son dadas totalmente al azar). (haciendo uso del principio multiplicativo) y utilizando la fórmula de permutación.
6. b) ¿Cuántas maneras diferentes hay de asignar los primeros tres premios de esta carrera de fórmula uno? (haciendo uso del principio multiplicativo) y utilizando la fórmula de permutación.

Solución:

$$P(n, r) = \frac{n!}{(n-r)!}$$

a). P_8 = 8! = 8x7x6x5x4x3x2x1= 40,320 maneras (Principio Multiplicativo)

$P(8,8) = \frac{8!}{(8-8)!} = \frac{8x7x6x5x4x3x2x1}{0! = 1} = 8x7x6x5x4x3x2! =$ 40,320 maneras

b). $P(n, r) = \frac{n!}{}$

$$(n-r)!$$

$P_8 = 8! = 8x7x6 = 336$ maneras (Principio Multiplicativo)

$P(8,3) = \dfrac{8!}{(8-3)!} = \dfrac{8x7x6x5!}{5!} = 8x7x6 = 336$ maneras de asignar las posiciones

TAREA:

1. ¿Cuántos puntos de tres coordenadas (x, y, z), será posibles generar con los dígitos 0, 1, 2, 4, 6, 9?
Si. a). No es posible repetir dígitos.
b). si es posible repetir dígitos.
2. a) ¿Cuántas maneras hay de asignar las cinco posiciones de un juego de un equipo de Basquetbol, si el equipo consta de doce integrantes?
b) ¿Cuantas maneras hay de asignar las posiciones de juego si una de ellas solo puede ser ocupada por José?
c) ¿Cuántas maneras hay de que se ocupen las posiciones de juego si es necesario que una de ellas este José y otra Antonio? (haciendo uso del principio multiplicativo) y utilizando la fórmula de permutación en los tres casos.
3. ¿Cuantas palabras de cinco se pueden formar con las letras de MUNDIAL?
4. ¿Encontrar el número total de permutaciones de los números 2, 4, 1, 6, 8, 3, 5, tomadas todas a la vez.
5. a) ¿Cuántas maneras diferentes hay de asignar las posiciones de salida de nueve aviones que participan en una carrera? (considere la salida de los aviones, son totalmente al azar).
b) ¿Cuántas maneras diferentes hay de asignar los cuatro primeros lugares de esta carrera? (haciendo uso del principio multiplicativo) y utilizando la fórmula de permutación en los tres casos.

Permutaciones con Repetición

Permutaciones con repetición: Es el número total de permutaciones que es posible obtener con "n" objetos, entre los que hay una cantidad "x_1" de objetos de cierto tipo, una cantidad "x_2" de objetos de un segundo tipo, una cantidad "x_3" de objetos de cierto tipo, así sucesivamente y una cantidad de "x_n" de objetos de tipo "n".

Son los distintos grupos que pueden formarse con esos "n" elementos de forma que:

1. Si están todos los elementos.
2. Si importa el orden.
3. Si se repiten los elementos.

$$nP_{x_1, x_2, x_3 \ldots, x_n} = \frac{n!}{x_1, x_2, x_3 \ldots, x_n}.$$

Donde:

P es la permutación

n es el tamaño de la muestra o número de elementos que es posible obtener con “n” objetos.

x_i es la i-esima observación de la variable “x” de la muestra i=1,2,3,…,n

n! es “n” factorial

Ejemplos:

1. Obtenga todas las señales posibles que se puedan diseñar con seis banderines, dos de los cuales son rojos, tres son verdes y uno morado.

Solución:

n = 6 banderines

x_1 = 2 banderines rojos

x_2 = 3 banderines verdes

x_3 = 1 banderín morado

$$nP_{x_1, x_2, x_3} = \frac{n!}{x_{1}!\ x_{2}!\ x_{3}!\ldots\ldots x_n}$$

$${}_6P_{2,3,1} = \frac{6!}{2!\ 3!\ 1!} = \frac{6x\overset{3}{5}x\overset{5}{4}x\overset{4}{3!}}{2x1!\ 3!\ 1!} = 3x5x4 = 60 \text{ señales diferentes}$$

2. ¿Cuántas claves de acceso a una computadora será posible diseñar con los números 1, 1, 1, 2, 3, 3, 3, 3,?

Solución:

n = 8

x_1 = 3 números uno

x_2 = 1 numero dos

x_3 = 4 números tres

$$nP_{x_1, x_2, x_3} = \frac{n!}{x_{1}!\ x_{2}!\ x_{3}!\ldots\ldots x_n}$$

$${}_8P_{3,1,4} = \frac{8!}{3!\ 1!\ 4!} = \frac{8x7x6x5x4!}{3x2x1!\ 1!\ 4!} = \overset{4}{4}x\overset{7}{7}x\overset{2}{2}x\overset{5}{5} = 280 \text{ señales diferentes}$$

3. ¿De cuantas maneras es posible plantar en una línea divisoria de un terreno dos nogales, cuatro manzanos y tres ciruelos?

Solución:

n = 9 arboles

x_1 = 2 nogales

x_2 = 4 manzanos

x_3 = 3 ciruelos

$$nP_{x_1, x_2, x_3} = \frac{n!}{x_{1}!\ x_{2}!\ x_{3}!\ldots\ldots x_n}$$

$${}_9P_{2,\ 4,\ 3} = \frac{9!}{2!\ 4!\ 3!} = \frac{\overset{3}{9}x\overset{4}{8}x\overset{7}{7}x\overset{3}{6}x\overset{5}{5}x4!}{2x1!\ 4!\ 3x2x1!} = 3x4x7x3x1 = 1{,}260 \text{ maneras}$$

4. Si un equipo de futbol soccer femenil participa en 12 juegos en una temporada, ¿Cuántas maneras hay de que entre esos doce juegos en que participa, obtenga 7 victorias, 3 empates, y 2 juegos perdidos?

Solución:

n = 12 juegos
x_1 = 7 victorias
x_2 = 3 empates
x_3 = 2 juegos perdidos

$$\mathbf{{}_nP_{x_1, x_2, x_3} = \frac{n!}{x_{1!}\ x_{2!}\ x_{3!} \ldots\ldots x_n}}$$

4 11 5 9 4

$${}_{12}P_{7,3,2} = \frac{12!}{7!, 3!, 2!} = \frac{12x11x10x9x8x7!}{7!\ 3x2x1!\ \ 2x1!} = 4x11x5x9x4 = 7{,}920 \text{ maneras}$$

5. ¿Con las letras A, A, A, B, B, C ,C, C, C, C, ¿Cuántas letras de diez (10), se pueden formar?
   Si entran todos los elementos
   Si importa el orden
   Si se repiten los elementos

Solución:

n= 10
x_1 = 3 letras A
x_2 = 2 letras B
x_3 = 5 letras C

$$\mathbf{{}_nP_{x_1, x_2, x_3} = \frac{n!}{x_{1!}\ x_{2!}\ x_{3!} \ldots\ldots x_n}}$$

5 3 4 7 6

$${}_{10}P_{3,2,5} = \frac{10!}{3!\ 2!\ 5!} = \frac{10x9x8x7x6x5!}{3x2x1!\ \ 2x1!\ \ 5!} = 5x3x4x7x6 = 2{,}520$$

TAREA: Encontrar los valores de las permutaciones con repetición de los siguientes problemas:

1. ${}_{12}P_{6,4,2}$ =
2. ${}_{14}P_{3,5,8}$ =
3. ${}_{10}P_{4,7,2}$ =
4. ${}_{13}P_{4,1,6}$ =
5. ${}_{16}P_{5,8,10}$ =

<u>Combinaciones</u>

Es la técnica de conteo que permite calcular el número de arreglos que pueden realizarse con todos o con una parte de los elementos de un solo conjunto, en donde no interesa el orden de los elementos.

En matemáticas usamos el lenguaje más preciso:

- Si el orden no importa, es una combinación.

- Si el orden sí importa es una permutación.

Ejemplo:

"Mi ensalada de frutas de manzanas, uvas y bananas" no importa en qué orden pusimos las frutas, podría ser bananas, uvas y manzanas o uvas, manzanas y bananas, es la misma ensalada, es una combinación.

"La combinación de la cerradura" es 472: ahora sí importa el orden.
"724" no funcionaria, ni "247" tiene que ser exactamente "472" para que sea una permutación.

Definición de combinación**.** Es una palabra que refiere al acto y consecuencia de combinar algo o de combinarse (es decir, unir, complementar o ensamblar cosas diversas para lograr un compuesto).

El concepto posee múltiples aplicaciones ya que las cosas factibles de combinar son de características y orígenes muy diversos.

Una combinación es una selección de objetos sin importar el orden en que se escojan, (el orden no es importante).

Existe solo una combinación de "n" elementos diferentes, tomando "r" a la vez, es una selección de "n" objetos de los "n" objetos, sin prestar atención al orden de elección. El número de combinaciones de "n" objetos, tomando "r" a la vez, considerados en grupos de "r" Sin embargo, el problema para determinar el número de combinaciones de "n" objetos considerados en grupos de "r". Este número se designara por C(n, r).

$$\mathbf{C(n, r) = \frac{n!}{(n-r)!\, r!}}$$

Donde:
C Es la combinación
n Es el número de elementos
r Es el número de elementos en grupos
n! n factorial

Ejemplos:

1. ¿Cuántas combinaciones de cuatro letras distintas pueden formarse con 16 letras distintas?

Solución:

$$\mathbf{C\,(n, r) = \frac{n!}{(n-r)!\, r!}}$$

$$C(16,4) = \frac{16!}{(16-4)!\, 4!} = \frac{16}{12!\;\; 4!} = \frac{\overset{4}{16}\text{x}\overset{5}{15}\text{x}\overset{7}{14}\text{x}\overset{13}{13}\text{x}12!}{12!\;\; 4\text{x}3\text{x}2\text{x}1!} = 4\text{x}5\text{x}7\text{x}13 = 1{,}820$$

2. ¿Encontrar el valor de la combinación C (8,5).

$$\mathbf{C\ (n, r) = \frac{n!}{(n-r)!\ r!}}$$

4 7 2

$$C\ (8,5) = \frac{8!}{(8-5)!\ 5!} = \frac{8x7x6x5!}{3!\ \ 5!} = \frac{8x7x6x5!}{3x2x1!\ \ 5!} = 4x7x2 = 56$$

3. ¿En cuántas formas puede seleccionarse un destaca miento de 6 elementos entre 20 soldados?

Solución:

$$\mathbf{C\ (n, r) = \frac{n!}{(n-r)!\ r!}}$$

4 19 3 17 4 5

$$C\ (20, 6) = \frac{20!}{(20-6)!\ 6!} = \frac{20x19x18x17x16x15x14!}{14!\ \ 6x5x4x3x2x1!} = \frac{4x19x3x17x4x5}{2}$$

$$C\ (20,6) = \frac{77,520}{2} = 38,760$$

4. ¿Cuántos comités de 7 personas pueden formar con un conjunto de 25 personas?

Solución: El número de comités es igual al número de combinaciones de 25 elementos tomados en grupos de 7.

$$\mathbf{C\ (n, r) = \frac{n!}{(n-r)!\ r!}}$$

1

5 4 23 11 3 5 19

$$C\ (25,7) = \frac{25!}{(25-7)!\ 7!} = \frac{25x24x23x22x21x20x19x18!}{18!\ \ 7x6x5x4x3x2x1!} = 5x4x23x11x1x5x19$$

C (25,7) = 480,700

5. ¿De cuantas maneras se pueden formar equipos de 6 integrantes de un grupo de 36 alumnos?

Solución:

$$\mathbf{C\ (n, r) = \frac{n!}{(n-r)!\ r!}}$$

6 7 17 11 8 31

$$C\ (36,6) = \frac{36!}{(36-6)!\ 6!} = \frac{36x35x34x33x32x31x30!}{30!\ \ 6x5x4x3x2x1!} = 6x7xx17x11x8x31 = 1'947,792$$

6. Encontrar el valor de la combinación C(13,3)

Solución:

C (n, r) = $\dfrac{n!}{(n-r)!\,r!}$

$$C(13,3)= \frac{13!}{(13-3)!\,3!} = \frac{13!}{10!\,3!} = \frac{\overset{13}{13}x\overset{\overset{2}{4}}{12}x\overset{11}{11}x10!}{10! \quad 3x2x1!} = 13x2x11=286$$

7. De cuantas maneras se puede formar con 9 personas una comisión de 5 miembros?

Solución:

C(n, r) = $\dfrac{n!}{(n-r)!}$

$$C(9,5) = \frac{9!}{(9-5)!5!} = \frac{\overset{3}{9}x\overset{2}{8}x\overset{7}{7}x\overset{3}{6}x5!}{4x3x2x1! \quad 5!} = 3x2x7x3=126$$

8. Encontrar el valor de la combinación: C(47,6)

Solución:

C (n, r) = $\dfrac{n!}{(n-r)!}$

$$C(47,6)= \frac{47!}{(47-6)!6!} = \frac{47!}{41! \quad 6!} = \frac{\overset{47}{47}x\overset{23}{46}x\overset{\overset{3}{9}}{45}x\overset{11}{44}x\overset{43}{43}x\overset{7}{42}x41!}{41! \quad 6x5x4x3x2x1!} = 47x23x3x11x43x7=$$

C(47,6)= 10’737,573

TAREA:

I. Encontrar las combinaciones de los conjuntos de datos:

1. C(23,8)
2. C(14,9)
3. C(21,13)
4. C(19,15)
5. C(17,10)

2.3 INTRODUCCIÓN A LA PROBABILIDAD

La teoría de la probabilidad es la parte de las matemáticas que se encarga del estudio de los fenómenos o experimentos aleatorios. Por experimento aleatorio entenderemos todo aquel experimento que cuando se le repite bajo las mismas condiciones iniciales, el resultado que se obtiene no siempre es el mismo. El ejemplo más sencillo y cotidiano de un experimento aleatorio es el de lanzar una moneda o un dado, y aunque estos experimentos pueden parecer muy sencillos, algunas personas los utilizan para tomar decisiones en sus vidas. En principio no sabemos cuál será el resultado del experimento aleatorio, así que por lo menos conviene agrupar en un conjunto a todos los resultados posibles. El espacio muestral (o espacio muestra) de un experimento aleatorio es el conjunto de todos los posibles resultados del experimento, y se le denota generalmente por la letra griega Ω (omega). En algunos textos se usa también la letra S para denotar al espacio muestral. Esta letra proviene del termino samplings pace de la lengua inglesa equivalente a espacio muestral. Llamaremos evento a cualquier subconjunto del espacio muestral y denotaremos a los eventos por las primeras letras del alfabeto en mayúsculas: A, B, C, etc. Ejemplo. Si un experimento aleatorio consiste en lanzar un dado y observar el número que aparece en la cara superior, entonces claramente el espacio muestral es el conjunto Ω = {1, 2, 3, 4, 5, 6}. Como ejemplo de un evento para este experimento podemos definir el conjunto A = {2, 4, 6}, que corresponde al suceso de obtener como resultado un número par. ◦ Si al lanzar un dado una vez obtenemos el número “4”, decimos entonces que se observó la ocurrencia del evento A = {2, 4, 6}, y si se obtiene por ejemplo el resultado “1” decimos que no se observó la ocurrencia del evento A. Puesto que los conceptos de espacio muestral y evento involucran forzosamente la terminología de conjuntos, recordaremos a continuación algunas operaciones entre estos objetos y algunas propiedades que nos serán de utilidad en el estudio de la probabilidad y la estadística

2.3.1 DEFINICIÓN Y EXPRESIÓN DE PROBABILIDAD

El concepto de probabilidad nace con el deseo del hombre de conocer con certeza los eventos futuros. Es por ello que el estudio de probabilidad surge como una herramienta utilizada por los nobles para ganar los juegos y pasatiempos de la época. El desarrollo de estas herramientas fue asignado a los matemáticos de la corte. Con el tiempo estas técnicas matemáticas se perfeccionaron y encontraron otros usos muy diferentes para la que fueron creadas. Actualmente se continuo con el estudio de nuevas metodologías que permitan maximizar el uso de la computación en el estudio de las probabilidades disminuyendo, de este modo, los márgenes de errores en los cálculos.

2.4 EVENTOS MUTUAMENTE EXCLUYENTES Y NO EXCLUYENTES

Dos o más eventos son mutuamente excluyentes o disjuntos, si no pueden ocurrir simultáneamente. Es decir, la ocurrencia de un evento impide automáticamente la ocurrencia del otro evento (o eventos).

Dos o más eventos son no excluyentes o conjuntos, cuando es posible que curran ambos. Esto no indica que necesariamente deban ocurrir estos eventos en forma simultánea. La regla de la Adición expresa que la probabilidad de ocurrencia de al menos dos sucesos A y B es igual a:
P(A o B) = P(A) U P(B) = P(A) + P(B).

Si A y B son mutuamente excluyentes: P(A o B) = P(A) P(B) P(A y B). Básica. Si A y B son no excluyentes siendo: P(A) = probabilidad de Ocurrencia del evento P(B) = probabilidad de ocurrencia del evento BP (A y B), eventos independientes dos o más eventos son independientes cuando la ocurrencia o no – ocurrencia de un evento no tiene efecto sobre la probabilidad de ocurrencia del otro evento (o eventos). Un caso típico de eventos independientes es el muestreo con reposición, es decir, una vez tomada la muestra se regresa de nuevo a la población donde se obtuvo.

Eventos Mutuamente Excluyentes

Se dice que dos o más eventos son mutuamente excluyentes si la ocurrencia de cualquiera de ellos excluye la de los otros. De modo que, si E_1 y E_2 son eventos mutuamente excluyentes, entonces $Pr\{E_1E_2\} = 0$

Si E_1+E_2 denota el evento de que ocurra E_1 o bien E_2 o ambos a la vez, entonces:

$$Pr\{E_1+E_2\}=Pr\{E_1\}+Pr\{E_2\} - Pr\{E_1E_2\}$$

En particular,

$Pr\{E_1+E_2\}=Pr\{E_1\}+Pr\{E_2\}$ para evento mutuamente excluyentes

Como extensión de estos, si $E_1,E_2,…..E_n$, son n eventos mutuamente excluyentes con probabilidades respectivas de ocurrencia E_1 o E_2 o…..E_n, es $p_1+p_2+…..p_n$

Se dice que dos eventos son mutuamente excluyentes si no tienen puntos muéstrales en común.

Se dice que dos o más eventos son mutuamente excluyentes si la ocurrencia de cualquiera de ellos excluye la de los otros. De modo que si E1 y E2 son eventos mutuamente excluyentes, entonces Pr{E1E2} = 0

Si E1+E2 denota el evento de que ocurra E1 o bien E2 o ambos a la vez, entonces:

Pr{E1+E2}=Pr{E1}+Pr{E2} – Pr{E1E2}

En particular,

Pr{E1+E2}=Pr{E1}+Pr{E2} para evento mutuamente excluyentes

Como extensión de estos, si E1,E2,…..En, son n eventos mutuamente excluyentes con probabilidades respectivas de ocurrencia E1 o E2 o…..En, es p1+p2+…..pn

Ejemplos:

1. Si A y B son dos sucesos mutuamente excluyentes y la probabilidad de A es 0,2 y que ocurran ambos sucesos es:

   Solución:

   La probabilidad pedida es $P(A\cap C)$. Como son eventos
   Mutuamente excluyentes, ambos no pueden suceder a la vez,
   $P(A\cap C)=0$.

2. Se tienen cinco libros de distintas materias: Matemáticas, Biología, Química, Física y Lenguaje. Si se toma uno de ellos, ¿Cuál es la Probabilidad de que este sea de Matemática o Física?

   Solución:

   Sean los eventos:
   A=Tomar el libro de Matemáticas.
   B=Tomar el libro de Física.

La probabilidad pedida es: P(AUB)= P(A) + P(B) – P(A∩B) como A y B son eventos mutuamente excluyentes, P(A∩B)= 0.
Por lo tanto, la probabilidad pedida nos queda:
P(AUB)= P(A) + P(B) – P(A∩B).
Como A y B son mutuamente excluyentes, P(A∩B)= 0
Por lo tanto, la probabilidad pedida nos queda:
P (A∩B)= (1/5)+(1/5) – 0 = 2/5

3. En la tabla adjuntada, X representa el número de hijos por familia en grupo de 20 familias seleccionadas al azar. Si de este grupo se elige al azar una familia, ¿Cuál es la probabilidad de que tenga uno o dos hijos?
   Solución:
   El total de familias con uno o dos hijos son 6+3=9, de un total de 20 familias. La probabilidad pedida es p=9/20 p=0.45

4. En una bolsa se tienen 3 bolitas verdes, dos amarillas y 4 naranjas,
   Cuál es la probabilidad de que al sacar una bolita esta sea naranja o verde?
   Solución:
   Hay 4 bolitas naranjas y 3 verdes, esto es, 7 casos favorables a lo pedido. Aplicando la definición de la Laplace: Casos favorables 7
   p=casos favorables/casos totales=7/9

5. En una bolsa se tienen 3 bolitas rojas, 2 blancas y 4 azules. Se saca una al azar, ¿Cuál es la probabilidad de que sea roja o azul?

Solución:

P= (roja o azul)= casos favorables / casos totales,
P= (cantidad de bolas rojas + cantidad de bolas azules)/cantidad total de bolas en la bolsa.
P= (3+4)/(3+2+4)
P= 7/9.
La probabilidad de que salga un 2 o un 3 es igual a 1/6.
Los **eventos** mutuamente **excluyentes** son aquellos que es
un **evento** sucede significa que el otro no puede ocurrir.
Dos o más **eventos** son no **excluyentes**, o conjuntos, cuando es posible
que ocurran ambos. Esto no indica que necesariamente deban ocurrir estos eventos en forma simultánea.

Ejemplos:

1. Sean E1 el suceso "sacar un as de una baraja" y E2 "sacar un rey".
   Entonces:
   Solución:
   $Pr\{E_1\} = \frac{4}{52} = \frac{1}{13}$ y $Pr = \{E2\} = \frac{4}{52} = \frac{1}{13}$. Encontrar la probabilidad de sacar un as o un rey
   en un solo ensayo: $Pr = \{E_1+E_2\} = Pr\{E1\}+Pr\{E2\} = \underline{1} + \underline{1} = \underline{2}$

13 13 13

Dado que un as y un rey no pueden obtenerse en una sola extracción; por lo tanto, son eventos mutuamente excluyentes.

2. Si E_1 es el evento "extraer un as" de una baraja y E_2 es el evento "extraer una espada". Entonces E_1 y E_2 no son eventos mutuamente excluyentes, porque se puede obtener el as de espadas. Luego la probabilidad de extraer un as o un espada o ambos es:

$$Pr\{E_1+E_2\} = Pr\{E_1\}+Pr\{E_2\} - Pr\{E_1E_2\} = \frac{4}{52} + \frac{13}{52} - \frac{1}{52} = \frac{16}{52} = \frac{4}{13}$$

<u>Ley Aditiva para eventos mutuamente excluyentes</u>

Los eventos A y B, subconjuntos del espacio muestral E de cierto experimento aleatorio, son mutuamente excluyentes si y solo si:

$$A \cap B = \phi$$

Ejemplo:

Suponga que un espacio muestral tiene 5 resultados experimentales igualmente posibles.

$E_1 \quad E_2 \quad E_3 \quad E_4 \quad E_5$

sean:

$A=\{E_1 \quad E_2\}$

$B=\{E_3 \quad E_4\}$

$C=\{E_2\ E_3 \quad E_5\}$

a). Determine P(A),P(B) y P(C)

b). Determine P(AUB), son A y B mutuamente excluyentes

c). Determine A^c, C^c $P(A^c)$ y $P(C^c)$

d). Determine AUB^c y $P(AUB^c)$

e). Determine P(BUC)

Resultados: 100/5 = 0.20 = 20%

a). P(A) = 0.20+0.20 = 0.40

P(B) = 0.20+0.20 = 0.40

P(C) = 0.20+0.20+0.20 = 0.60

b). P(AUB) = P(A)+P(B) = 0.40+0.40= 0.80 si son mutuamente excluyentes

c). $A^c = \{E_3 \quad E_4 \quad E_5\} = P(A^c) = 0.20+0.20+0.20 = 0.60$

Cc $= \{E_1 \quad E_2\} = P(C^c) = 0.20+0.20 = 0.40$

d). $AUB^c = \{E_1 \quad E_2 \quad E_5\} = P(AUB^c) = 0.20+0.20+0.20 = 0.60$

e). BUC $= \{E_2 \quad E_3 \quad E_4 \quad E_5\} = 0.20+0.20+0.20+0.20 = 0.80$

2.5 PROBABILIDAD CONDICIONAL EVENTOS DEPENDIENTES E INDEPENDIENTES

A la probabilidad de que un evento B se dé cuando se sabe que algún otro evento A se ha presentado se llama **probabilidad condicional** y se le escribe esta expresión $P(B \mid A)$, por lo común, se lee "la probabilidad de que B ocurra dado que ocurrió A", o simplemente "la probabilidad de B, dado A".

Considérese el evento B de obtener un cuadrado perfecto cuando se lanza un dado. El dado se ha considerado de tal forma que los números pares tienen dos veces más posibilidad de presentarse que los números impares. Con base en el espacio muestral S = {1, 2, 3, 4, 5, 6} con probabilidades de 1/9 y 2/9 para los impares y pares, respectivamente, la probabilidad de que ocurra B es 1/3. Supóngase que se sabe que el lanzamiento resulto en un número mayor que 3. Se tiene un espacio muestral reducido A = {4, 5, 6}, el cual es subconjunto de S. Para encontrar la probabilidad de que B ocurra, con relación al espacio A, se deberá asignar primero nuevas probabilidades proporcionalmente a los elementos de A proporcionalmente a sus probabilidades originales, de tal forma que su suma sea 1. Al asignar una probabilidad de w al número impar en A y una probabilidad de 2w a los dos números pares, se tiene que 5w = 1 o w = 1/5. En relación con el espacio A, se encuentra que B contiene solo el elemento 4. Si este evento se representa con el símbolo $B \mid A$, se tiene que $B \mid A = \{4\}$, y de aquí:

$$P(B/A) = \frac{2}{3}$$

Este ejemplo muestra que los eventos pueden tener probabilidades diferentes cuando se consideran relativos a diferentes espacios muéstrales. También se puede escribir:

$$P(B/A) = \frac{2}{5} = \frac{2/9}{5/9} = \frac{P(A \cap B)}{P(A)}$$

Donde la $P(A \cap B)$ y la $P(A)$ se encuentran a partir del espacio muestral Original S. En otras palabras, una probabilidad condicional relativa a un Sub-espació A de S puede calcularse directamente a partir de las probabilidades que se asignan a los elementos del espacio muestral S.

La probabilidad condicional de B, dado A, que se indica por P(B/A), se
Define:

$$P(B/A) = \frac{P(A \cap B)}{P(A)} \qquad \text{si } P(A) \geq 0$$

Eventos Dependientes e Independientes

A grandes rasgos podemos decir que las variables son símbolos que formarán parte de fórmulas o funciones, dentro del terreno matemático. Pueden tomar diferentes valores y ahí tendremos que mencionar dos de las principales: variable dependiente e independiente.

Además de contar lo que cada una de ellas significa, nada como una serie de ejemplos para poder entender a la perfección todas sus funciones. ¡Verás cómo una vez que se entiende, ya no parece tan complicado como en un primer momento parecía!

Definición de Variable Dependiente e Independiente

Como ya hemos avanzado, la variable dependiente e independiente son las dos variables más importantes dentro de cualquier tipo de investigación. Para saber la función que tiene cada una y a grandes rasgos, podemos decir que la variable independiente es la causa de algo, mientras que la dependiente será el efecto de ese algo. A modo de ejemplo, el consumo de azúcar hará que nuestro peso incremente. Así que, se traduce como que el tomar azúcar sería la variable independiente y el incremento de peso, la variable dependiente.

Variable dependiente y sus ejemplos

Los valores que adopte una variable dependiente, siempre van a estar ligados a otra. Es decir, ésta dependerá siempre de la otra variable, de ahí su nombre. Por lo que el valor de la misma, irá acorde a la modificación de la otra variable. Al ir en directa relación con la variable independiente, hará que los errores en la investigación sean menores. Las variables dependientes pueden tomar valores de tipo numéricos. Ahí mencionaríamos tanto las variables cuantitativas como cualitativas.

Toda explicación siempre se entiende mejor con grandes ejemplos. Si haces un viaje largo en coche, en el que recorrerás unos 600 kilómetros, diremos que la velocidad es la variable independiente. Mientras que la duración del viaje sería la variable dependiente. ¿Por qué?, pues porque la duración del trayecto va a depender de la velocidad que llevemos. No es lo mismo ir a 80 km/h que a 120 km/h. Se supone que cuando vamos un poco más rápido, siempre dentro de los límites establecidos, el trayecto se terminará antes.

Lo mismo ocurre cuando vamos a comprar. No siempre pagamos el mismo dinero por hacer la compra. Todo dependerá de la cantidad de productos que hayamos elegido. Por lo que, de nuevo, la variable dependiente será el dinero final que nos marque el ticket y que depende de los productos así como de sus cantidades. Otros ejemplos a tener en cuenta:

Después de varias horas de ejercicio físico (variable independiente), nos sentiremos cansados (variable dependiente o efecto del ejercicio).

Si comemos poco o nada durante varias horas (variable independiente), tendremos hambre (variable dependiente o efecto de no haber comido).

Cuando realizas un trabajo, te pagan 20 euros. En este caso la variable dependiente será el dinero que ganes, porque si haces más trabajos, te pagarán el doble o el triple de lo mencionado.

Variable independiente

A la variable independiente también se le conoce como como ‘manipulada’, ya que debido a ella, puede dar como resultado varios ejemplos de variables dependientes. Se dice que en un experimento no suele haber más de dos variables independientes. Porque si no, los resultados pueden no ser del todo fiables. Se trata de una variable que está aislada de otros factores y es por este motivo por el cual, hay una manipulación de carácter experimental. Consiguiendo así resultados que pueden ser analizados. Hay que decir que, en una función, el valor de la variable independiente se puede fijar de manera libre y es un tipo de valor que no depende de ninguna otra.

Número de horas de que tiene un día. Es algo que no depende de ningún tipo de estación, sino que es un valor predeterminado. Eso sí, por ejemplo, las horas de sol ya dependerán del mes o de la estación en la que nos encontremos.

La deshidratación es el efecto o variable dependiente de las horas que has dejado al cuerpo sin proporcionarle agua. Así que, las horas sin beber es la variable independiente.

La cantidad de los productos vendidos en una tienda, es también independiente. Ya que las ganancias serían las variables dependientes, porque como su nombre indica, el resultado dependerá de muchos factores.

Combinando ejemplos de variables dependientes e independientes
Si ya tenemos claro qué es una variable dependiente, así como la independiente y sus ejemplos, nada como combinar ambas opciones. Quizás de este modo, lleguemos a darles un último repaso y a aclararnos todavía un poco más. Una forma de poner en práctica todo lo que hemos aprendido.

En un examen de matemáticas, consigues 5 puntos por cada pregunta contestada correctamente.

Variable dependiente: El número de puntos que obtengas.
Variable independiente: El número de preguntas que has respondido bien.
Compras varias cajas de galletas. Cada una cuesta 3 euros.
Variable dependiente: La cantidad de dinero que gastas en las galletas.
Variable independiente: El número de cajas que compres.
Contratas un nuevo servicio de telefonía que cuesta 40 euros cada mes.
Variable dependiente: El precio total que pagues por el servicio.
Variable independiente: El tiempo, es decir, los meses que vas a La regla de la multiplicación o regla del producto, permite encontrar la probabilidad de que ocurra el evento A y el evento B al mismo tiempo (probabilidad conjunta). Esta regla depende de si los eventos son dependientes o independientes.

La regla de la multiplicación o regla del producto**,** permite encontrar la probabilidad de que ocurra el evento A y el evento B al mismo tiempo (probabilidad conjunta). Esta regla depende de si los eventos son dependientes o independientes.

La regla de la multiplicación establece que la probabilidad de ocurrencia de dos o más eventos estadísticamente independientes es igual al producto de sus probabilidades individuales.

P(A y B) = P (AB) = P(A) P (B), si A y B, son independientes.
P(A y B) = P (AB) = P(A) P (B/A) =, si A y B, son dependientes.

La regla de la multiplicación se refiere a la determinación de la probabilidad de la ocurrencia conjunta de A y b.

Existen dos opciones de esta regla:

1. Si los eventos son independientes:

   P(A y B) = P(A o B) = P (A) P (B)

2. Si los eventos son dependientes:

Es la probabilidad de A multiplicada por la probabilidad condicional de B dado A.

P(A Y B) = P (A) P (B/A)

Si la posición de los dos eventos se invierte, se obtiene un valor equivalente.

P(A y B) = P (B/A) = P (B) P(A/B)

Ejemplo:
Si una moneda equilibrada se lanza dos veces, la probabilidad de que ambos lanzamientos den por resultado una "cara" es:

(1/2) x (1/2) = 1/4

Eventos Dependientes
Dos eventos A y B son dependientes**,** si la ocurrencia de uno de ellos afecta la ocurrencia del otro. Para eventos dependientes, la regla de la multiplicación establece que:

$$P(A \cap B) = P(A) \times P(B|A)$$

$$\mathrm{P}(A \cap B) = P(B) \times P(A|B)$$

Ejemplo 1:
Una caja contiene 2 canicas azules y 3 rojas. Si se extraen dos canicas al azar sin reposición, ¿cuál es la probabilidad de que las dos sean azules?
Solución:
Dado que las canicas serán extraídas de la misma caja, y que las canicas que se extraigan, no serán devueltas a la caja (no hay reposición), entonces, se trata de eventos dependientes.

Evento A: obtener una canica azul en la primera extracción.
Evento B: obtener una canica azul en la segunda extracción.
Por la regla de la multiplicación, sabemos que:

$$P(A \cap B) = P(A) \times P(B|A)$$

$$P(A \cap B) = \frac{2}{5} \cdot \frac{1}{4} = \frac{2}{20} = 0{,}1 = 10\ \%$$

Eventos Independientes

Dos eventos A y B son independientes**,** si la ocurrencia de uno de ellos no afecta la ocurrencia del otro, es decir, cuando los eventos A y B no están relacionados. Para eventos independientes, la regla de la multiplicación establece que:

$$P(A \cap B) = P(A) \times P(B)$$

Esto se debe, a que, en los eventos independientes, la ocurrencia de un evento, no afecta a la ocurrencia del otro:

$$P(A|B) = P(A) \quad \wedge \quad P(B|A) = P(B)$$

Ejemplo 2:

En un colegio, la probabilidad de que un alumno seleccionado al azar hable inglés es de 0,20; mientras que la probabilidad de que un alumno juegue fútbol es de 0,80. El hecho de que un alumno hable inglés, no afecta en nada que juegue fútbol; por lo tanto, se trata de eventos independientes.

- Evento A: que el alumno hable inglés. **P(A) = 0,20.**
- Evento B: que el alumno juegue fútbol. **P(B) = 0,80.**

Usamos la regla de la multiplicación para eventos independientes:

$$P(A \cap B) = P(A) \times P(B)$$

$$P(A \cap B) = 0{,}20 \times 0{,}80 = 0{,}16 = 16\ \%$$

Ejemplo 3:

Sabiendo que P(A) = 0,70; P(B) = 0,50; y además, P(A∩B) = 0,40; determinar si son eventos dependientes o independientes.

Solución:

En los eventos independientes, se cumple que:

$$P(A \cap B) = P(A) \times P(B)$$

En este caso:

- P(A∩B) = 0,40.
- P(A) × P(B) = 0,70 × 0,50 = 0,35.

Podemos ver que 0,40 es diferente de 0,35; entonces:

$$P(A \cap B) \neq P(A) \times P(B)$$

Podemos concluir que no son eventos independientes, es decir, **son eventos**

Ejemplo 4:
Una caja contiene 2 canicas azules y 3 rojas. Si se extraen dos canicas al azar sin reposición, ¿cuál es la probabilidad de que las dos sean azules?
Solución:
Dado que las canicas serán extraídas de la misma caja, y que las canicas que se extraigan, no serán devueltas a la caja (no hay reposición), entonces, se trata de eventos dependientes.

- Evento A: obtener una canica azul en la primera extracción.
- Evento B: obtener una canica azul en la segunda extracción.

Por la regla de la multiplicación, sabemos que:

$$P(A \cap B) = P(A) \times P(B|A)$$

$$P(A \cap B) = \frac{2}{5} \cdot \frac{1}{4} = \frac{2}{20} = 0{,}1 = 10\ \%$$

TAREA:
1**–** Usando las siguientes 2 variables de un estudio, “comentarios positivos” y “autoestima”, con el ejercicio sugerido, se leería de la siguiente manera: Los comentarios positivos causan un cambio en la autoestima y es imposible que la autoestima pueda causar cambios en los comentarios positivos.

Desde un punto de vista más lógico y científico, la proposición anterior tiene mucho sentido y funciona para ilustrar la identificación y diferenciación entre variables dependientes e independientes.

Como se dijo en puntos anteriores, con un estudio mucho más profundo desde el punto de vista social o psicológico, se pudiera debatir los casos en los que tener una buena autoestima puede causar efectos positivos en las personas que podrían traducirse en comentarios positivos.

2- En “La exposición a más luz solar aumenta los niveles de felicidad en trabajadores que permanecen todo el día en oficinas cerradas”, al usar el ejercicio sugerido, exposición al sol resultaría como la variable independiente y nivel de felicidad sería la dependiente.

La independiente se podría controlar con el tiempo de exposición (horas, días, semanas), y la dependiente con una escala múltiple donde se le pregunte a los trabajadores al final del día simplemente cómo se sienten.

3- En la pregunta “¿Cuáles son los beneficios o agravantes de las redes sociales en niños?”, claramente se puede identificar a las redes sociales como la variable independiente, porque se asume que causa un efecto beneficioso o agravante en los niños. Ese efecto es lo que se propone como objeto de estudio, por consiguiente, es la variable dependiente.

4- ¿Cuánta agua fluye por un grifo en diferentes aperturas?: La variable independiente sería la apertura del grifo de agua controlado como cerrado, poco abierto, medio abierto y completamente abierto. La variable dependiente sería el flujo de agua medido en litros por minuto.

5**–** Un motor eléctrico gira más rápido al aumentarle el voltaje: voltaje de electricidad controlado en voltios, variable independiente. Velocidad de rotación medida en revoluciones por minuto, variable dependiente.

2.6 TEOREMA DE BAYES

El Teorema de Bayes es un procedimiento que nos permite expresar la probabilidad condicional de un evento aleatorio A dado B, en términos de la distribución de probabilidad del evento B dado A y la distribución de probabilidad de solo A.

Este teorema es de mucha utilidad, ya que gracias a él podemos relacionar la probabilidad de que un evento A ocurra sabiendo que ocurrió B, con la probabilidad de que ocurra lo contrario, es decir, que ocurra B dado A.

El teorema de Bayes fue una proposición plateada por el reverendo Thomas Bayes, un teólogo inglés del siglo XVIII quien también fue matemático. Fue autor de varios trabajos en teología, pero en la actualidad es conocido por un par de tratados matemáticos, entre los cuales destaca como resultado principal el ya referido Teorema de Bayes.

Bayes se ocupó de este teorema en un trabajo titulado “An Essay towards solving a Problem in the Doctrine of Chances” (Un ensayo para resolver un problema en la doctrina de las posibilidades), publicado en 1763, y sobre el cual se han desarrollado grandes estudios con aplicaciones en diversas áreas de conocimiento.

Primero, para una mayor compresión de este teorema, son necesarias algunas nociones básicas de teoría de probabilidad, especialmente el teorema de la multiplicación para probabilidad condicional, el cual establece que

$$P(E \cap A) = P(E)P(A|E)$$

Para E y A eventos arbitrarios de un espacio muestral S.

Y la definición de particiones, la cual nos dice que si tenemos A1 ,A2,…, An eventos de un espacio muestral S, estos formarán una partición de S, si los Ai son mutuamente exclusivos y su unión es S.

Teniendo esto, sea B otro evento. Entonces podemos ver a B como

$$B = S \cap B = (A_1 \cup A_2 \cup \ldots \cup A_n) \cap B$$
$$= (A_1 \cap B) \cup (A_2 \cap B) \cup \ldots (A_n \cap B)$$

Donde las Ai intersectadas con B son eventos mutuamente exclusivos.Y en consecuencia.

$$P(B) = P(A_1 \cap B) + P(A_2 \cap B) + \cdots + P(A_n \cap B)$$

Luego, aplicando el teorema de la multiplicación

$$P(B) = P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + \cdots + P(A_n)P(B|A_n)$$

Por otra parte, la probabilidad condicional de Ai dado B se define por

Sustituyendo de manera adecuada tenemos que para cualquier i

$$P(A_i|B) = \frac{P(A_i \cap B)}{P(B)}$$

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + \cdots + P(A_n)P(B|A_n)}$$

Con base en la definición de Probabilidad condicionada se obtiene la Fórmula de Bayes, también conocida como la Regla de Bayes.

Donde B es el suceso sobre el que tenemos información previa y A(n) son los distintos sucesos condicionados. En la parte del numerador tenemos la probabilidad condicionada, y en la parte de abajo la probabilidad total. En cualquier caso, aunque la fórmula parezca un poco abstracta, es muy sencilla. Para demostrarlo, utilizaremos un ejemplo en el que en lugar de A(1), A(2) y A(3), utilizaremos directamente A, B y C.

Ejercicios:

1. Una empresa de celulares tiene dos máquinas A y B. El 54% de los celulares producidos son hechos por la máquina A y el resto por la máquina B. No todos los celulares producidos están en buen estado.

La proporción de celulares defectuosos hechos por A es 0.2 y por B es 0.5. ¿Cuál es la probabilidad de que un celular de dicha fábrica sea defectuoso? ¿Cuál es la probabilidad de que, sabiendo que un celular es defectuoso, proceda de la máquina A?

Solución:
Aquí, se tiene un experimento que se realiza en dos partes; en la primera parte ocurren los eventos:

A: celular hecho por la máquina A.

B: celular hecho por la máquina B.

Ya que la máquina A produce el 54% de los celulares y el resto los produce la máquina B, se tiene que la máquina B produce el 46% de los celulares. Las probabilidades de estos eventos son dadas, a saber:

P(A)=0,54.
P(B)=0,46.
Los eventos de la segunda parte del experimento son:

B: celular defectuoso.

E: celular no defectuoso.

Como se dice en el enunciado, las probabilidades de estos eventos dependen del resultado obtenido en la primera parte:

P(D|A)=0,2.

P(D|B)=0,5.

Utilizando estos valores, se puede determinar también las probabilidades de los complementos de estos eventos, es decir:

P(E|A) = 1 – P(D|A)

= 1 – 0,2 = 0,8 y

p(E|B) = 1 – P(D|B)

= 1 – 0,5 = 0,5.

Ahora, el evento B se puede escribir como sigue:

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

Utilizando el Teorema de la multiplicación para probabilidad condicional resulta:

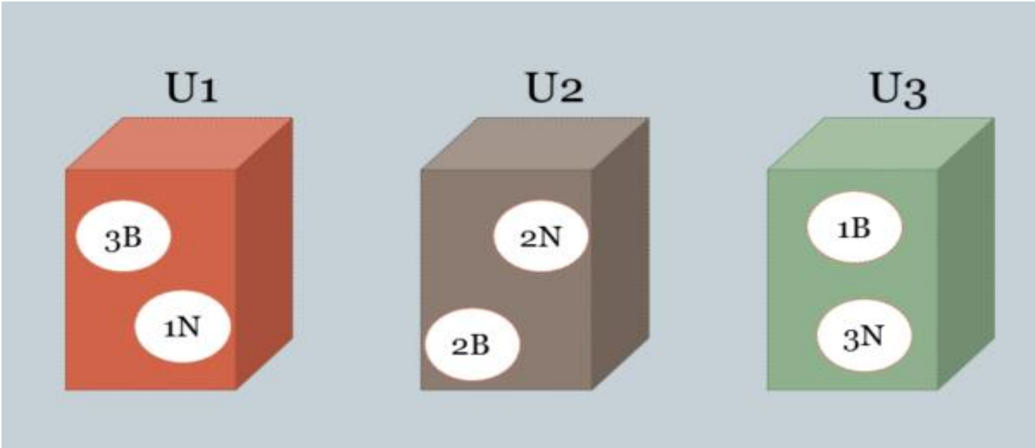
$$\begin{aligned} P(D) &= P(D|A)P(A) + P(D|B)P(B) \\ &= (0.2)(0.54) + (0.5)(0.46) \\ &= 0.338 \end{aligned}$$

Con lo cual se responde la primera pregunta.
Ahora solo falta calcular P(A|D), para lo cual se aplica el Teorema de Bayes:

$$\begin{aligned} P(A|D) &= \frac{P(D|A)P(A)}{P(D|A)P(A) + P(D|B)P(B)} \\ &= \frac{(0.2)(0.54)}{(0.2)(0.54) + (0.5)(0.46)} \\ &= 0.319 \end{aligned}$$

Gracias al Teorema de Bayes se puede afirmar que la probabilidad de que un celular haya sido hecho por la máquina A, sabiendo que el celular es defectuoso, es de 0.319.

2. Tres cajas contienen bolas blancas y negras. La composición de cada una de ellas es la siguiente: U1 = {3B, 1N}, U2 = {2B, 2N}, U3 = {1B, 3N}.



Se elige al azar una de las cajas y se extrae de ella una bola al azar la cual resulta ser blanca. ¿Cuál es la caja con mayor probabilidad de haber sido elegida?

Solución

Mediante U1, U2 y U3, representaremos también la caja elegida.
Estos sucesos constituyen una partición de S y se verifica que P(U1) = P(U2) = P(U3) = 1/3 ya que que la elección de la caja es al azar.
Si B={la bola extraída es blanca}, tendremos P(B|U1) = 3/4, P(B|U2) = 2/4 , P(B|U3) = 1/4 .
Lo que deseamos obtener es la probabilidad de que la bola haya sido sacada de la caja Ui sabiendo que dicha bola fue blanca, es decir, P(Ui |B), y ver cuál de los tres valores fue el más alto para conocer de cuál caja ha sido más probable la extracción de la bola blanca.

Aplicando el teorema de Bayes a la primera de las cajas:

$$P(U_1|B) = \frac{\frac{1}{3} * \frac{3}{4}}{\frac{1}{3} * \frac{3}{4} + \frac{1}{3} * \frac{2}{4} + \frac{1}{3} * \frac{1}{4}} = \frac{3}{6}$$

Y para las otras dos:

P(U2|B) = 2/6 y P(U3|B) = 1/6.

Luego, la primera de las cajas es la que tiene una mayor probabilidad de haber sido elegida para la extracción de la bola blanca.

3. Una empresa tiene una fábrica en Estados Unidos que dispone de tres máquinas A, B y C, que producen envases para botellas de agua. Se sabe que la máquina A produce un 40% de la cantidad total, la máquina B un 30%, y la máquina C un 30%. También se sabe que cada máquina produce envases defectuosos. De tal manera que la máquina A produce un 2% de envases defectuosos sobre el total de su producción, la máquina B un 3%, y la máquina C un 5%. Dicho esto, se plantean dos cuestiones:

P(A) = 0,40 **P(D/A) =** 0,02
P(B) = 0,30 **P(D/B) =** 0,03
P(C) = 0,30 **P(D/C) =** 0,05

Si un envase ha sido fabricado por la fábrica de esta empresa en Estados Unidos ¿Cuál es la probabilidad de que sea defectuoso?

Se calcula la probabilidad total. Ya que, a partir los diferentes sucesos, calculamos la probabilidad de que sea defectuoso.

P(D) =[P(A) x P(D/A)] + [P(B) x P(B/A)] + [P(C) x P(C/A)] = [0,4 x 0,02] + [0,3 x 0,03] + [0,3 x 0,05] = 0,032

Expresado en porcentaje, diríamos que la probabilidad de que un envase fabricado por la fábrica de esta empresa en Estados Unidos sea defectuoso es del 3,2%.

4. Siguiendo con la pregunta anterior, si se adquiere un envase y este es defectuoso ¿Cuáles es la probabilidad de que haya sido fabricado por la máquina A?¿Y por la máquina B?¿Y por la máquina C?

Aquí se utiliza el teorema de Bayes. Tenemos información previa, es decir, sabemos que el envase es defectuoso. Claro que, sabiendo que es defectuoso, queremos saber cuál es la probabilidad de que se haya producido por una de las máquinas.

$P(A/D) = [P(A) \times P(D/A)] / P(D) = [0{,}40 \times 0{,}02] / 0{,}032 = 0{,}25$
$P(B/D) = [P(B) \times P(D/B)] / P(D) = [0{,}30 \times 0{,}03] / 0{,}032 = 0{,}28$
$P(C/D) = [P(C) \times P(D/C)] / P(D) = [0{,}30 \times 0{,}05] / 0{,}032 = 0{,}47$

Sabiendo que un envase es defectuoso, la probabilidad de que haya sido producido por la máquina A es del 25%, de que haya sido producido por la máquina B es del 28% y de que haya sido producido por la máquina C es del 47%.

5. En la sala de pediatría de un hospital, el 60% de los pacientes son niñas. De los niños el 35% son menores de 24 meses. El 20% de las niñas tienen menos de 24 meses. Un pediatra que ingresa a la sala selecciona un infante al azar.

a. Determine el valor de la probabilidad de que sea menor de 24 meses.
b. Si el infante resulta ser menor de 24 meses. Determine la probabilidad que sea una niña.

SOLUCIÓN:
Se definen los sucesos:
Suceso H: seleccionar una niña.
Suceso V: seleccionar un niño.

Suceso M: infante menor de 24 meses.

En los ejercicios de probabilidad total y teorema de bayes, es importante identificar los sucesos que forman la población y cuál es la característica que tienen en común dichos sucesos. Estos serán los sucesos condicionados.

a. En este caso, la población es de los infantes. Y la característica en común es que sean menores de 24 meses. Por lo tanto, la probabilidad de seleccionar un infante menor de 24 meses es un ejemplo de probabilidad total. Su probabilidad será:

$$P(M)=P(H) \times P(M/H) + P(V) \times P(M/V) = 0.6 \times 0.2+0.4 \times 0.35 = 0.26 \text{ o } 26\%$$

b). Para identificar cuando en un ejercicio se hace referencia al teorema de bayes, hay que partir de reconocer esta es una probabilidad condicionada y que la característica común de los sucesos condicionantes ya ha ocurrido. Entonces, la probabilidad de que sea niña una infanta menor de 24 meses será:

$$P(H/M) = \frac{P(H) \times P(M/H)}{P(H/M) \times P(M/H) + P(V) \times P(M/V)} = \frac{0.6 \times 0.2}{0.6x0.2+0.4x0.35} = \frac{0.12}{0.26} = 0.46 \text{ o } 46\%$$

7. Un médico cirujano se especializa en cirugías estéticas. Entre sus pacientes, el 20% se realizan correcciones faciales, un 35% implantes mamarios y el restante en otras cirugías correctivas. Se sabe, además, que son de género masculino el 25% de los que se realizan correcciones faciales, 15% implantes mamarios y 40% otras cirugías correctivas. Si se selecciona un paciente al azar, determine:

a). Determine la probabilidad de que sea de género masculino
b). Si resulta que es de género masculino, determine la probabilidad que se haya realizado una cirugía de implantes mamarios.

SOLUCIÓN:
Se definen los sucesos:
Suceso *F*: pacientes que se realizan cirugías faciales
Suceso *M*: pacientes que se realizan implantes mamarios
Suceso *O*: pacientes que se realizan otras cirugías correctivas
Suceso *H*: pacientes de género masculino

a). La probabilidad de que sea de género masculino se refiere a un problema de probabilidad total, ya que es el suceso condicionado y las cirugías los condicionantes. Dicho valor será:

$P(H) = P(F) \times P(H/F) + P(M) \times P(H/M) + P(O) \times P(O/H)$
$P(H) = 0.2 \times 0.25 + 0.35 \times 0.15 + 0.45 \times 0.40 = 0.28$ o 28%

b). Como el suceso condicionado ha ocurrido entonces se aplica el teorema de bayes, luego, el valor de la probabilidad será:

$$P(M/H) = \frac{P(M) \times P(H/M)}{P(F) \times P(H/F) + P(M) \times P(H/M) + P(O) \times P(O/H)} =$$
$$P(M/H) = \frac{0.35 \times 0.15}{0.2 \times 0.25 + 0.35 \times 0.15 + 0.45 \times 0.40} = \frac{0.0525}{0.2825} = 0.19 \text{ o } 0.19\%$$

8. Un Doctor dispone de tres equipos electrónicos para realizar ecosonogramas. El uso que le da a cada equipo es de 25% al primero, 35% el segundo en y 40% el tercero. Se sabe que los aparatos tienen probabilidades de error de 1%, 2% y 3% respectivamente. Un paciente busca el resultado de una ecografía y observa que tiene un error. Determine la probabilidad de que se ha usado el primer aparato.

SOLUCIÓN:

Se definen los sucesos:

Suceso *P*: seleccionar el primer aparato

Suceso *S*: seleccionar el segundo aparato

Suceso *T*: seleccionar el tercer aparato

Suceso *E*: seleccionar un resultado con error

Se puede observar que la pregunta es sobre determinar la probabilidad de que un examen errado sea del primer aparato, es decir, ya ha ocurrido el error. Por lo tanto, debemos recurrir al teorema de bayes. Claro está, que es necesario de igual forma obtener la probabilidad de que los aparatos produzcan un resultado erróneo, por lo tanto:

$$P(P/E) = \frac{P(P) \times P(E/P)}{P(P) \times P(E/P) + P(S) \times P(E/S) + P(T) \times P(E/T)} =$$

$$P(P/E) = \frac{0.25 \times 0.01}{0.03\ 0.25 \times 0.01 + 0.35 \times 0.02 + 0.4 \times 0.03} = \frac{0.0025}{0.0215} = 0.116 = 0.12 \text{ o } 12\%$$

TAREA:

1. Tres máquinas, A, B y C, producen el 45%, 30% y 25%, respectivamente, del total de las piezas producidas en una fábrica. Los porcentajes de producción defectuosa de estas máquinas son del 3%, 4% y 5%.

a). Seleccionamos una pieza al azar; calcula la probabilidad de que sea defectuosa.
b). Tomamos, al azar, una pieza y resulta ser defectuosa; calcula la probabilidad de haber sido producida por la máquina B.
c). ¿Qué máquina tiene la mayor probabilidad de haber producido la citada pieza defectuosa?

2. La máquina con mayor probabilidad de haber producido la pieza defectuosa es A. Tenemos tres urnas: A con 3 bolas rojas y 5 negras, B con 2 bolas rojas y 1 negra y C con 2 bolas rojas y 3 negras. Escogemos una urna al azar y extraemos una bola. Si la bola ha sido roja, ¿cuál es la probabilidad de haber sido extraída de la urna A?

3. El 20% de los empleados de una empresa son ingenieros y otro 20% son economistas. El 75% de los ingenieros ocupan un puesto directivo y el 50% de los economistas también, mientras que los no ingenieros y los no economistas solamente el 20% ocupa un puesto directivo. ¿Cuál es la probabilidad de que un empleado directivo elegido al azar sea ingeniero?

4. La probabilidad de que ha ya un accidente en una fábrica que dispone de alarma es 0.1. La probabilidad de que suene esta sí se ha producido algún incidente es de 0.97 y la probabilidad de que suene si no ha sucedido ningún incidente es 0.02 En el supuesto de que ha ya funcionado la alarma, ¿cuál es la probabilidad de que no haya habido ningún incidente?
Sean los sucesos: I = Producirse incidente. A = Sonar la alarma.

2.7 VALOR ESPERADO O ESPERANZA MATEMÁTICA

La Esperanza Matemática (también llamada media o valor esperado) es el valor medio de un fenómeno aleatorio.

La Esperanza Matemática se representa por E(x). Cuando la variable aleatoria es discreta (toma solamente números discretos: 1, 2, 3, 4...no continuos) la esperanza matemática es igual a la suma de los productos de la probabilidad de cada suceso por el valor de dicho suceso:

$$E(x) = x_1 \cdot p_1 + x_2 \cdot p_2 + \cdots + x_i \cdot p_i = \sum_{i=1}^{n} x_i \cdot p_i$$

Los nombres de **esperanza matemática** y **valor esperado** tienen su origen en los juegos de azar y hacen referencia a la ganancia promedio esperada por un jugador cuando hace un gran número de apuestas.

Si la **esperanza matemática** es **cero**, E(x) = 0, el **juego** es **equitativo**, es decir, no existe ventaja ni para el jugador ni para la banca.

En estadística el valor esperado o esperanza matemática (o simplemente esperanza) de una variable aleatoria es la suma de la probabilidad de cada suceso multiplicado por su valor. Por ejemplo en un juego de azar el valor esperado es el beneficio medio.

Si todos los sucesos son de igual probabilidad la esperanza es la media aritmética
El valor esperado o esperanza, abreviado como EV, de un movimiento es la ganancia o pérdida media resultante de una situación teniendo en cuenta todos los resultados posibles y sus probabilidades.

Cálculo de la esperanza
Valor Esperado o Esperanza Matemática

La esperanza matemática o valor esperado de una variable aleatoria discreta es la suma del producto de la probabilidad de cada suceso por el valor de dicho suceso.

Los nombres de esperanza matemática y valor esperado tienen su origen en los juegos de azar y hacen referencia a la ganancia promedio esperada por un jugador cuando hace un gran número de apuestas.

Si la esperanza matemática es cero, E(x) = 0, el juego es equitativo, es decir, no existe ventaja ni para el jugador ni para la banca.

El valor esperado de una Variable Aleatoria X es el promedio ponderado de todos los valores posibles de la misma. No de los pesos son las probabilidades asociadas con los valores.

$E(x) = \mu = E\ xf\ (x)$

<u>Varianza</u>

Es un promedio ponderado de las de las desviaciones al cuadrado.

Varianza = $E\ (x - \mu)^2\ f\ (x)$

Ejemplos:

a). Si una persona compra una papeleta en una rifa, en la que puede ganar de 5,000 C, o un segundo premio de 2,000 con probabilidades de 0.001 y 0.003, ¿Cuál sería el precio justo a pagar por la papeleta?

$E(x) = 5{,}000 \times 0.001 + 2{,}000 \times 0.003 = 11C$

b). Un jugador lanza dos monedas gana 1 o 2 C, SI aparecen uno o dos caras. Por otra parte, pierde 5 C, si no aparece cara. Determinar la esperanza matemática del juego y si este es favorable.

$E(x) = \{(c,c), (c,x), (x,c), (x,x)\}$

$P(+2) = 2/4$
$P(+1) = 1/4$
$P(-5) = 1/4$
$E(x) = 1 \times 2/4 + 2 \times \frac{1}{4} - 5 \times \frac{1}{4} = -1/4$ Es desfavorable

Donde f(x) es, respectivamente, la función de probabilidad o la función densidad de probabilidad y g(x*)* es cualquier función de valores reales que está definida para todos los valores posibles de x.

Veamos algunos ejemplos para entender mejor el concepto de esperanza matemática:

Ejemplos:
1. Supongamos que en el juego de los dados si sale 1, 2 o 3 pierdo un dólar, si sale un 4 o un 5 no gano nada y si sale 6 gano 2 dólares. ¿Cuánto puedo esperar ganar si juego 100 veces seguidas?
SOLUCION:

$$E(x) = \Sigma\, x_i \cdot p_i = x_1 \cdot p_1 + x_2 \cdot p_2 + x_3 \cdot p_3 + ... + x_n \cdot p_n$$

La probabilidad de cada suceso es igual: $p_1=1/6$, $p_2=1/6$, $p_3=1/6$, $p_4=1/6$, $p_5=1/6$, p6=1/6.
Los valores de los sucesos son: $x_1 = x_2 = x_3 = -1$, $x_4 = x_5 = 0$, $x_6 = 2$
$E(x) = -1\cdot 1/6 + -1 \cdot 1/6 + -1 \cdot 1/6 + 0 \cdot 1/6 + 0 \cdot 1/6 + 2 \cdot 1/6 =$
$E(x) = -1/6 -1/6 -1/6 +2/6 = -1/6$
$100 \cdot (-1/6) = -100/6 = -16.7$ dólares → si se tira el dado 100 veces puedo esperar perder unos 16.7 dólares de media.

2. Sea un juego de mesa en el que al tirar un dado avanzo tantas casillas como salga en el dado. Se gana al llegar a la casilla 100. ¿Cuántas jugadas se puede esperar necesitar para llegar a la meta?

La probabilidad de cada suceso es igual: $p_1=1/6$, $p_2=1/6$, $p_3=1/6$, $p_4=1/6$, $p_5=1/6$ $p_6=1/6$.

Los valores de los sucesos son: $x_1 = 1$, $x_2 = 2$, $x_3 = 3$, $x_4 = 4$, $x_5 = 5$, $x_6 = 6$

$E(x) = 1\cdot 1/6 + 2 \cdot 1/6 + 3 \cdot 1/6 + 4 \cdot 1/6 + 5 \cdot 1/6 + 6 \cdot 1/6 = 1/6 + 2/6 +3/6 +4/6 + 5/6 + 6/6 = = 21/6 = 3.5$ $E(x) = 3.5$ → de media se avanza 3.5 casillas cada vez que se tira el dado.

100 casillas / 3.5 casillas cada tirada = 28.6 tiradas de media para llegar a la meta.
Si una persona compra una papeleta en una rifa, en la que puede ganar de 5.000 € ó un segundo premio de 2000 € con probabilidades de: 0.001 y 0.003. ¿Cuál sería el precio justo a pagar por la papeleta?

$E(x) = 5000 \cdot 0.001 + 2000 \cdot 0.003 = 11$ €
Un jugador lanza dos monedas. Gana 1 ó 2 € si aparecen una o dos caras. Por otra parte pierde 5 € si no aparece cara. Determinar la esperanza matemática del juego y si éste es favorable.

$E = \{(c,c);(c,x);(x,c);(x,x)\}$
$p(+1) = 2/4$
$p(+2) = 1/4$
$p(-5) = 1/4$
$E(x)= 1 \cdot 2/4 + 2 \cdot 1/4 - 5 \cdot 1/4 = -1/4$.

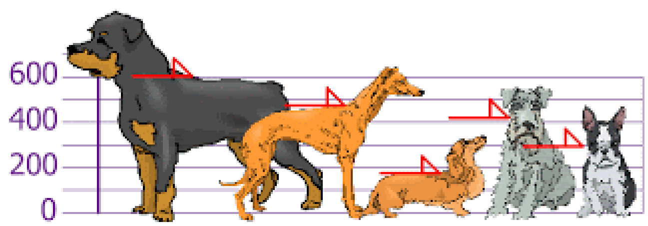
Es desfavorable
Desviación estándar
4. La desviación estándar (σ) mide cuánto se separan los datos.
La fórmula es fácil: es la raíz cuadrada de la varianza. Así que, "¿qué es la varianza?"

<u>Varianza</u>
La varianza (que es el cuadrado de la desviación estándar: σ^2) se define así:
Es la media de las diferencias con la media elevadas al cuadrado.
En otras palabras, sigue estos pasos:

Calcula la media (el promedio de los números)
Ahora, por cada número resta la media y eleva el resultado al cuadrado (la diferencia elevada al cuadrado).

Ahora calcula la media de esas diferencias al cuadrado.

EJEMPLO
Tú y tus amigos han medido las alturas de sus perros (en milímetros):

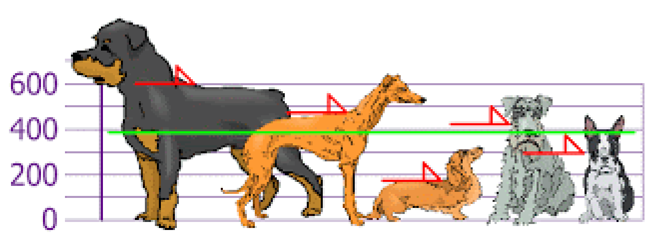


Las alturas (de los hombros) son: 600mm, 470mm, 170mm, 430mm y 300mm.Calcula la media, la varianza y la desviación estándar.

Respuesta:

$$\text{Media} = \frac{600 + 470 + 170 + 430 + 300}{5} = \frac{1970}{5} = 394$$

así que la altura media es 394 mm. Vamos a dibujar esto en el gráfico:



Para calcular la varianza, toma cada diferencia, elévala al cuadrado, y haz la media:
Varianza: σ^2 = =

$$\frac{206^2 + 76^2 + (-224)^2 + 36^2 + (-94)^2}{5} \qquad \frac{108{,}520}{5} = 21{,}704$$

Así que la varianza es 21,704.

Y la desviación estándar es la raíz de la varianza, así que:

Desviación estándar: **σ = √21,704 = 147**

Propiedades de la varianza

Si X es una variable aleatoria con función de probabilidad o densidad f(x), la varianza de una función de la variable X , m(x) , se calcula según la expresión:

$$\sigma^2_{m(x)} = E\left[\left(m(X) - \mu_{m(X)}\right)^2\right] = E\left[\left(m(X) - E[m(X)]\right)^2\right]$$

Casos concretos:

1. Cuando a todos los valores de una variable se les suma una constante, la varianza de la variable conserva el mismo valor (ver imagen en las propiedades de la media)

$$\forall b \in R \quad m(X) = X + b \Rightarrow \sigma^2(X + b) = \sigma^2_X$$

2. Cuando a todos los valores de una variable se les multiplica por una constante, la varianza de la variable queda multiplicada por el valor de la constante elevado al cuadrado (ver imagen en las propiedades de la media)

$$\forall a \in R \quad m(X) = a \cdot X \;\Rightarrow\; \sigma^2(a \cdot X) = a^2 \cdot \sigma^2_X$$

3. Si X e Y son dos variables aleatorias con función de densidad o probabilidad conjunta f(x,y), la varianza de la función m(x,y) = a X ± b Y, donde a y b son constantes reales se calcula como:

$$\sigma^2(aX \pm bY) = a^2 \cdot \sigma^2_X + b^2 \cdot \sigma^2_Y \pm 2 \cdot a \cdot b \cdot \sigma_{XY}$$

En el caso de que a = b = 1

$$\sigma^2(X \pm Y) = \sigma^2_X + \sigma^2_Y \pm 2 \cdot \sigma_{XY}$$

Si además ocurre que X e Y sean independientes $\sigma_{xy} = 0$, luego

$$X\,Y \text{ independientes } \Rightarrow \sigma^2(X \pm Y) = \sigma^2(X) + \sigma^2(Y)$$

TAREA:

Sea X una variable aleatoria discreta cuya función de probabilidad es:

X	p_i
0	0,1
1	0,2
2	0,1
3	0,4
4	0,1
5	0,1

1. Calcular, representar gráficamente la función de distribución
2. Calcular las siguientes probabilidades:

$p(X < 4.5)$

$p(X \geq 3)$

$p(3 \leq X < 4.5)$ Solución

3. Sabiendo que $p(X \leq 2) = 0.7$ y $p(X \geq 2) = 0.75$. Hallar la esperanza matemática, la varianza y la desviación típica Solución
4. Un jugador lanza dos monedas. Gana 1 ó 2 € si aparecen una o dos caras. Por otra parte pierde 5 € si no aparece cara. Determinar la esperanza matemática del juego y si éste es favorable Solución
5. Se lanza un par de dados. Se define la variable aleatoria X como la suma de las puntuaciones obtenidas. Hallar la función de probabilidad, la esperanza matemática y la varianza Solución
6. Un jugador lanza un dado corriente. Si sale 1 o número primo, gana tantos cientos de euros como marca el dado, pero si no sale número primo, pierde tantos cientos de euros como marca el dado. Determinar la función de probabilidad y la esperanza matemática del juegoSolución
7. Si una persona compra una papeleta en una rifa, en la que puede ganar de 5.000 € ó un segundo premio de 2000 € con probabilidades de: 0.001 y 0.003. ¿Cuál sería el precio justo a pagar por la papeleta Solución

3. TIPOS DE DISTRIBUCIONES, VARIABLES ALEATORIAS DISCRETAS Y CONTINUAS.

Distribución de Probabilidades

Una distribución de probabilidad indica toda la gama de valores que pueden representarse como resultado de un experimento si éste se llevase a cabo.

Es decir, describe la probabilidad de que un evento se realice en el futuro, constituye una herramienta fundamental para la prospectiva, puesto que se puede diseñar un escenario de acontecimientos futuros considerando las tendencias actuales de diversos fenómenos naturales

Toda distribución de probabilidad es generada por una variable (porque puede tomar diferentes valores) aleatoria x (porque el valor tomado es totalmente al azar), y puede ser de dos tipos:

Variables Aleatorias

Una variable aleatoria es un conjunto o subconjunto de datos agrupados para poder obtener datos tales como la media de la modo en una estadística de un muestreo que funciona con una regla de correspondencia, función que asigna un único número real a cada resultado de un espacio maestral en un experimento. Cuantifica los resultados de un experimento aleatorio y también se le llama variable de azar o variable estocástica, que significa cantidad que puede tomar varios valores imprevistos.

Dado un experimento aleatorio cualquiera cuyos sucesos elementales posibles pueden identificarse fácilmente mediante un número real, se denomina Variable Aleatoria, X, al conjunto de estos números.

Ejemplo.- Sea el experimento aleatorio de averiguar la marca de tabaco que preferirá un individuo entre las posibles marcas: X, Y, Z.

En este caso la asociación de un número para cada suceso elemental posible del experimento no es inmediata. En consecuencia, se establece una correspondencia entre el conjunto de los sucesos elementales posibles y el conjunto de los números reales, del modo siguiente:

Al suceso elemental preferir la marca X se le hace corresponder el número 1; al suceso elemental preferir la marca Y se le hace corresponder el número 2; al suceso elemental preferir la marca Z se le hace corresponder el número 3.

La variable aleatoria X será: X = (1,2,3).

El número asociado a cada suceso elemental puede ser cualquiera dentro del conjunto de los números reales, con la condición única de que a sucesos elementales distintos le correspondan números también distintos. Se comprueba fácilmente que la correspondencia así definida entre el conjunto de los posibles sucesos elementales de un experimento aleatorio y el conjunto de los números reales es una aplicación inyectiva.

Esperanza Matemática o Valor Esperado

El valor esperado de una Variable Aleatoria X es el promedio ponderado de todos los valores posibles de la misma. No de los pesos son las probabilidades asociadas con los valores.

Para calcular el valor esperado de una variable aleatoria por su correspondiente probabilidad y luego sumar los términos resultantes.

La esperanza matemática o valor esperado de una variable aleatoria tiene sus orígenes en los juegos de azar, debido a que los apostadores deseaban saber cuál era su esperanza de ganar repetidamente un juego, por lo tanto, el valor esperado representa la cantidad de dinero promedio que el jugador está dispuesto a ganar o perder después de un número grande de apuestas.

$E(x) = \mu = E\ xf\ (x)$

Varianza

Es un promedio ponderado de las de las desviaciones al cuadrado.

Varianza $= E\ (x - \mu)^2 f\ (x)$

Clasificación de las Variables Aleatorias
Las variables aleatorias pueden ser continuas o discontinuas. En este último caso se denomina también discretas.

<u>Variable Aleatoria Continua</u>
Si X es una Variable aleatoria continua, puede tomar cualquier valor de un intervalo continuo o dentro de un campo de variación dado. Las probabilidades de que ocurra un valor dado x están dadas por una función de densidad de probabilidad de que X quede entre a y b. El área total bajo la curva es 1.

Porque puede tomar tanto valores enteros como fraccionarios y un número infinito de ellos dentro de un mismo intervalo.

Por ejemplo:

x es la Variable que nos define la concentración en gramos de plata de algunas muestras de mineral (14.8 gr, 12.1, 10.0, 42.3, 15.0, 18.4, 19.0, 21.0, 20.8, …, n)

Propiedades de una variable aleatoria continua

p(x) Las probabilidades asociadas a cada uno de los valores que toma x deben ser mayores o iguales a cero. Dicho de otra forma, la función de densidad de probabilidad deberá tomar solo valores mayores o iguales a cero.

El área definida bajo la función de densidad de probabilidad deberá ser de 1.

Ejemplo.- Sea el experimento aleatorio consistente en medir la altura que es capaz de saltar cada miembro de un conjunto de personas. En este experimento, cada miembro del conjunto observado da lugar a un número, por lo que se toma como variable aleatoria el conjunto de las medidas de las alturas que son capaces de saltar las distintas personas.

En el supuesto que una persona hubiera saltado 105 cm y otra 106 cm, no existiría ninguna razón para que otra no hubiera saltado un valor intermedio cualquiera entre las dos anteriores, como 105.5 cm. Se trata de una variable aleatoria continua.

Distribución de Variable Continua

Se denomina variable continua a aquella que puede tomar cualquiera de los infinitos valores existentes dentro de un intervalo. En el caso de variable continua la distribución de probabilidad es la integral

La función de probabilidad de una variable aleatoria es la función que asigna a los valores que toma la variable aleatoria su correspondiente probabilidad $P(X = k)$, mientras que la función de distribución $F(X)$ es la probabilidad $P(X \leq k)$. En este tema vamos a trabajar con la distribución de probabilidad Binomial que describe experimentos que cumplen una serie de condiciones: • El modelo de distribución Binomial define experimentos consistentes en realizar ensayos repetidos e independientes. Cada uno de estos experimentos presenta dos posibles resultados que denominamos éxito o fracaso, cuya probabilidad se mantiene constante en las diferentes pruebas. • La variable se define como X "nº de veces que ocurre el suceso A en n experimentos", y viene determinada por dos parámetros: • n = tamaño muestral, número de experimentos realizados.

• $p = P[A]$ = probabilidad de que tenga lugar el suceso A.

• La función de probabilidad de la distribución B(n, p) es:

P[X = x] = x n ·px·(1 – p)(n – x) En el caso de la binomial, la variable toma valores comprendidos entre cero y n. Los parámetros que caracterizan la distribución binomial son:) Media o esperanza matemática: X = E[X] = ∑ · n ii xp 1 Donde xi son los posibles valores que toma la variable aleatoria y pi la probabilidad asociada a cada uno de ellos.

3.1 BINOMIAL

Una Distribución Binomial es una distribución discreta que modela el número de eventos en un número de ensayos fijo. Cada ensayo tiene dos resultados posibles y evento es el resultado de interés en un ensayo.

La Distribución Binomial se utiliza para describir un proceso donde los resultados se pueden etiquetar como un evento o un no evento y cuando esté interesado en la ocurrencia de un evento y no en su magnitud. La Distribución Binomial se usa frecuentemente en control de calidad, sondeos de opinión pública, investigaciones médicas y seguros.

El número de eventos (X) en n ensayos sigue una distribución binomial si se cumplen las siguientes condiciones:

- El número de ensayos es fijo.
- Cada ensayo es independiente de otros ensayos.
- Cada ensayo tiene uno de dos resultados: evento o no evento.
- La probabilidad de un evento es igual para cada ensayo.

En estadística, la distribución binomial es una distribución de probabilidad discreta que cuenta el número de éxitos en una secuencia de n ensayos de Bernoulli independientes entre sí, con una probabilidad fija p de ocurrencia del éxito entre los ensayos. Un experimento de Bernoulli se caracteriza por ser dicotómico, esto es, solo dos resultados son posibles. A uno de estos se denomina «éxito» y tiene una probabilidad de ocurrencia p y al otro, «fracaso», con una probabilidad[$q=1-p$. En la distribución binomial el anterior experimento se repite n veces, de forma independiente, y se trata de calcular la probabilidad de un determinado número de éxitos. Para $n = 1$, la binomial se convierte, de hecho, en una distribución de Bernoulli.

Para representar que una variable aleatoria X sigue una distribución binomial de parámetros n y p, se escribe:

X ~ B (n , p) {\displaystyle X\sim B(n,p)\,}

La distribución binomial es la base del test binomial de significación estadística.

Las siguientes situaciones son ejemplos de experimentos que pueden modernizarse por esta distribución:

- Se lanza un dado diez veces y se cuenta el número *X* de tres obtenidos: entonces $X \sim B(10, 1/6)$

Una distribución se denomina binomial cuando se cumplen las condiciones siguientes:

1. El experimento aleatorio de base se repite n veces, y todos los resultados obtenidos son mutuamente independientes.

2. En cada prueba se tiene una misma probabilidad de éxito (suceso A), expresada por p. Asimismo, existe en cada prueba una misma probabilidad de fracaso (suceso), que es igual a q = 1 - p.

3. El objetivo de la distribución binomial es conocer la probabilidad de que se produzca un cierto número.

Existen muchas situaciones en las que se presenta una experiencia binomial. Cada uno de los experimentos es independiente de los restantes (la probabilidad del resultado de un experimento no depende del resultado del resto). El resultado de cada experimento ha de admitir sólo dos categorías (a las que se denomina éxito y fracaso). El valor de ambas posibilidades ha de ser constante en todos los experimentos, y se denotan como p y q respectivamente, o p y 1-p de forma alternativa.

Se designa por X a la variable que mide el número de éxitos que se han producido en los "n" experimentos.

Cuando se dan estas circunstancias, se dice que la variable X sigue una distribución de probabilidad binomial, y se denota B(n, p).

De distribución, siempre se esperan dos tipos de resultados

Las características de esta Distribución Binomial son:

a). En los experimentos que tienen este tipo de distribución, siempre se esperan dos tipos de resultados, ejemplo: Defectuoso, no defectuoso, pasa, no pasa, etc., denominados arbitrariamente “éxito” (que es lo que se espera que ocurra), o fracaso (lo contrario del éxito).

b). Las probabilidades asociadas a cada uno de estos resultados son constantes, es decir no cambian.

c). Cada uno de los ensayos o repeticiones del experimento son independientes entre sí.

d). El número de ensayos o repeticiones del experimento (n) es constante.

Si “p” es la probabilidad de que ocurra un suceso en un solo intento (llamada probabilidad de éxito) y $q = 1 – p$, es la probabilidad de que no ocurra en un solo intento (llamada probabilidad de fracaso), entonces la probabilidad de que el suceso ocurra exactamente “x” veces en N intentos (0 sea, x éxitos y N - X Fracasos), viene dada por:

$$P(x) = \frac{N}{X} p^{x} q^{N-X} = \frac{N!}{X!\,(N-X)!} p^{x} q^{N-X}$$

Donde:
N = número de ensayos/experimentos
X = número de éxitos
P = probabilidad de éxito en cualquier prueba
q = probabilidad de fracaso (1-p)
N-X = número de fracasos y se trata de determinar p(x).

El arreglo más simple de los de este tipo es x éxitos seguidos de N-X fracasos. La probabilidad de este resultado, debido a la condición de independencia en cada ensayo, es $p^{x} q^{N-X}$. Cualquier otro arreglo con x éxitos y N-X fracasos tendrá la misma probabilidad y lo que queda entonces es determinar el número de estos arreglos; para esto se puede considerar que se trata de permutaciones de N elementos, donde uno de los elementos se repite x veces y el otro N-X veces, por lo que se tienen entonces de acuerdo con la fórmula de permutaciones con repetición, $\frac{N!}{X!\,(N-X)!}$ arreglos.

Ejemplos:

1. La probabilidad de obtener exactamente 2 caras en 6 tiradas de una moneda es:

Solución:
N = 6
X = 2
P = q = 1

$$P(x) = \frac{N}{X} p^{x} q^{N-X} = \frac{N!}{X!\,(N-X)!} p^{x} q^{N-X}$$

2

$P(x) = \binom{6}{2} \times \left(\frac{1}{2}\right)^{2} \times \left(\frac{1}{2}\right)^{6-2} = \frac{6!}{2!(6-2)!} \times \left(\frac{1}{2}\right)^{2+6-2}$ $\quad p(2) = \frac{6!}{2!\,4!} \times \left(\frac{1}{2}\right)^{6}$ $\quad p(2) = \frac{6x5x4x3x2!}{2!\ \ 4x3x2x1!} \times \left(\frac{1}{2}\right)^{6}$

3 5 1 1

$P(2) = \frac{6x5x4x3!}{4x3!x2x1!} \times \left(\frac{1}{2}\right)^{6}$ $\quad P(2) = 3x5x1x1 \times \frac{1}{64}$ $\quad p(2) = \frac{15}{64}$

2. La probabilidad de obtener al menos 4 caras en 6 tiradas de una moneda:

Solución:

a).	b).	c).
N = 6	N = 6	N = 6
X = 4	X = 5	X = 6
P y q = $\frac{1}{2}$	p y q = $\frac{1}{2}$	p y q = $\frac{1}{2}$

$P(x) = \binom{N}{X} p^{x} q^{N-X} = \frac{N!}{X!\,(N-X)!} p^{x} q^{N-X}$

a),

$P(x) = \binom{6}{4} p^{4} q^{6-4} = \frac{6!}{4!\,(6-4)!} \times \left(\frac{1}{2}\right)^{4} \times \left(\frac{1}{2}\right)^{6-4}$ $\quad p(4) = \frac{6x5x4!}{4!\ \ 2!} \times \left(\frac{1}{2}\right)^{4+6-4}$ $\quad p(4) = \frac{6x5x4!}{4!\ \ 2x1!} \times \left(\frac{1}{2}\right)^{6}$

3 5

$P(4) = \frac{6x5!}{2x1!} \times \left(\frac{1}{2}\right)^{6} -=$ $\quad p(4) = 3x5 = \left(\frac{1}{2}\right)^{6}$ $\quad p(4) = 15 \times \left(\frac{1}{2}\right)^{6} -$ $\quad p(4) = \frac{15}{64}$

b).

$P(x) = \binom{N}{X} p^{x} q^{N-X} = \frac{N!}{X!\,(N-X)!} p^{x} q^{N-X}$

N = 6

X = 5

Pyq = $\frac{1}{2}$

$P(x) = \binom{6}{5} p^{5} q^{6-5} = \frac{6!}{5!\,(6-5)!} \times \left(\frac{1}{2}\right)^{5} \times \left(\frac{1}{2}\right)^{6-5}$ $\quad p(5) = \frac{6!}{5!\ 1!} \times \left(\frac{1}{2}\right)^{5+6-5}$ $\quad p(5) = \frac{6x5!}{5!\ 1!} \times \left(\frac{1}{2}\right)^{6}$

$P(5) = \frac{6}{1} \times \left(\frac{1}{2}\right)^{6}$ $\quad p(5) = \frac{6}{64}$ $\quad p(5) = \frac{6}{64}$

c).

$P(x) = \binom{N}{X} p^{x} q^{N-X} = \frac{N!}{X!\,(N-X)!} p^{x} q^{N-X}$

N = 6

X = 6

Pyq = $\frac{1}{2}$

$P(x) = \binom{6}{6} p^{6} q^{6-6} = \frac{6!}{6!\,(6-6)!} \times \left(\frac{1}{2}\right)^{6} \times \left(\frac{1}{2}\right)^{6-6}$ $\quad p(6) = \frac{6!}{6!\ 0!} \times \left(\frac{1}{2}\right)^{6+6-6}$ $\quad p(6) = \frac{1!}{1!} \times \left(\frac{1}{2}\right)^{6}$

$P(6) = 1 \times \underline{1^{6}}$ $\quad p(6) = \underline{1}$ $\quad p(5) = \underline{1}$

2 64 64

$$P(x) = \frac{15}{64} + \frac{6}{64} + \frac{1}{64} = \frac{22}{64} = \frac{11}{32} \qquad p(x) = \frac{11}{32}$$

3. Imaginemos que un 80% de personas en el mundo han visto el partido de la final del último mundial de futbol. Tras el evento, 4 amigos se reúnen a conversar. ¿Cuál es la probabilidad de que 3 de ellos hayan visto.

Solucion:

N = 4 es el total de la muestra que buscamos

X = número de éxitos, que en este caso es igual a 3, dado que buscamos la probabilidad de que 3 de los 4 amigos lo hayan visto,

p = probabilidad de éxito (0.8)

q = probabilidad de fracaso (0.2). Este resultado se obtiene al restar 1-p

$$P(x) = \frac{N}{X}\ p^{x} q^{N-X} = \frac{N!}{X!\,(N-X)!} p^{x} q^{N-X}$$

$$P(x) = \frac{4!}{3!} \times 0.8^3 x 0.2^{4-3} \qquad P= \frac{4!}{3!(4-3)!} \times 0.8^3 x 0.2^{4-3} \qquad P(3) = \frac{4!}{!\ 1!} x\ 0,8^3 x 0.2^{\ 4-3}$$

$$P(3) = \frac{4x3!}{3!\ 1!} x 0.8^3 x 0.2^1 \qquad P(3) = \frac{4!}{1!} \times 0.512x0.2 \qquad P(3) = 4! \times 0.512x0.2$$

P(3) = 4 x 0.1024 = 0.4096 P(3) = 0.4096

4. La contaminación constituye un problema en la fabricación de discos de almacenamiento óptico. El número de partículas de contaminación que ocurre en un disco óptico tiene una distribución de Poisson y el número promedio de partículas por centímetro cuadrado de superficie del disco es 0.1. El área de un disco bajo estudio es 100 centímetros cuadrados.

a). Encuentre la probabilidad de que ocurran 12 partículas en el área del disco bajo estudio.

b). La probabilidad de que ocurran cero partículas en el área del disco bajo estudio.

c). Determine la probabilidad de que 12 o menos partículas ocurran en el área del disco bajo estudio.

Solución:

Sea que x denote el número de partículas en el área de un disco bajo estudio.

Puesto que el número promedio de partículas es 0.1 partículas por cm^2 .

E(x)=100 cm^2 x0.1 partículas/ cm^2 = 10 partículas

Por lo tanto:

a). $P(x=12)=\frac{(10)^{12}(2.7182)^{-10}}{12!} = \frac{(1)^{12}(4.5413)^{-5}}{12x11x10x\ldots x1!} = \frac{45´413,598.77}{479´001,600.00} = 0.09480$

b). La probabilidad de que ocurran cero partículas en el área del disco bajo estudio es: $P(x=0)= \underline{(10)^0(2,7182)^{-10}} = \underline{(1)\ (4.5413)} = 4.5413$

0! 1

(c) Determine la probabilidad de que 12 o menos partículas ocurran en el área del disco bajo estudio. La probabilidad es

$$P(X \leq 12)=P(x=0)+P(x=1)+\ldots+P(x=12)=\sum_{i=1}^{n} = \frac{(10)^{12}\,(2.7182)^{-10}}{12!} = \frac{(1)(4,5413)^{-5}}{479´001,600}$$

p(x=12) = 0.0948

3.1.1 PROPIEDADES: MEDIA, VARIANZA Y DESVIACIÓN ESTÁNDAR

Se lanza una moneda al aire 1000 veces con probabilidad de obtener una cara de 0.5, podríamos esperar 1000(0.5) = 500 caras como promedio o valor esperado.

La media de la distribución binomial es $\mu = n\,p$ sólo para distribución binomial. La La varianza de la distribución binomial es $\mu^2 = n\,p$ y la desviación estándar $\sigma = \sqrt{npq}$. Las medidas de variación más comunes son la varianza y su raíz cuadrada o sea, la desviación estándar, la cual mide la variabilidad promedio de las variaciones cuadráticas de la media.

Ejemplos:

1. ¿Cuál es la media y la desviación estándar de las caras en 400 lanzamientos de una moneda balanceada?

Solución:

$\mu = n\,p$ $\mu = 400(0.5) = 200$

$\sigma = \sqrt{npq}$ $\sigma^2 = 400(0.5)\,(0.5) = 100$ y $\sigma = \sqrt{100}$ $\sigma = 10$

2. Número de caras en 484 lanzamientos de una moneda

Solución:

$\mu = n\,p$ $\mu = 484(0.5) = 242$

$\sigma = \sqrt{npq}$ $\sigma^2 = 484(0.5)\,(0.5) = 121$ $\sigma = \sqrt{121}$ $\sigma = 11$

3. El nutriera de 3's obtenido en 720 lanzamientos de un dado

Solución:

$\mu = n\,p$ $\mu = 720(1/6) = 120$

$\sigma = \sqrt{npq}$ $\sigma^2 = 120(1/6)\,(1/6) = 20$ $\sigma = \sqrt{20}$ $\sigma = 4.47$

4. 600 personas invitadas a una fiesta con probabilidad dé asistir del 0.30

Solución:

$\mu = n\,p$ $\mu = 600(0.30) = 180$

$\sigma = \sqrt{npq}$ $\sigma^2 = 600(0.30)\,(0.70) = 126$ $\sigma = \sqrt{126}$ $\sigma = 11.22$

5. Artículos defectuosos en una muestra de n = 600 partes dé maquinaria con probabilidad de defecto de 0.04

Solución:

$\mu = n\,p$ $\mu = 600(0.04) = 24$

$\sigma = \sqrt{npq} = 600(0\text{-}04)\,(0.96) = 23.04$ $\sigma = \sqrt{23.04}$ $\sigma = 4.8$

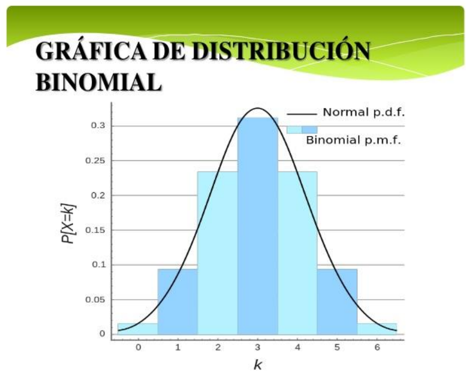
6. 800 estudiantes que no les gusta la comida que sirven en la cafetería de la universidad con el 0.65 de probabilidad de no gustarles.

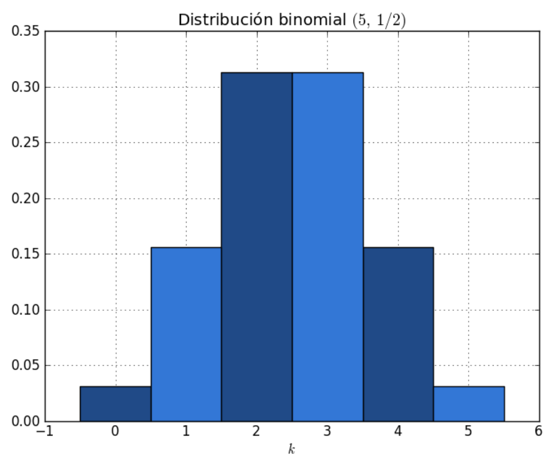
Solución:

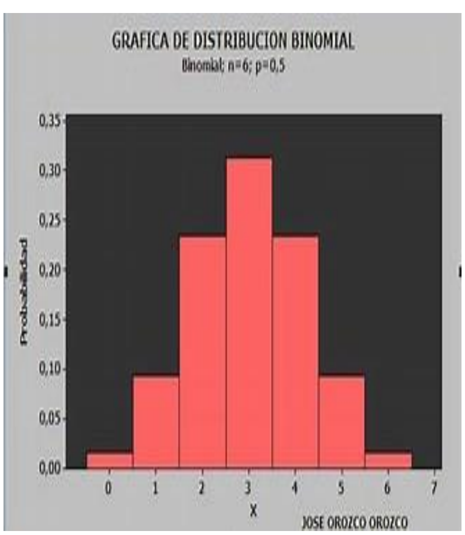
$\mu = n\,p$ $\mu = 800(0.65) = 520$

$\sigma = \sqrt{npq}$ $\sigma^2 = 800(0.65)\,(0.35) = 182$ $\sigma = \sqrt{182}$ $\sigma = 13.49$

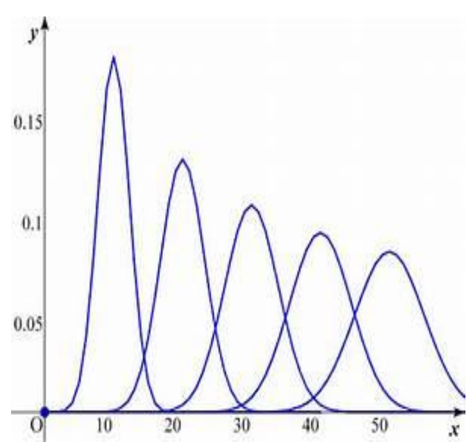
3.1.2 GRAFICA







Grafica Binomial



3.2 POISSON

En teoría de probabilidad y estadística, la distribución de Poisson es una distribución de probabilidad discreta que expresa, a partir de una frecuencia de ocurrencia media, la probabilidad de que ocurra un determinado número de eventos durante cierto período de tiempo. Concretamente, se especializa en la probabilidad de ocurrencia de sucesos con probabilidades muy pequeñas, o sucesos "raros".

La función de densidad de probabilidad de la distribución de Poisson es

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

donde:

k es el número de ocurrencias del evento o fenómeno (la función nos da la probabilidad de que el evento suceda precisamente *k* veces).

λ es un parámetro positivo que representa el número de veces que se espera que ocurra el fenómeno durante un intervalo dado. Por ejemplo, si el suceso estudiado tiene lugar en promedio 4 veces por minuto y estamos interesados en la probabilidad de que ocurra *k* veces dentro de un intervalo de 10 minutos, usaremos un modelo de distribución de Poisson con $\lambda = 10\times 4 = 40$. *e* es la base de los logaritmos naturales (e = 2.71828...)

Esta distribución es una de las más importantes distribuciones de variable discreta. Sus principales aplicaciones hacen referencia a la modelización de situaciones en las que nos interesa determinar el número de hechos de cierto tipo que se pueden producir en un intervalo de tiempo o de espacio, bajo presupuestos de aleatoriedad y ciertas circunstancias restrictivas. Otro de sus usos frecuentes es la consideración límite de procesos dicotómicos reiterados un gran número de veces si la probabilidad de obtener un éxito es muy pequeña.

Esta distribución de Poisson. Es una de las más importantes distribuciones de variable discreta. Sus principales aplicaciones hacen referencia a la modelización de situaciones en las que nos interesa determinar el número de hechos de cierto tipo que se pueden producir en un intervalo de tiempo o de espacio, bajo presupuestos de aleatoriedad y ciertas circunstancias restrictivas. Otro de sus usos frecuentes es la consideración límite de procesos dicotómicos reiterados un gran número de veces si la probabilidad de obtener un éxito es muy pequeña.

Esta distribución se puede hacer derivar de un proceso experimental de observación en el que tengamos las siguientes características:

En este tipo de experimentos los éxitos buscados son expresados por unidad de área, tiempo, pieza, etc, etc,:

-Número de defectos de una tela por m^2
-Número de aviones que aterrizan en un aeropuerto por día, hora, minuto, etc, etc.
-Número de bacterias por cm^2 de cultivo
-Número de llamadas telefónicas a un conmutador por hora, minuto, etc, etc.
-Número de llegadas de embarcaciones a un puerto por día, mes, etc, etc.

Para determinar la probabilidad de que ocurran x éxitos por unidad de tiempo, área, o producto, la fórmula a utilizar sería: $P(X, \lambda) = \dfrac{\lambda^x e^{-\lambda}}{X!}$

donde:

p(x, λ) = probabilidad de que ocurran x éxitos, cuando el número promedio de ocurrencia de ellos es λ.

λ = media o promedio de éxitos por unidad de tiempo, área o producto

e=2.718

x = variable que nos denota el número de éxitos que se desea que ocurra

Hay que hacer notar que en esta distribución el número de éxitos que ocurren por unidad de tiempo, área o producto es totalmente al azar y que cada intervalo de tiempo es independiente de otro intervalo dado, así como cada área es independiente de otra área dada y cada producto es independiente de otro producto dado.

Ejemplos:

1. Si un banco recibe en promedio 6 cheques sin fondo por día, ¿cuáles son las probabilidades de que reciba, a) cuatro cheques sin fondo en un día dado, b) 10 cheques sin fondos en cualquiera de dos días consecutivos?

   a). x = variable que nos define el número de cheques sin fondo que llegan al banco en un día cualquiera = 0, 1, 2, 3, ..., etc, etc.
   λ = 6 cheques sin fondo por día
   e = 2.718

Solución: $P(X{=}X) = \dfrac{\lambda^X e^{-\lambda}}{X!}$

$$P(X{=}4, \lambda{=}6) = \frac{(6)^4(2{,}718)^{-6}}{4!} = \frac{(1{,}296)(0.002480294702)}{4x3x2x1!} = \frac{3.21461934}{24} = 0.133935$$

   b). x = variable que nos define el número de cheques sin fondo que llegan al banco en dos días consecutivos = 0, 1, 2, 3, ..., etc., etc.

   λ = 6 x 2 = 12 cheques sin fondo en promedio que llegan al banco en dos días consecutivos

   Nota: l siempre debe de estar en función de x siempre o dicho de otra forma, debe "hablar" de lo mismo que X.

Solución: $P(X, \lambda) = \dfrac{\lambda^x e^{-\lambda}}{X!}$

$$p(X{=}10, \lambda{=}12) = \frac{(12)^{10}(2.718)^{-12}}{10!} = \frac{(6.191736)^{10}(6.151861809)^{-0.6}}{10x9x8x7x6x5x4x3x2x1!} = \frac{380{,}907.0683}{3´628{,}800.00}$$

   $p(x{=}10, \lambda{=}12) = 0.104967778$

2. En la inspección de hojalata producida por un proceso electrolítico continuo, se identifican 0.2 imperfecciones en promedio por minuto. Determine las

probabilidades de identificar a) una imperfección en 3 minutos, b) al menos dos imperfecciones en 5 minutos, c) cuando más una imperfección en 15 minutos.

a). x = variable que nos define el número de imperfecciones en la hojalata por cada 3 minutos = 0,1, 2, 3,.., etc., etc. λ = 0.2 x 3 =0.6 imperfecciones en promedio por cada 3 minutos en la hojalata.

Solución: $$P(X, \lambda) = \frac{\lambda^{x} e^{-\lambda}}{X!}$$

$$p(x=1,\lambda=0.6) = \frac{(0.6)^{1}(2.718)^{-0.6}}{1!} = \frac{(0.6)(0.548845)}{1} = 0.329307$$

b). x = variable que nos define el número de imperfecciones en la hojalata por cada 5 minutos = 0, 1, 2, 3, ..., etc., etc.

λ = 0.2 x 5 =1 imperfección en promedio por cada 5 minutos en la hojalata.

Solución:

$$p(x=2, 3, 4, \text{etc}\ldots \lambda=1) = 1- p(x=0.1\ \lambda= 1) = \frac{1-((1)^{0}(2.718)^{-1}}{0!} + \frac{(1)(2,718)^{-1}}{1}$$

p(x = 0,1, λ = 1) =1-(0.2718+0.2718) = 1 (-0.2718 – 0.2718) =

p(x=0.1, λ =1) = (-0.2718- 0.2718) = -0.5436   p(x=0.1, λ= 1)= 0.5436

c). x = variable que nos define el número de imperfecciones en la hojalata por cada 15 minutos = 0, 1, 2, 3, ..., etc., etc.

λ = 0.2 x 15 = 3 imperfecciones en promedio por cada 15 minutos en la hojalata.

Solución:

$$p(x=0.1,\ \lambda= 3)= p(x=0, \lambda=3) + p(x=1, \lambda=3)=\frac{(3)^{0}(2.718)^{-3}}{0!} + \frac{(3)^{1}(2,718)^{-3}}{1!}$$

p(x = 0,1, λ = 3) = 0.149407671 + 0.149407671

p(x= 0.1, λ=3) = 0.298815343

3. Si ya se conoce que solo el 3% de los alumnos de Contabilidad son muy inteligentes

¿Calcular la probabilidad de que si tomamos 100 alumnos al azar 5 de ellos sean muy inteligentes. n = 100P = 0.03lambda = 100 * 0.03 = 3. Si un banco recibe en promedio

6 cheques sin fondo por día, ¿cuáles son las probabilidades de que reciba.

a). cuatro cheques sin fondo en un día dado,

b). 10 cheques sin fondos en cualquiera de dos días consecutivos?

Solución:

a). x = variable que nos define el número de cheques sin fondo que llegan al banco en un día cualquiera = 0, 1, 2, 3, etc. λ = 6 cheques sin fondo por días

Solución: $$P(X, \lambda) = \frac{\lambda^{x} e^{-\lambda}}{X!}$$

$p= (x=4,\ \lambda=6)=\frac{(6)^4\,(2.718)^{-6}}{4!}=\frac{(1,296)(0.00248)}{4x3x2x1!}=\frac{3.214461934^{-3}}{24}=0.1339359^{-4}$

b). x= variable que nos define el número de cheques sin fondo que llegan al banco en dos días consecutivos = 0, 1, 2, 3, ..., etc., etc.

λ = 6 x 2 = 12 cheques sin fondo en promedio que llegan al banco en dos días consecutivos.

Nota:

λ siempre debe de estar en función de x siempre o dicho de otra forma, debe "hablar" de lo mismo que x.

Solución: $P(X, \lambda) = \frac{\lambda^x\, e^{-\lambda}}{X!}$

Datos: x=10 λ=12 e=2.7182

$p = (x=10,\ \lambda=12) = \frac{(12)^{10}(2.718)^{12}}{10!} = \frac{(6.191736423)^{10}(162´552.416)}{10x9x8x7x6x5x4x3x2x1!} =$

$p = \frac{1.006481715^{16}}{3´628,800.00} = 2.773593791^{10}$

4. Ciertos autos llegan a una garita de peaje a razón de 6 autos por hora, calcular:
   a). La probabilidad de que lleguen 5 autos en una hora.
   b). La probabilidad de que llegue a lo más 2 autos en 30 minutos.
   a). La variable X, es el número de autos que llega a una garita de peaje por hora.

Solución: $P(X, \lambda) = \frac{\lambda^x\, e^{-\lambda}}{X!}$

Datos:

λ = 6 autos por hora

e=2.7182

6 – 60 min.

λ´ - 30 mi

$\lambda´= \frac{6x30}{60} = \frac{180}{60} = 3$

$p(x=5) = \frac{(6)^5(2.718)^{-6}}{5!} = \frac{(7,776)(2.47919)^{-3}}{5x4x3x2x1!} = \frac{19.2782}{120} = 0.1606$ $p(x=5) = 0.0106$

b). La variable x, es el número de autos que llegue a lo más 2 autos en 30 min.

Solución: $P(X, \lambda) = \frac{\lambda^x\, e^{-\lambda}}{X!}$

Datos:

p(x menor igual) en 30 minutos.

e=2.7182

λ = 6 autos por hora

6 – 60 min.

λ´ - 30 min.

$\lambda´ = \frac{6x30}{60} = \frac{180}{60} = 3$

p(x=0) + p(x=1) + p(x=2)

$$p(x) = \frac{(3)^0(2.7182)^{-3}}{0!} + \frac{(3)^1(2.7182)^{-3}}{1!} + \frac{(3)^2(2.7182)^{-3}}{2!}$$

$$p(x) = \frac{(3)(0.0498)}{1} + \frac{(3)(0.0498)}{1} + \frac{(9)(0.0498)}{2x1!}$$

$p(x) = 0.0498 + 0.1494 + 0.2241 = 0.4233$ $\quad p(x) = 0.4233$

$p(x \leq l\ 2) = 0.4233$

5. En una tienda llega un cliente a cada minuto ¿Cuál es la probabilidad que lleguen 3 en próximos 5 minutos.

Solución: $P(X, \lambda) = \frac{\lambda^x e^{-\lambda}}{X!}$

Datos: x=0 λ= -5 e = 2.7182

$$p(3) = \frac{(5)^3 (2.7182)^{-5}}{3!} = \frac{(125)(6.73896)^{-5}}{3x2x1!} = \frac{0.842370}{6} = 0.14039$$

$p(x) = 0.14039 = 14.04\%$

7. Cada hora nacen 8 bebes en la ciudad ¿Cuál es la probabilidad que en los próximos 15 minutos no nazcan ningún bebe?

Solución: $P(X, \lambda) = \frac{\lambda^x e^{-\lambda}}{X!}$

Datos: x=0 λ=8/hora λ=2/15min. ,,,,,,,,,,,,,,,,,,,,,,, e = 2.7128

$$p(0) = \frac{(2)^0(2.7182)^{-2}}{0!} = \frac{(1)(0.13534)}{1} = 0.13534$$

8. A un conmutador de la oficina principal de la compañía llegan llamadas a un promedio de dos por minuto y se sabe que tiene la distribución de Poisson.
   Si el operador esta distraído por un minuto, ¿Cuál es la probabilidad de que el número de llamadas no respondidas sea:
   a) ¿cero?
   b) ¿por lo menos una?
   c) ¿Entre 3 y 5, inclusive?
   ¿Cuál es la probabilidad de que se pierdan cero llamadas?
   Representa el número de resultados que ocurren en un intervalo de tiempo dado o en un área o volumen especifico.

Solución: $P(X, \lambda) = \frac{\lambda^x e^{-\lambda}}{X!}$

a). Datos: x=0 λ=2 llamadas/min. e= 2.7182

$$p(x=0) = \frac{(2)^0(2.7182)^{-2}}{0!} = \frac{(1)(0.1353)}{1} = 0.1353 \quad p(x) = 13.53\%$$

b). Datos: (x mayor igual 1) λ=2 llamadas/min. e = 2,7182

$p(x=1) = \dfrac{(2)^1(2.7182)^{-2}}{1!} = \dfrac{(2)(0.1353)}{1} = 0.2706$

$p(x=2) = \dfrac{(2)^2(2.7182)^{-2}}{2!} = \dfrac{(4)(0.13534)}{2x1!} = \dfrac{0.54136}{2} = 0.27068$

$p(x=3) = \dfrac{(2)^3(2.7182)^{-2}}{3} = \dfrac{(8)(0.13534)}{3x2x1!} = \dfrac{1.08272}{6} = 0.18045$

$p(x \geq 1) = p(x=1) + p(x=2) + p(x=3)$

$p(x \geq 1) = 0.2706 + 0.27068 + 0.18274 = 0.72181$

$p(x) = 0.72181$

c). Datos: 3 mayor igual X menor igual 5 λ= 2 llamadas /min. e = 2.7182

$p(x=3) = \dfrac{(2)^3(2.7182)^{-2}}{3!} = \dfrac{(8)(0.13534)}{3x2x1!} = \dfrac{1.0827}{6} = 0.18045$

$p(x=4) = \dfrac{(2)^4(2.7182)^{-2}}{4!} = \dfrac{(16)(0.13534)}{4x3x2x1!} = \dfrac{2.16549}{24} = 0.090228$

$p(x=5) = \dfrac{(2)^5(2.7182)^{-2}}{5!} = \dfrac{(32)(0,13534)}{5x4x3x2x1!} = \dfrac{4.3309}{120} = 0,03609$

$P(3 \leq 5) = 0.030669$

9. Una compañía telefónica recibe llamadas a razón de 4 por minuto. Calcular la probabilidad de:
   a). Recibir 2 llamadas en un minuto
   b). No recibir ninguna llamada en un minuto
   c). Recibir menos de 3 llamadas en un minuto
   d). Recibir más de 3 llamadas en un minuto

Solución:

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

Datos: X = 2 λ = 4 llamadas /minuto e = 2,7182

a). $p(x=2) = \dfrac{(4)^2(2.7182)^{-4}}{2!} = \dfrac{(16)(0.0183)}{2x1!} = \dfrac{0.29308}{2} = 0.14654$

b). $p(x=0) = \dfrac{(4)^0(2.7182)^{-4}}{0!} = \dfrac{(1)(0.0183)}{1} = 0.0183$

c). $p(x=1) = \dfrac{(4)^1(2.7182)^{-4}}{1!} = \dfrac{(4)(0.0183)}{1} = 0.07327$

p(x menor 3) = p(x=0) + p(x=1) + p(x=2)

p(x menor 3) = 0.0183 + 0.0732 + 0.14654 = 0.2380

d). $p(x=3) = \dfrac{(4)^3(2.7182)^{-4}}{3!} = \dfrac{(64)(0.0183174}{3x2x1!} = \dfrac{1.17234}{6} = 0.19539$

$p(x \leq 3) = p(x=0) + p(x=1) + p(x=2) + p(x=3)$

$p(x \leq 3) = 0.0183 + 0.07327 + 0.14654 + 0.19539$

$p(x \leq 3) = 0.4333$

TAREA:

1. En una clínica el promedio de atención es 16 pacientes por 4 horas, encuentre la probabilidad que en 30 minutos se atiendan menos de 3 personas y que en 180 minutos se atiendan 12 pacientes. Usamos la distribución de Poisson.

2. En una tienda los clientes llegan al mostrador conforme una distribución de Poisson con un promedio de 10 por hora, en una hora dada, ¿Cuál es la probabilidad de que lleguen al menos 5 clientes?

3. Con objeto de revisar la calidad en el pulido de un lente cierta compañía acostumbra determinar el número de manchas en la superficie considerando el lente defectuoso si 3 o más de tales manchas asperezas y otros tipos de defectos aparecen en él, Si la tasa media es de 2 defectos por cm^2, ¿calcule la probabilidad de que un lente de 4 cm^2, que ha sido revisado no se le catalogue como defectuoso? Un lente no se cataloga defectuoso, si tiene menos de 3 defectos.

4. Los reportes de crímenes recientes indican que 3.2 de los robos de vehículos motorizados ocurren cada minuto en estados unidos. Suponga que la distribución de los robos por minuto puede calcularse con la distribución de probabilidad de Poisson.

   a)¿calcule la probabilidad de que ocurran cuatro robos exactamente en un minuto.
   b). ¿cuál es la probabilidad de que en un cuarto de hora cualquiera ocurran exactamente 45 robos?

5. Suponga que la agencia de protección ambiental (APA) es quien establece los estándares para Garantizar la calidad de las emisiones de aire por parte de las empresas. El límite máximo Permitido de cobre en las emisiones es de 10 partículas por millón y usted trabaja en una empresa Donde el valor medio en sus emisiones es de cuatro partículas por millón.

   a). Si se define X como el número de partículas por millón en una muestra ¿Cuál es la desviación estándar de X en su empresa?
   b). Si el número medio de partículas por millón en su empresa es efectivamente de cuatro por millón ¿Tendría usted temor de que la agencia lo multe por contaminar el aire?

3.3 PROPIEDADES: MEDIA, VARIANZA Y DESVIACIÓN ESTÁNDAR

La Media, varianza y desviación estándar de una variable aleatoria Se llama esperanza de la variable aleatoria discreta X, al número:

$$E[X] = x_1p_1 + x_2p_2 + ... + x_np_n \quad x_1, x_2, ..., x_n$$

son los valores de la variable aleatoria y p1, p2, ..., pn, las probabilidades respectivas.

La esperanza de una variable aleatoria X también se representa por µ, y se llama media de la distribución. Por tanto, "esperanza de la variable aleatoria" y "media de la distribución" son expresiones equivalentes.

$$\mu = \overset{n}{\underset{i=1}{E}}\, p_i\, x_i \; i=1 = E\,[X]$$

El conocimiento de la media de la distribución no es suficiente para caracterizar la distribución, ya que hay distribuciones con la misma media y distintas unas de otras. 147 Para medir la dispersión de los valores de una variable aleatoria X respecto de su media µ, se define el siguiente estadístico llamado varianza:

$$V[X] = E\,[(x,\, \mu)^2]$$

Es decir:

$$V[X] = (x_{1,}\, \mu)^2\, p_1 + (x_2,\, \mu)^2\, p_2 + ... + (x_2,\, \mu)^2\, p_n$$

Puesto que la varianza no podría medirse en las mismas unidades que la variable, utilizamos la raíz cuadrada de la varianza y a este número la llamamos desviación típica.

$$Desv[X] = V[X]$$

$$Desv[X] = (x_{1,}\, \mu)^2 p_1 + x(x_2,\, \mu)^2 p_2 + ... + \;(\, x_n \,!\, \mu)^2\, p_n$$

EJEMPLO:

1. Calcular la media y la varianza del número de hijos varones de una familia con dos hijos.

Solución:
E={VV, VH, HV, HH}
X={0, 1, 2}= "número de hijos varones de una familia con dos hijos"
P_1 = P(X = 0) = 1/ 4
P_2 = P(X =1) = 2 / 4 =1 / 2
P_3 = P(X = 2) = 1/ 4
P(x)= 1/4 +1/2 +1/4 =1
En promedio, una familia con dos hijos tiene un hijo varón con una varianza de 1/2.

2. Tras una intervención quirúrgica de un tipo determinado, el equipo médico mantuvo en el hospital a unos pacientes cinco días y a otros ocho. De éstos últimos no regresó ninguno al hospital y el coste de cada uno ascendió a 90.000 pts., mientras que, de los dados de alta a los cinco días, las dos terceras partes no regresaron al hospital y el coste por cada individuo fue de 50.000 pts. El otro tercio restante tuvo que regresar al hospital ocasionando unos gastos totales por individuo de 150.000 pts. En términos puramente económicos, ¿es preferible dar de alta a los enfermos a los cinco o a los ocho días?

Solución: Se trata de calcular el coste promedio en ambos casos. En el supuesto de que los pacientes estén ingresados 8 días, el coste promedio es de 90.000 pts., y en el supuesto de que los pacientes estén 5 días, la variable aleatoria se distribuye de la siguiente forma:

X	50.000	150.000
P(X)	2/3	1/3

El coste promedio en este caso será:

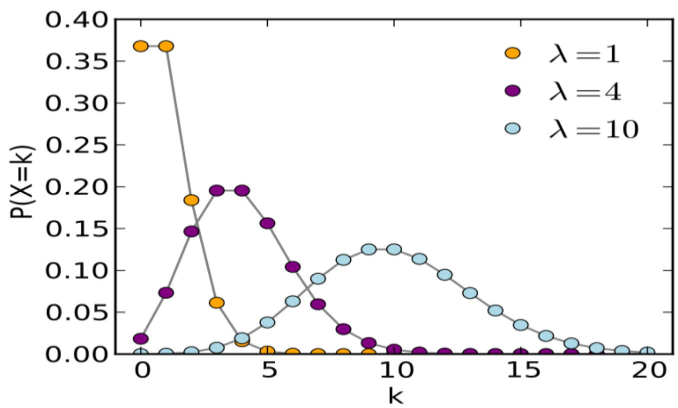
$$E[X] = 50.000\frac{(2)}{3} + 150.000\frac{(1)}{3} = 83.330\text{pts.}$$

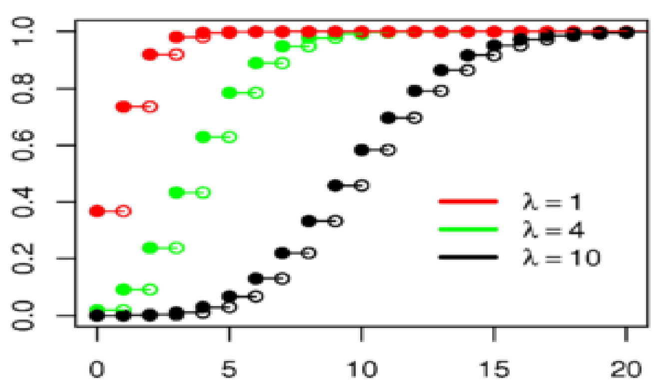
Puesto que 83.333 < 90.000, esto indica que es preferible, desde el punto de vista económico, tener ingresados a los pacientes cinco días.

La varianza la calculamos de la siguiente forma:

$$V[X] = (50.000 - 83.000)^2\frac{(2)}{3} + (150.000 - 83.330)^2\frac{(1)}{3} = 2,2109$$

3.4 GRAFICA





El eje horizontal es el índice *x*. La función solamente está definida en valores enteros de *k*. Las líneas que conectan los puntos son solo guías para el ojo y no indican continuidad.

Función de densidad de probabilidad

3.5 HIPERGEOMETRICA

La distribución hipergeométrica es aquella en la que se considera la existencia de éxitos y/o fracasos en una población conocida, y de la cual se extrae una muestra sin remplazo donde también existen éxitos o fracasos.

La Distribución Hipergeométrica se utiliza para calcular la probabilidad de obtener determinado número de éxitos en un espacio muestral de n ensayos.

La distribución hipergeométrica se emplea para muestreos sin reemplazo de una población finita cuya probabilidad cambia a lo largo del ensayo.

Si en una población de N elementos se tienen k éxitos, la probabilidad de que en una muestra aleatoria de n elementos seleccionados sin reemplazo se tengan x éxitos está dada por:

$$p(x) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}} = \frac{C_x^k \; C_{n-x}^{N-k}}{C_n^N} \qquad \text{con } x \leq k \qquad \text{Donde}$$

N = número de elementos en la población
n = número de elementos en la muestra
k = número de éxitos en la población
x = número de éxitos en la muestra

Además C_n^N es el número de maneras en que se puede tomar una muestra *n* de la población *N*

C_x^k es el número de formas en que se toman x éxitos del total *r* éxitos que hay en la población

C_{n-x}^{N-k} es el número de maneras en que se puede tomar *n* - x fracasos del total *N* - *r* de la población.

La media (esperanza) y desviación estándar de la distribución hipergeométrica están dadas por:

$$\text{Media} \quad \mu = n\left(\frac{k}{N}\right)$$

$$\text{Desviación estándar } \sigma = \sqrt{n\left(\frac{k}{N}\right)\left(\frac{N-k}{N}\right)\left(\frac{N-n}{N-1}\right)} = \sqrt{\mu\left(\frac{N-k}{N}\right)\left(\frac{N-n}{N-1}\right)}$$

Los experimentos que tienen este tipo de distribución tienen las siguientes características:

Al realizar un experimento con este tipo de distribución, se esperan dos tipos de resultados.

Las probabilidades asociadas a cada uno de los resultados no son constantes.
Cada ensayo o repetición del experimento no es independiente de los demás.

El número de repeticiones del experimento (n) es constante.

¿Qué es la distribución hipergeométrica?

La distribución hipergeométrica es una distribución discreta que modela el número de eventos en una muestra de tamaño fijo cuando usted conoce el número total de elementos en la población de la cual proviene la muestra. Cada elemento de la muestra tiene dos resultados posibles (es un evento o un no evento). Las muestras no tienen reemplazo, por lo que cada elemento de la muestra es diferente. Cuando se elige un elemento de la población, no se puede volver a elegir. Por lo tanto, la probabilidad de que un elemento sea seleccionado aumenta con cada ensayo, presuponiendo que aún no haya sido seleccionado.

Utilice la distribución hipergeométrica para muestras obtenidas de poblaciones relativamente pequeñas, sin reemplazo. Por ejemplo, la distribución hipergeométrica se utiliza en la prueba exacta de Fisher para probar la diferencia entre dos proporciones y en muestreos de aceptación por atributos cuando se toman muestras de un lote aislado de tamaño finito.

La distribución hipergeométrica se define por 3 parámetros: tamaño de la población, conteo de eventos en la población y tamaño de la muestra.

Por ejemplo, usted recibe un envío de pedido especial de 500 etiquetas. Supongamos que el 2% de las etiquetas es defectuoso. El conteo de eventos en la población es de 10 (0.02 * 500). Usted toma una muestra de 40 etiquetas y desea determinar la probabilidad de que haya 3 o más etiquetas defectuosas en esa muestra. La probabilidad de que haya 3 o más etiquetas defectuosas en la muestra es de 0.0384.

Los experimentos que tienen este tipo de distribución tienen las siguientes características:

a) Al realizar un experimento con este tipo de distribución, se esperan dos tipos de resultados.
b) Las probabilidades asociadas a cada uno de los resultados no son constantes.
c) Cada ensayo o repetición del experimento no es independiente de los demás.
d) El número de repeticiones del experimento (*n*) es constante.

Ejemplo de cálculo de las probabilidades hipergeométricas

Supongamos que hay diez automóviles disponibles para que usted los pruebe

(N = 10) y cinco de ellos tienen motores turbo (x = 5). Si prueba tres de los vehículos (n = 3), ¿cuál es la probabilidad de que dos de los tres que probará tengan motores turbo?

1. Elija Calc > Distribuciones de probabilidad > Hipergeométrica.
2. Elija Probabilidad.
3. En Tamaño de la población (N), ingrese *10*. En Conteo de eventos en la población (M), ingrese *5*. En Tamaño de la muestra (n), ingrese *3*.
4. Elija Constante de entrada e ingrese *2*.
5. Haga clic en Aceptar.

La probabilidad de que seleccione exactamente dos automóviles con motores turbo de forma aleatoria cuando pruebe tres de los diez vehículos es 41.67%.

Ejemplos:

1. En una urna o recipiente hay un total de *N* objetos, entre los cuales hay una cantidad *a* de objetos que son defectuosos, si se seleccionan de esta urna *n* objetos al azar, y sin reemplazo, ¿cuál es la probabilidad de obtener *x* objetos defectuosos?

Solución:

$$p(x,n) = \frac{{}_{a}C_{x} * {}_{N-a}C_{n-x}}{{}_{N}C_{n}}$$

donde:
p(*x*,*n*) = probabilidad de obtener *x* objetos defectuosos de entre *n* seleccionados
${}_{a}C_{x} * {}_{N-a}C_{n-x} =$ muestras de n objetos en donde hay *x* que son defectuosos y n-x buenos
${}_{N}C_{n} = \delta =$ todas las muestras posibles de seleccionar de *n* objetos tomadas de entre *N* objetos en total = espacio muestral

Considerando que en la urna hay un total de 10 objetos, 3 de los cuales son defectuosos, si de seleccionan 4 objetos al azar, ¿cuál es la probabilidad de que 2 sean defectuosos?

Solución:
N = 10 objetos en total
a = 3 objetos defectuosos
n = 4 objetos seleccionados en muestra
x = 2 objetos defectuosos deseados en la muestra

$$p(x=2,n=4)=\frac{{}_3C_2 * {}_{10-3}C_{4-2}}{{}_{10}C_4}=\frac{{}_3C_2 * {}_7C_2}{{}_{10}C_4}=\frac{\frac{3!}{(3-2)!2!} * \frac{7!}{(7-2)!2!}}{\frac{10!}{(10-4)!4!}}=$$

$$=\frac{\frac{3!}{1!2!} * \frac{7!}{5!2!}}{\frac{10!}{6!4!}}=\frac{\frac{3x2x1!}{1!2!} * \frac{7x6x5!}{5!2!}}{\frac{10x9x8x7x6!}{6!4!}}=\frac{3x2x7x6}{10x9x8x7} * \frac{4!}{2!2!}=$$

donde:

$$\frac{3x2x7x6}{10x9x8x7}=$$

probabilidad asociada a cada muestra de 4 objetos que se seleccionaron, con lo que se demuestra que las probabilidades no son constantes

$$\frac{4!}{2!2!}=$$

formas o maneras de obtener 2 objetos defectuosos entre los 4 seleccionados = muestras de 4 objetos entre los que 2 son defectuosos.

Como se observa en el desarrollo de la solución del problema, la pretensión es demostrar que las probabilidades asociadas a cada uno de los resultados no son constantes.

Luego la probabilidad de obtener 2 objetos defectuosos entre los 4 seleccionados al azar sería:

$$=\frac{3x2x7x6}{10x9x8x7} * \frac{4!}{2!2!}=\frac{252}{5040} * \frac{24}{4}=\frac{6048}{20160}=0.30$$

2. Para evitar que lo descubran en la aduana, un viajero ha colocado 6 tabletas de narcótico en una botella que contiene 9 píldoras de vitamina que son similares en apariencia. Si el oficial de la aduana selecciona 3 tabletas aleatoriamente para analizarlas.

   a) ¿Cuál es la probabilidad de que el viajero sea arrestado por posesión de narcóticos?,
   b) ¿Cuál es la probabilidad de que no sea arrestado por posesión de narcóticos?

Solución:
a) Datos N = 9+6 =15 total de tabletas
a = 6 tabletas de narcótico
n = 3 tabletas seleccionadas

x = 0, 1, 2, o 3 tabletas de narcótico = variable que nos indica el número de tabletas de narcótico que se puede encontrar al seleccionar las 3 tabletas

p(viajero sea arrestado por posesión de narcóticos) = p(de que entre las 3 tabletas seleccionadas haya 1 o más tabletas de narcótico)

$$= p(x = 1,2 ó 3 tabletas; n = 3) = \frac{{}_6C_1 * {}_9C_2}{{}_{15}C_3} + \frac{{}_6C_2 * {}_9C_1}{{}_{15}C_3} + \frac{{}_6C_3 * {}_9C_0}{{}_{15}C_3} =$$

$$= \frac{(6)(36)}{455} + \frac{(15)(9)}{455} + \frac{(20)(1)}{455} = \frac{216 + 135 + 20}{455} = \frac{371}{455} = 0.81538$$

otra forma de resolver;

p(el viajero sea arrestado por posesión de narcóticos) = 1 – p(de que entre las tabletas seleccionadas no haya una sola de narcótico)

$$= 1 - p(x = 0; n = 3) = 1 - \frac{{}_6C_0 * {}_9C_3}{{}_{15}C_3} =$$

$$= 1 - \frac{(1)(84)}{455} = -0.184615 = 0.815385$$

b) p(no sea arrestado por posesión de narcóticos)

$p(x=0; n=3) = \frac{{}_6C_0 * {}_9C_3}{{}_{15}C_3}$

$$= p(x = 0; n = 3) = \frac{{}_6C_0 * {}_9C_3}{{}_{15}C_3} =$$

$$= \frac{(1)(84)}{455} = 0.184615$$

3. De un lote de 10 proyectiles, 4 se seleccionan al azar y se disparan. Si el lote contiene 3 proyectiles defectuosos que no explotarán, ¿cuál es la probabilidad de que, a) los 4 exploten?, b) al menos 2 no exploten?

Solución:

a) Datos:

N = 10 proyectiles en total

a = 7 proyectiles que explotan

n = 4 proyectiles seleccionados

x = 0, 1, 2, 3 o 4 proyectiles que explotan = variable que nos define el número de proyectiles que explotan entre la muestra que se dispara

$p(x=4; n=4) = \frac{{}_7C_4 * {}_3C_0}{{}_{10}C_4} = \frac{(35)(1)}{210} = \frac{35}{210} = 0.16667$

b) Datos:

N = 10 proyectiles en total

a = 3 proyectiles que no explotan
n = 4 proyectiles seleccionados
x = 0, 1, 2 o 3 proyectiles que no explotan
p(al menos 2 no exploten) = p (2 o más proyectiles no exploten)=p(x=2 o 3; n=4)

$$p(x = 2 \text{ o } 3; n=4) = \frac{{}_3C_2 * {}_7C_2 + {}_3C_3 * {}_7C_1}{{}_{10}C_4} = \frac{(3)(21)}{210} = \frac{63 + 7}{210} = \frac{70}{210} = 0.3333$$

4 a)¿Cuál es la probabilidad de que una mesera se rehúse a servir bebidas alcohólicas únicamente a dos menores de edad si verifica aleatoriamente solo 5 identificaciones de entre 9 estudiantes, de los cuales 4 no tienen la edad suficiente?

b) ¿Cuál es la probabilidad de que como máximo 2 de las identificaciones pertenezcan a menores de edad?

Solución:
a). Datos:
N = 9 total de estudiantes
a = 4 estudiantes menores de edad
n = 5 identificaciones seleccionadas
x = variable que nos define el número de identificaciones que pertenecen a personas menores de edad
x = 0, 1, 2, 3 o 4 identificaciones de personas menores de edad

$$p(x=2; n=5) = \frac{{}_4C_2 * {}_5C_3}{{}_9C_5} = \frac{(3)(10)}{126} = 0.238095$$

b). Datos:
N = 9 total de estudiantes
a = 4 estudiantes menores de edad
n = 5 identificaciones seleccionadas
x = variable que nos define el número de identificaciones que pertenecen a personas menores de edad

x = 0, 1, 2, 3 o 4 identificaciones de personas menores de edad

$$P(x= 0, 1, 2 ; n=5) = \frac{{}_4C_0 * {}_5C_5 + {}_4C_1 * {}_5C_4 + {}_4C_2 * {}_5C_3}{{}_9C_5} = (1)(1)+(4)(5)+(6)(10)$$

$$P(x=0,1; n=5)= \frac{1 + 20 + 60}{126} = \frac{81}{126} = 0.64286$$

3. Un lote contiene 100 piezas de un proveedor de tubería local y 200 unidades de un proveedor de tubería del estado vecino. Si se seleccionan 4 piezas al azar y sin reemplazo.

a) ¿Cuál es la probabilidad de que todas sean del proveedor local?

Datos: N = 300 | X = 100 | n = 4 | x = 4

$$P(x=4)=\frac{C_{4-4}^{300-100}C_{4}^{100}}{C_{4}^{300}}=\frac{\left(\frac{200!}{(200-0)!\,0!}\right)\left(\frac{100!}{(100-4)!\,4!}\right)}{\frac{300!}{(300-4)!\,4!}}$$

$$=\frac{\left(\frac{200!}{200!\,0!}\right)\left(\frac{100!}{96!\,4!}\right)}{\frac{300!}{296!\,4!}}=\frac{25\times 33\times 49\times 97}{75*299*149*99}=0.0119$$

$$P(X\geq 2)=P(X=2)+P(x=3)+P(X=4)$$

$$P(X\geq 2)=\frac{C_{4-2}^{300-100}C_{2}^{100}}{C_{4}^{300}}+\frac{C_{4-3}^{300-100}C_{3}^{100}}{C_{4}^{300}}+\frac{C_{4-4}^{300-100}C_{4}^{100}}{C_{4}^{300}}$$

$$=\frac{\left(\frac{200!}{(200-2)!\,2!}\right)\left(\frac{100!}{(100-2)!\,2!}\right)+\left(\frac{200!}{(200-1)!\,1!}\right)\left(\frac{100!}{(100-3)!\,3!}\right)+\left(\frac{200!}{(200-0)!\,0!}\right)\left(\frac{100!}{(100-4)!\,4!}\right)}{\frac{300!}{(300-4)!\,4!}}$$

$$=\frac{\left(\frac{200!}{188!\,2!}\right)\left(\frac{100!}{98!\,2!}\right)+\left(\frac{200!}{199!\,1!}\right)\left(\frac{100!}{97!\,3!}\right)+\left(\frac{200!}{200!\,0!}\right)\left(\frac{100!}{96!\,4!}\right)}{\frac{300!}{296!\,4!}}$$

$$P(X\geq 2)=0.408$$

c) ¿Cuál es la probabilidad de que al menos una pieza de la muestra sea del proveedor local?

Datos: N = 300 | X = 100 | n = 4 | x = 1, 2, 3 o 4

$$P(X\geq 1)=1-P(x=0)=1-\frac{C_{4-0}^{300-100}C_{0}^{100}}{C_{4}^{300}}$$

$$=1-\frac{\left(\frac{200!}{196!\,0!}\right)\left(\frac{100!}{100!\,0!}\right)}{\frac{300!}{296!\,4!}}=1-0.1955=0.8045$$

4. Considerando que en la urna hay un total de 10 objetos, 3 de los cuales son defectuosos, si de seleccionan 4 objetos al azar, ¿cuál es la probabilidad de que 2 sean defectuosos?

Datos: N = 10 | X = 3 | n = 4 | x = 2

$$P(x=2)=\frac{C_{4-2}^{10-3}C_{2}^{3}}{C_{4}^{10}}=\frac{\left(\frac{7!}{5!\,2!}\right)\left(\frac{3!}{1!\,2!}\right)}{\frac{10!}{6!\,4!}}$$

$$P(x=2)=\frac{3}{10}=0.3$$

a) ¿Cuál es la probabilidad de que una mesera se rehúse a servir bebidas alcohólicas únicamente a dos menores de edad si verifica aleatoriamente solo 5 identificaciones de entre 9 estudiantes, de los cuales 4 no tienen la edad suficiente?

b) ¿Cuál es la probabilidad de que como máximo 2 de las identificaciones pertenezcan a menores de edad?

a) Datos N=9 X=4 |n=5| x=2

$$P(x=2)=\frac{C_{5-2}^{9-4}C_{2}^{4}}{C_{5}^{9}}=\frac{\left(\frac{5!}{(5-3)!\,3!}\right)\left(\frac{4!}{(4-2)!\,2!}\right)}{\frac{9!}{(9-5)!\,5!}}$$

$$=\frac{\left(\frac{5!}{2!\,3!}\right)\left(\frac{4!}{2!\,2!}\right)}{\frac{9!}{4!\,5!}}=\frac{60}{126}=0.4762$$

b) Datos: N = 9 | X = 4 | n = 5 | x = 0, 1, 2

$$P(x=0)+P(x=1)+P(x=2)=\frac{C_{5-0}^{9-4}C_0^4}{C_5^9}+\frac{C_{5-1}^{9-4}C_1^4}{C_5^9}+\frac{C_{5-2}^{9-4}C_2^4}{C_5^9}$$

$$=\frac{\left(\frac{5!}{(5-0)!\,0!}\right)\left(\frac{4!}{(4-0)!\,0!}\right)+\left(\frac{5!}{(5-1)!\,1!}\right)\left(\frac{4!}{(4-1)!\,1!}\right)+\left(\frac{5!}{(5-3)!\,3!}\right)\left(\frac{4!}{(4-2)!\,2!}\right)}{\frac{9!}{(9-5)!\,5!}}$$

$$=\frac{\left(\frac{5!}{5!\,0!}\right)\left(\frac{4!}{4!\,0!}\right)+\left(\frac{5!}{4!\,1!}\right)\left(\frac{4!}{3!\,1!}\right)+\left(\frac{5!}{2!\,3!}\right)\left(\frac{4!}{2!\,2!}\right)}{\frac{9!}{4!\,5!}}=\frac{81}{126}=0.6429$$

3.6 PROPIEDADES: MEDIA, VARIANZA Y DESVIACIÓN ESTÁNDAR

La distribución Hipergeometrica es especialmente útil en todos aquellos casos en los que se extraigan muestras o se realicen experiencias repetidas sin devolución del elemento extraído o sin retornar a la situación experimental inicial. Es una distribución fundamental en el estudio de muestras pequeñas de poblaciones pequeñas y en el cálculo de probabilidades de juegos de azar. Tiene grandes aplicaciones en el control de calidad para procesos experimentales en los que no es posible retornar a la situación de partida. Las consideraciones a tener en cuenta en una distribución hipergeométrica:

- El proceso consta de "n" pruebas, separadas o separables de entre un conjunto de "N" pruebas posibles.
- Cada una de las pruebas puede dar únicamente dos resultados mutuamente excluyentes.
- El número de individuos que presentan la característica A (éxito) es "k".
- En la primera prueba las probabilidades son: P(A)= p y P(A)= q; con p+q=1.

En estas condiciones, se define la variable aleatoria *X = "nº de éxitos obtenidos"*. La función de probabilidad de esta variable sería:

$$p(X=x)=\frac{\binom{k}{x}\cdot\binom{N-k}{n-x}}{\binom{N}{n}} \qquad \begin{array}{l} N = \textit{tamaño de población} \\ K = n^{o}\ \textit{individuos que...} \\ n = \textit{tamaño de la muestra} \\ x = \textit{valor que toma la variable} \end{array}$$

La media, varianza y desviación típica de esta distribución vienen dadas por:

$$\mu = np$$

$$\sigma^2 = npq \cdot \frac{N-n}{N-1}$$

$$\sigma = \sqrt{npq \cdot \frac{N-n}{N-1}}$$

Cuando N es muy grande, como criterio se suele considerar N > 10n, la distribución hipergeométrica se puede aproximar por la binomial ***B(n, p)*** con ***p = k/N***.

En la siguiente escena puedes observar la función de probabilidad de la distribución hipergeométrica. Puedes cambiar los diferentes parámetros que configuran dicha distribución y observar cómo cambia esta función a medida que se varía alguno de ellos. Extrae tus propias conclusiones. Así mismo, puedes utilizar también la escena como calculadora directa que permite resolver situaciones concretas que se puedan plantear en problemas específicos. Lógicamente hay un límite para los valores de la población de manera que la escena funcione con fluidez (valores menores de 200).

EJEMPLOS:

1: Supongamos la extracción aleatoria de 8 elementos de un conjunto formado por 40 elementos totales (cartas baraja española) de los cuales 10 son del tipo A (salir oro) y 30 son del tipo complementario (no salir oro).

Si realizamos las extracciones sin devolver los elementos extraídos y llamamos X al número de elementos del tipo A (oros obtenidos) que extraemos en las 8 cartas; X seguirá una distribución hipergeométrica de parámetros 40, 8, 10/40. H (40, 8, 0, 25).

Para calcular la probabilidad de obtener 4 oros:

$$p(X=4) = \frac{\binom{10}{4} \cdot \binom{30}{4}}{\binom{40}{8}} = 0{,}07$$

2. De cada 20 piezas fabricadas por una máquina, hay 2 que son defectuosas. Para realizar un control de calidad, se observan 15 elementos y se rechaza el lote si hay alguna que sea defectuoso. Vamos a calcular la probabilidad de que el lote sea rechazado.

$N = 20$

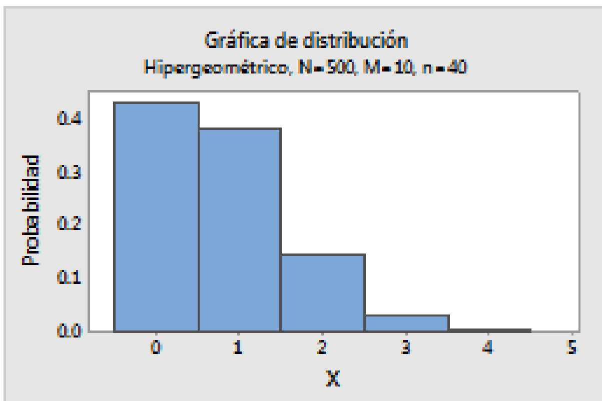
$n = 15$

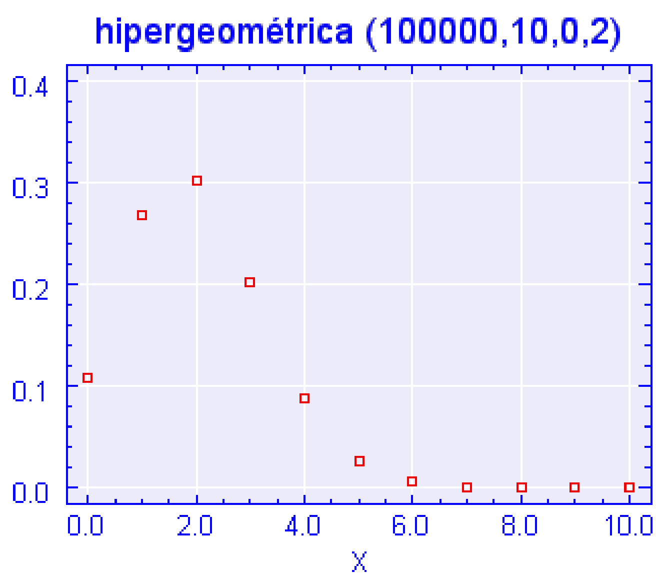
$X =$ número de piezas defectuosa s de las 15 escogidas

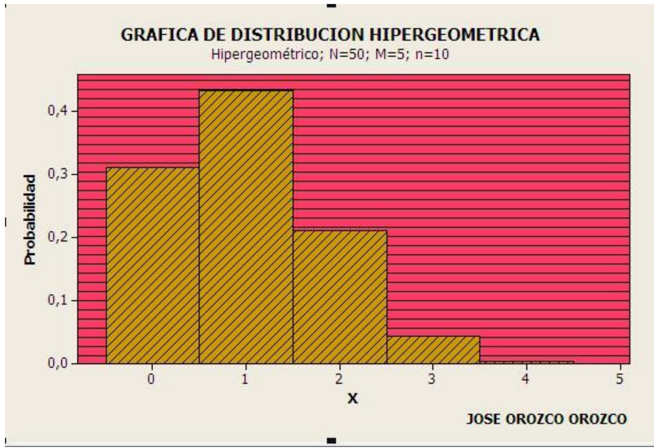
$P(X \geq 1) = 1 - p(X < 1) = 1 - p(X = 0)$

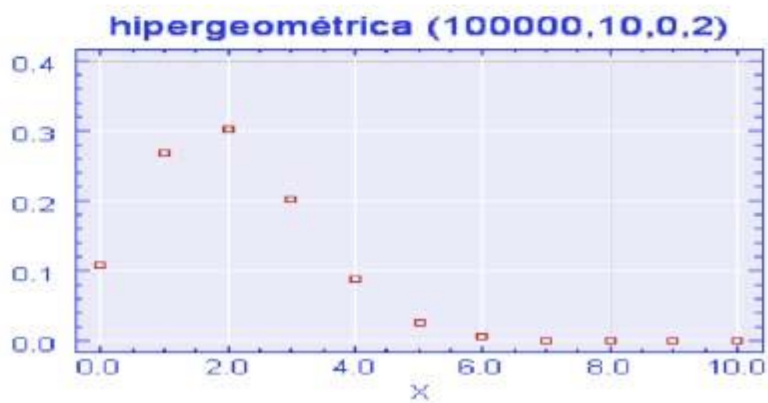
$$1 - \frac{\binom{2}{0} \cdot \binom{20-2}{15}}{\binom{20}{15}} = 1 - \frac{816}{15504} = \frac{18}{19} = 0{,}947$$

3.7 GRAFICA









3.8 NORMAL Y LOGARÍTMICO-NORMAL

En estadística y probabilidad se llama distribución normal, distribución de Gauss, distribución gaussiana o distribución de Laplace-Gauss, a una de las distribuciones de probabilidad de variable continua que con más frecuencia aparece en estadística y en la teoría de probabilidades.

La gráfica de su función de densidad tiene una forma acampanada y es simétrica respecto de un determinado parámetro estadístico. Esta curva se conoce como campana de Gauss y es el gráfico de una función gaussiana.[]

La importancia de esta distribución radica en que permite modelar numerosos fenómenos naturales, sociales y psicológicos. []Mientras que los mecanismos que subyacen a gran parte de este tipo de fenómenos son desconocidos, por la enorme cantidad de variables incontrolables que en ellos intervienen, el uso del modelo normal puede justificarse asumiendo que cada observación se obtiene como la suma de unas pocas causas independientes.

De hecho, la estadística descriptiva sólo permite describir un fenómeno, sin explicación alguna. Para la explicación causal es preciso el diseño experimental, de ahí que al uso de la estadística en psicología y sociología sea conocido como método correlacional.

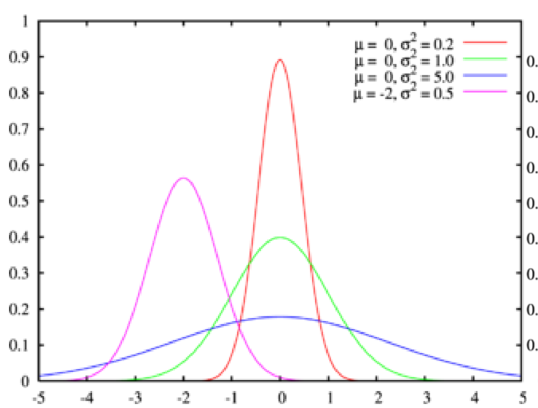
La distribución normal también es importante por su relación con la estimación por mínimos cuadrados, uno de los métodos de estimación más simples y antiguos.

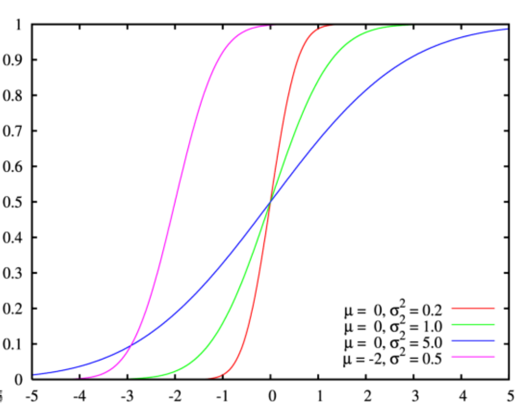
Algunos ejemplos de variables asociadas a fenómenos naturales que siguen el modelo de la normal son:

- caracteres morfológicos de individuos como la estatura;
- caracteres fisiológicos como el efecto de un fármaco;
- caracteres sociológicos como el consumo de cierto producto por un mismo grupo de individuos;
- caracteres psicológicos como el cociente intelectual;
- nivel de ruido en telecomunicaciones;
- errores cometidos al medir ciertas magnitudes;
- etc.

La distribución normal también aparece en muchas áreas de la propia estadística. Por ejemplo, la distribución muestral de las medias muestrales es aproximadamente normal, cuando la distribución de la población de la cual se extrae la muestra no es normal.[] Además, la distribución normal maximiza la entropía entre todas las distribuciones con media y varianza conocidas, lo cual la convierte en la elección natural de la distribución subyacente a una lista de datos resumidos en términos de media muestral y varianza. La distribución normal es la más extendida en estadística y muchos test estadísticos están basados en una "normalidad" más o menos justificada de la variable aleatoria bajo estudio.

En probabilidad, la distribución normal aparece como el límite de varias distribuciones de probabilidad continuas y discretas.





La línea verde corresponde a la probabilidad distribución normal estándar Función de densidad de probabilidad.

Función de distribución de

Una distribución normal de media μ y desviación típica σ se designa por N (μ, σ). Su gráfica es la campana de Gauss: El área del recinto determinado por la función y el eje de abscisas es igual a la unidad. Esto es porque la probabilidad de que un suceso ocurra entre todas las posibilidades es un 100%, o sea 1. La integral, entre menos y más infinito, de la función de densidad de probabilidad es 1. Al ser simétrica respecto al eje que pasa por x = μ, deja un área igual a 0.5 a la izquierda y otra igual a 0.5 a la derecha. Esta función de densidad de probabilidad tiene una expresión parecida a g:g(x)= e-x2. La distribución Normal o curva Normal de Gauss es:

Como consecuencia de la Propiedad 1; es posible relacionar todas las variables aleatorias normales con la distribución normal estándar.
Si,$X \sim N(\mu, \sigma^2)$ entonces

$$Z = \frac{X - \mu}{\sigma}$$

es un variable aleatoria normal estándar: $Z \sim N(0, 1)$. .

La transformación de una distribución $X \sim N(\mu, \sigma)$ en una $N(0, 1)$ se llama normalización, estandarización o tipificación de la variable X.

Una consecuencia importante de esto es que la función de distribución de una distribución normal es, por consiguiente,

$$\Pr(X \leq x) = \Phi\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{2}\left(1 + \operatorname{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right).$$

A la inversa, si es una distribución normal estándar, $Z \sim N(0, 1)$, entonces

$$X = \sigma Z + \mu$$

es una variable aleatoria normal tipificada de media μ y varianza σ^2.

La distribución normal estándar está tabulada (habitualmente en la forma del valor de la función de distribución Φ) y las otras distribuciones normales pueden obtenerse como transformaciones simples, como se describe más arriba, de la distribución estándar. De este modo se pueden usar los valores tabulados de la función de distribución normal estándar para encontrar valores de la función de distribución de cualquier otra distribución normal.

Ejemplos:
1. La temperatura durante setiembre está distribuida normalmente con media 18,7ºC y desviación standard 5ºC. Calcule la probabilidad de que la temperatura durante setiembre esté por debajo de 21ºC.

Solución.
Datos: μ = 18,7ºC σ = 5ºC x = 21ºC

$$z = \frac{x - \mu}{\sigma}$$

z = $\underline{21 - 18{,}7}$ = $\underline{2.3}$ = 0.46

5 5

0.46 Ahora vamos a la tabla y para el valor de Z = 0,46 tenemos que la probabilidad es de 0,6772

2. Una población normal tiene una media de 80 una desviación estándar de14.0

a). Calcule la probabilidad de un valor localizado entre 75.0 y 90.0

p $(75 \leq x \leq 90)$

Solución:

Datos: $\mu = 80$ $\sigma = 14$ $x = 90$

$$z = \frac{x - \mu}{\sigma}$$

$$z = \frac{90-80}{14} = \frac{10}{14} = 0.71 = 0.7611$$

$$z = \frac{75-80}{14} = \frac{-5}{14} = -0.36 = 0.3594$$

p $(75 \leq x \leq 90) = 0.7611 - 0.3594 = 0.4017$

b). Calcule la probabilidad de un valor de 75.0 ó menor. $p(x \leq 75)$

$$z = \frac{75-80}{14} = \frac{-5}{14} = -0.36 = 0.3594$$

c). Calcule la probabilidad de un valor localizado entre 55.0 y 70.0

p $(55 \leq x \leq 70)$

$$z = \frac{70-80}{14} = \frac{-10}{14} = \frac{-10}{14} = -0.71 = 0.2389$$

$$z = \frac{55-80}{14} = \frac{-25}{14} = -1.79 = 0.0367$$

p $(55 \leq x \leq 70) = 0.2389 - 0.0367 = 0.2022$

3. Los montos de dinero que se piden en las solicitudes de préstamos en Down River Federal Savings tiene una distribución normal, una media de $70,000y una desviación estándar de $20,000. Esta mañana se recibió una solicitud de préstamo. ¿Cuál es la probabilidad de que:

Solución:

Datos: $\mu = \$70,000$ $\sigma = \$20,000$ $x = 80,000$

$$z = \frac{x - \mu}{\sigma}$$

a). El monto solicitado sea de $80,000 o superior $p(x \geq 80,000)$

Solución:

$z = \frac{80,000 - 70,000}{20,000} = \frac{10,000}{20,000} = 0.50 = 0.6915$

p(x ≥ 80,000) = 1– 0.6915=**0.3085**

b). El monto solicitado oscile entre $65,000 y $80,000
p (65,000 ≤ x ≤ 80,000)

$z = \frac{80,000 - 70,000}{20,000} = \frac{10,000}{20,000} = 0.50 = 0.6915$

$z = \frac{65,000 - 70,000}{20,000} = \frac{-50,000}{20,000} = -0,25 = 0.4013$

p (65,000 ≤ x ≤ 80,000) = 0.6915 – 0.4013 =0.2902

c). El monto solicitado sea de $65,000 o superior. p(x ≥65,000)

$z = \frac{65,000 - 70,000}{20,000} = \frac{-50,000}{20,000} = -0,25 = 0.4013$

p(x≥ 65,000) = 1–0.4013 =0.59873.

4. Entre las ciudades de Estados Unidos con una población de más de250,000 habitantes, la media del tiempo de viaje de ida al trabajo es de24.3 minutos. El tiempo de viaje más largo pertenece a la ciudad de Nueva York, donde el tiempo medio es de 38.3 minutos. Suponga que la distribución de los tiempos de viaje en la ciudad de Nueva York tiene una distribución de probabilidad normal y la desviación estándar es de 7.5minutos.

a). ¿Qué porcentaje de viajes en la ciudad de Nueva York consumen menos de 30 minutos? p(x ≤ 30)

Solución:
Datos; $\mu = 38.3$ $\sigma = 7.5$ $x = 30$

$z = \frac{x - \mu}{\sigma}$

$z = \frac{30 - 38.5}{7.5} = \frac{-8.3}{7.5} = -1.11 = \underline{0.1335}$

p(x ≤ 30) = 0.1335 = 13.35%

b). ¿Qué porcentaje de viajes consumen entre 30 y 35 minutos? p (30 ≤ x ≤ 35)

$z = \frac{35 - 38.3}{7.5} = \frac{-3.3}{7.5} = -0.44 = 0.3300$

$z = \frac{x - \mu}{\sigma} = \; = \frac{30 - 38.5}{7.5} = \frac{-8.3}{7,5} = -1.11 = \underline{0.1335}$

p(30 ≤ x ≤ 35) = 0.3300 – 0.1335 = 0.1965 = 19.65%

c). ¿Qué porcentaje de viajes consumen entre 30 y 40 minutos? p (30 ≤ x ≤ 40)

$$z = \frac{40 - 38.3}{7.5} = \frac{1.7}{7.5} = 0.23 = 0.5910$$

$p\ (30 \leq x \leq 40) = 0.5910 - 0.1335 = 0.4575 = 45.75\%$.

TAREA:

1. Una distribución normal tiene una media de 80 y una desviación estándar de 14. Determine el valor por encima del cual se presentará 80% de las observaciones.

2. Una distribuidora de cerveza entra al semana un promedio de 135 cajas con una desviación típica de 44 al semana calcular en dos momentos P(≤180) y P(≤120).

3. De los siguiente datos media 60 y una desviación de 22 calcular la dos probabilidad P(≤110) y P(≤25) y solución los siguiente datos media 100 y una desviación de 20 calcular la dos probabilidad P(≤142) y P(≤160)

4. Una población normal tiene una media de 80 una desviación estándar de 14.0 Calcule la probabilidad de un valor localizado entre 75.0 y 90.0 $P(75 \leq x \leq 90)$.

5. Una población normal tiene una media de 80 una desviación estándar de 14.0. Calcule la probabilidad de un valor de 75.0 ó menor $P(x \leq 75)$.

6. Una población normal tiene una media de 80 una desviación estándar de 14.0 Calcule la probabilidad de un valor localizado entre 55.0 y 70.0 $p(55 \leq x \leq 70)$.

LOGARITMICO NORMAL

En probabilidades y estadísticas, la distribución normal logarítmica es una distribución de probabilidad de una variable aleatoria cuyo logaritmo está normalmente distribuido.

Es decir, si X es una variable aleatoria con una distribución normal, entonces exp (X) tiene una distribución log-normal.

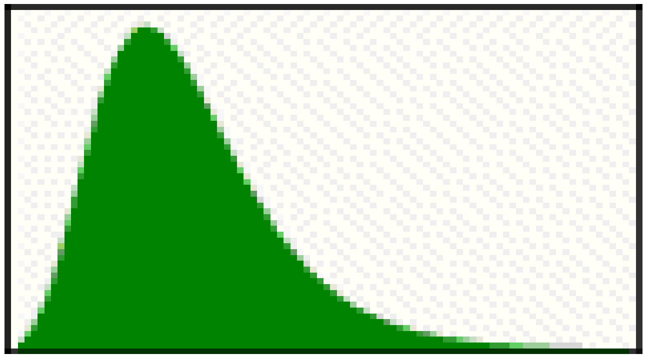
La base de una función logarítmica no es importante, ya que loga X está distribuida normalmente si y solo si logb X está distribuida normalmente, solo se diferencian en un factor constante.

La distribución log-normal tiende a la función densidad de probabilidad {\displaystyle f(x;\mu ,\sigma)={\frac {1}{x\sigma {\sqrt {2\pi }}}}e^{-(\ln(x)-\mu)^{2}/2\sigma ^{2}}}para Log-normal también se escribe log normal o log-normal o distribución de Tinaut.

Una variable puede ser modelada como log-normal si puede ser considerada como un producto multiplicativo de muchos pequeños factores independientes. Un ejemplo típico es un retorno a largo plazo de una inversión: puede considerarse como un producto de muchos retornos diarios.

La distribución log-normal tiende a la función densidad de probabilidad

$$f(x;\mu,\sigma) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-(\ln(x)-\mu)^2/2\sigma^2}$$



Lognormal

para X = 0, donde μ y σ, son la media y la desviación estándar del logaritmo de variable. El valor esperado es:

$E(X) = e^{\mu+\sigma 2/2}$ y la varianza es $\mathrm{var}(X)=(e^{\sigma 2} - 1)e^{2\mu+\sigma 2}$.

Definición

La distribución logarítmica normal se utiliza en una amplia variedad de aplicaciones. La distribución se aplica en casos donde una transformación logarítmica natural tiene como resultado una distribución normal.

Descripción

La distribución se ha aplicado en una amplia variedad de campos en los que se incluyen las ciencias Físicas, las biológicas, las sociales y la ingeniería. En las aplicaciones de esta última la distribución logarítmica se ha utilizado para describir el tiempo de falla en la ingeniería de confiabilidad y el tiempo de reparación en la ingeniería y mantenimiento.

La variable aleatoria continua! tiene una distribución logarítmica normal si la variable aleatoria Y = ln (X) tiene una distribución normal con media μ (miu) y desviación estándar σ(sigma). La función de densidad de X que resulta es:

$$f(x;\mu,\sigma) = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{1}{2\sigma^2}[\ln(x)-\mu]^2}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

Mediante el cambio de variable y ln X e identificando el integrando como la densidad normal con **M = α y σ = β**, encontramos que la densidad buscada es dada por:

$$f(x;\mu,\sigma) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-(\ln(x)-\mu)^2/2\sigma^2}$$

$$f(x;\mu,\sigma) = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{1}{2\sigma^2}[\ln(x)-\mu]^2} \\ 0, \end{cases}$$

t

$$F\left(\frac{In\ b - \alpha}{\beta}\right) - F\left(\frac{In\ a - \alpha}{\beta}\right)$$

Variables: α = alfa β = beta µ (miu) σ = sigma

La distribución logarítmico-normal es continua. Se suele utilizar a menudo en situaciones en las que los valores se sesgan positivamente, por ejemplo, para determinar precios de acciones, precios de propiedades inmobiliarias, escalas salariales y tamaños de depósitos de aceite.

Parámetros
Ubicación, Media, Desviación estándar
De forma predeterminada, la distribución logarítmico normal utiliza la media aritmética y la desviación estándar. En el caso de aplicaciones en las que hay datos históricos disponibles, resulta más adecuado utilizar la desviación estándar logarítmica y la media logarítmica o la media geométrica y la desviación estándar geométrica. Estas opciones están disponibles en el menú Parámetros de la barra de menús. Tenga en cuenta que el parámetro de ubicación está siempre en el espacio aritmético.

Nota:
Si tiene datos históricos disponibles con los que definir una distribución logarítmico normal, es importante calcular la media y la desviación estándar de los logaritmos

de los datos y, a continuación, introducir estos parámetros de logaritmo mediante el menú Parámetros (Ubicación, Media logarítmica y Desviación estándar logarítmica). Calcular la media y la desviación estándar directamente en los datos sin procesar no le dará la distribución logarítmico-normal correcta. También puede optar por utilizar la función de ajuste de distribución que se describe en Ajuste de distribuciones a datos históricos.

Para obtener más información sobre estos parámetros alternativos, consulte la sección de distribución logarítmico-normal en la guía de ejemplos y referencia de Oracle Crystal Ball. Para obtener más información acerca de este menú, consulte Uso de conjuntos de parámetros alternativos.

Condicionales

La distribución logarítmico-normal se utiliza cuando se dan las siguientes condiciones: Los límites superiores e inferiores son ilimitados, pero la variable incierta no puede estar por debajo del valor del parámetro de ubicación.

La distribución se ha sesgado positivamente, con la mayoría de los valores próximos al límite inferior.

El logaritmo natural de la distribución es una distribución normal.

Distribución log-normal

Utilice la distribución log-normal si el logaritmo de la variable aleatoria está distribuido normalmente. Utilícese cuando las variables aleatorias sean mayores que 0. Por ejemplo, la distribución log-normal se usa para el análisis de fiabilidad y en aplicaciones financieras, como modelar el comportamiento de las acciones.

La distribución log-normal es una distribución continua que se define por sus parámetros de ubicación y escala. La distribución log-normal de 3 parámetros se define por sus parámetros de ubicación, escala y valor umbral.

La forma de la distribución log-normal es similar a la forma de las distribuciones log-logística y de Weibull.

Por ejemplo, la siguiente gráfica ilustra la distribución log-normal para escala=1.0, ubicación=0.0 y valor umbral=0.0.

1. Como parte de un análisis de riesgo con respecto a una planta de energía nuclear, los ingenieros deben modelar la resistencia de los soportes de un generador de vapor en función de su capacidad de resistir la aceleración máxima ocasionada por los temblores. La opinión de los expertos sugiere que el logaritmo ln (resistencia) se distribuye normalmente con $\mu = 4{,}0$ y

$\sigma = 0.09$ calcúlese la probabilidad la probabilidad de que los soportes resistan una aceleración máxima de 33.

Solucion:

$$1 - F\left(\frac{\ln(33) - 4.0}{4.0}\right) = 1 - (-1.68) = 0.9535$$

2. Se sabe que históricamente la concentración de contaminantes producidos por plantas químicas exhiben un comportamiento que se parece a una distribución logarítmica normal esto es importante cuando se consideran problemas respecto de la obediencia de las regulaciones gubernamentales. Suponga que la concentración de cierto contaminante, en partes por millón, tiene una distribución logarítmica normal con parámetros $\mu = 3.2$ y $\sigma = 1$ ¿Cuál es la probabilidad de que la concentración exceda 8 partes por millón?

Solución:

$$P(x \geq 8) = F\left(\frac{\ln(8) - 3.2}{3.2}\right) = F(-1.12) = 0.1314$$

3.- Si una persona compra una papeleta en una rifa, en la que puede ganar de 5.000 € ó un segundo premio de 2000 € con probabilidades de: 0.001 y 0.003. ¿Cuál sería el precio justo a pagar por la papeleta?

Solución:

$$E(x) = 5000 \cdot 0.001 + 2000 \cdot 0.003 = 11\ €$$

4.- Un jugador lanza dos monedas. Gana 1 ó 2 € si aparecen una o dos caras. Por otra parte, pierde 5 € si no aparece cara. Determinar la **esperanza matemática** del juego y si éste es favorable.

Solución:

$E = \{(c,c);(c,x);(x,c);(x,x)\}$

$p(+1) = 2/4$

$p(+2) = 1/4$

$p(-5) = 1/4$

$E(x) = 1 \cdot 2/4 + 2 \cdot 1/4 - 5 \cdot 1/4 = -1/4$. Es desfavorable

5.- En una ciudad, la temperatura máxima durante el mes de junio está distribuida normalmente con una media de 26º y una desviación típica de 4º.

Calcular el número de días que se "espera", tengan temperatura máxima comprendida entre 22º y 28º.

6. Como se trata de una distribución Normal, tipificamos (estandarizamos) los valores 22 y 28:

$z_1 = (22 - 26) / 4 = -1$

$z_2 = (28 - 26) / 4 = 0,5$

Entonces la probabilidad de que en un día de junio la temperatura máxima esté entre 22 y 28º es:

$p(22< x < 28) = p(-1 < z < 0,5) = 0, 5328$
Y el número esperado (esperanza) de días es:

Solución:
$E(x) = n * p = 30 * 0, 5328 \approx 16$ días

7. Si X es el número de puntos obtenidos al lanzar un dado de seis caras, encontremos el valor esperado de la variable aleatoria $Y = X^2$.
La función de probabilidad de X es f(x) = 1/6 si X€{1, 2, 3, 4, 5, 6}. La función de probabilidad de $Y = X^2$ es entonces f(y) = 1/6 si Y€{1, 4, 9,16, 25, 36}, así

Solución:
$E(Y) = 1/6*1 + 1/6*4 + 1/6*9 + 1/6*16 + 1/6*25 + 1/6*36 = 1^2*P(X=1) + 2^2*P(X= 2) + 3^2*P(X= 3) + 4^2*P(X= 4) + 5^2*P(X= 5) + 6^2*P(X= 6)$; $E(Y)=$ å $X^2*P(X=x)$

8. Supongamos ahora que X es una v.a. que tiene función de probabilidad

f(x) = 1/6 si X€ {-2,-1,0,1,2,3} y $Y = X^2$.

La función de probabilidad de Y es f(y) = 2/6 si YE{1, 4} y f(y) = 1/6 si Y€{0, 9}. Entonces E(Y) = 2/6*1 + 2/6*4 + 1/6*0 + 1/6*9. Esta ecuación puede escribirse de la siguiente manera:

Solución:
$E(Y) = 2/6*1 + 2/6*4 + 1/6*0 + 1/6*9 = 1*P(Y=1) + 4*P(Y=4) + 0*P(Y=0) + 9*P(Y=1)$
$E(Y) = 1^2*P(X=1$ ó $X=-1) + 2^2*P(X=2$ ó $X=-2) + 0^2*P(X=0) + 3^2*P(X=3)$
$E(Y) =$ å $X^2*P(X=x)$

A través de estos ejemplos vemos que no es necesario calcular la función de probabilidad de Y, sólo tenemos que usar la función de probabilidad de X y los valores obtenidos al aplicar la función $Y = g(X) = X^2$. Esto es cierto aún en el caso en que la función no es uno-uno

TAREA:
1. El departamento de control de calidad de una fábrica de baterías, somete a prueba de funcionamiento una muestra de 5200 baterías, obteniéndose una duración media de 9000 horas y una desviación estándar de 1200 horas. El estudio determina que 364 de las baterías son rechazadas por no presentar la duración mínima.

a). Determinar la duración mínima para no ser rechazadas por el proceso de control.
b). Cuántas baterías tienen una duración mayor que la duración mínima pero menor de 10800 horas

Respuesta:
a). 7224 tomando Z=-1.48
b) 0,86319 tomando P(X<vida mínima) = 0,07 Calculado a partir del texto

2. Una empresa produce ciertas partes de acero especial para la industria espacial, las longitudes de esas partes se distribuyen normalmente con una media de 4.5 cm, y una dispersión de 0.25 cm, si las especificaciones permiten una tolerancia de ± p cm en la longitud. Determinar el valor de "p" para que la especificación cubra el 95 % de las mediciones

Respuesta:
p=0,49

3. Supóngase que la demanda mensual de un cierto producto se encuentre aproximado por una distribución normal con media 150 y varianza 56. ¿Qué tan grande debe ser el inventario disponible a principio de mes para que la probabilidad de que la existencia se agote no sea mayor de 0.08?

Respuesta:
161 (redondeado) tomando Z=1,41

4. El diámetro de un cable eléctrico armado, está distribuido normalmente con media de 19.559 mm y una desviación de 0.254 mm.

a). Cuál es la probabilidad de que el cable exceda 20.193 mm.

b). Si las especificaciones de Ingeniería son 19.812 ± 0.508. ¿Qué porcentaje de piezas son defectuosas?

Respuesta:
a). 0,00621 tomando Z=2,5 b) 0,16

5. Un gerente que investiga el tiempo que necesitan los operarios para ensamblar una pieza, determinó que los datos (tiempo en s.) tenían una distribución normal con media de 75 y desviación de 6 s.

a). En cuántos segundos se puede esperar que un operario ensamble la pieza si sólo el 10% de todos los operarios pueden ensamblarla en ese tiempo mínimo.

b). ¿Calcule la probabilidad de que un operario requiera más de 60 segundos para ensamblar la pieza?

Respuesta:
a). 67,32 b). 0,99379
6. Una V.A. tiene una distribución Normal con desviación estándar de 15. Si la proba de que asuma un valor mayor que 82.5 es 0.1788. Calcular la media.

Respuesta:

68,25 con Z=0,95

7. La vida útil de cierto tipo de bombillo se distribuye normalmente, con media de 900 horas y una varianza de 400 $horas^2$. Determine la probabilidad que uno de esos bombillos

a). Se queme antes de 925 horas
b). Dure entre 870 y 940 horas
c). Tenga una vida útil superior a las 947 horas

Respuesta:
a). 0,89435 b). 0,91044 c). 0,00939

8. Unos cilindros de concreto tienen una resistencia a la compresión con distribución log-Normal, con β=1 yα=5. Calcule la probabilidad de que la resistencia sea mayor de 300 Kg/cm^3. Calcular la esperanza matemática y la varianza de la distribución Log Normal

Respuesta:
0,24196 E(X)= 244,69 V(X)=102,889)

9. Se supone que los tiempos fuera de servicio (en min.) de un equipo siguen una distribución Log normal. Calcular la mantenibilidad para un tiempo de 195 min. Los tiempos fueron:162 , 175 , 169 , 181 , 192 , 165 , 191 , 204 , 168 , 178 , 182 , 176

Respuesta:

0,90658. Calcular la media de los ln y la desviación estándar de los logaritmos neperianos

10. Se supone que los tiempos entre fallas de un equipo (en horas) siguen una distribución Log normal. Calcular la confiabilidad para un tiempo de 42 horas. Los tiempos fueron:32 , 30 , 82 , 58 , 42 , 65 , 51 , 48 , 41 , 51 , 36 , 39

Respuesta:
0,62172

3.9 PROPIEDADES: MEDIA, VARIANZA Y DESVIACIÓN ESTÁNDAR

Una variable puede ser modelada como log-normal si puede ser considerada como un producto multiplicativo de muchos pequeños factores independientes. Un ejemplo típico es un retorno a largo plazo de una inversión: puede considerarse como un producto de muchos retornos diarios.

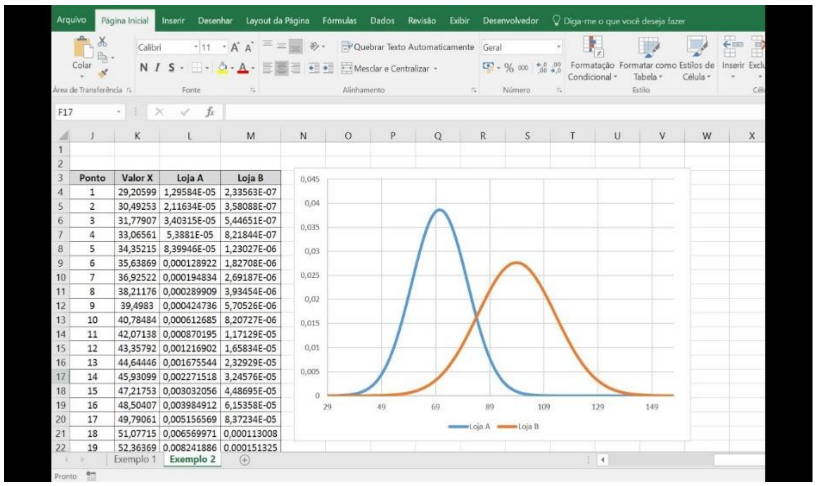
La distribución log-normal tiende a la función densidad de probabilidad para, donde y son la media y la desviación estándar del logaritmo de variable. El valor esperado es:

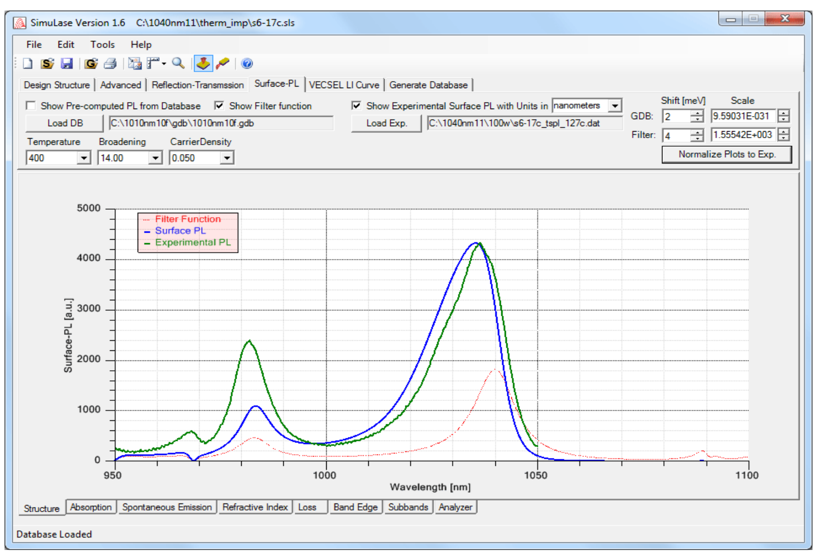
$$f(x;\mu,\sigma) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-(\ln(x)-\mu)^2/2\sigma^2}$$

$$\mathrm{E}(X) = e^{\mu+\sigma^2/2}$$

$$\mathrm{var}(X) = (e^{\sigma^2} - 1)e^{2\mu+\sigma^2}$$

3.10 Grafica





3.11 Aproximación de la Normal a la Binomial

En este caso se estarán calculando probabilidades de experimentos Binomiales de una forma muy aproximada con la distribución Normal, esto puede llevarse a cabo si *n* y *p* = p(éxito) no es muy cercana a *0 y 1*, o cuando *n* es pequeño y *p* tiene un valor muy cercano a ½ ; esto es,

$$P(x,n,p)={}_{n}C_{x}\,p^{x}q^{n-x} \cong p\left(z=\frac{x-np}{\sqrt{npq}}\right)$$

Donde:
x = variable de tipo discreto; solo toma valores enteros
p = np = media de la distribución Binomial
s = $\sqrt{npq}$ = desviación estándar de la distribución Binomial

Cuando ocurren las condiciones anteriores, la gráfica de la distribución Binomial, es muy parecida a la distribución Normal, por lo que es adecuado calcular probabilidades con la Normal en lugar de con la Binomial y de una forma más rápida.

En resumen, se utiliza la aproximación Normal para evaluar probabilidades Binomiales siempre que *p* no esté cercano a *0 o 1*. La aproximación es excelente cuando *n* es grande y bastante buena para valores pequeños de *n* si *p* está razonablemente cercana a ½. Una posible guía para determinar cuándo puede utilizarse la aproximación Normal es tener en cuenta el cálculo de *np*y *nq*. Sí ambos, *np y nq* son *mayores o iguales a 5*, la aproximación será buena.

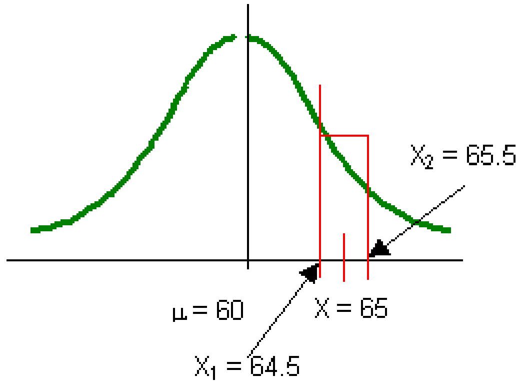
Antes de empezar a resolver problemas con la aproximación Normal, es bueno aclarar que se están evaluando probabilidades asociadas a una variable discreta x, con una distribución que evalúa variables de tipo continuo como es la Normal,

Por lo que z sufre un pequeño cambio como se muestra a continuación:

$$z=\frac{(x\pm 1/2)-\mu}{\sigma}$$

¿Por qué vamos a sumar o a restar ½ a x?

Este es un factor de corrección debido a que se está evaluando una variable discreta con una distribución continua, por lo que hay que delimitar claramente desde que punto se va a evaluar la variable, dicho de otra forma, en que límite de la barra (inferior o superior) nos debemos posicionar para determinar la probabilidad requerida, cada barra de probabilidad a evaluar tiene como base la unidad, ese es el porqué del □ ½.



Ejemplos:

1. La probabilidad de que un paciente se recupere de una rara enfermedad de la sangre es de 0.4. Si se sabe que 100 personas han contraído esta enfermedad, ¿Cuál es la probabilidad de que:

a) al menos 30 sobrevivan?,

b) más de 46 sobrevivan?,

c) menos de 50 no sobrevivan?

Solución:

a)

n = 100

p = p(paciente se recupere) = 0.40

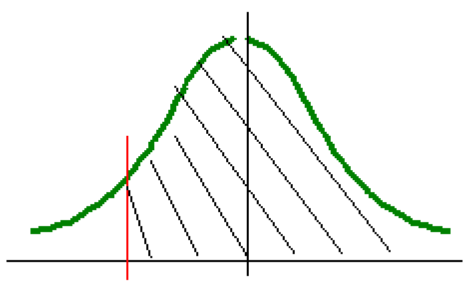
q = p(paciente no se recupere) = 1 – p = 1 – 0.40 = 0.60

□ = *np* = (100)(0.40) = 40 pacientes se recuperen

□ = $\sqrt{npq} = \sqrt{100(0.40)(0.60)} = 4.899$ pacientes que se recuperan

x = variable que nos define el número de pacientes que se recuperan

x = 0, 1, 2,....,100 pacientes que se recuperan

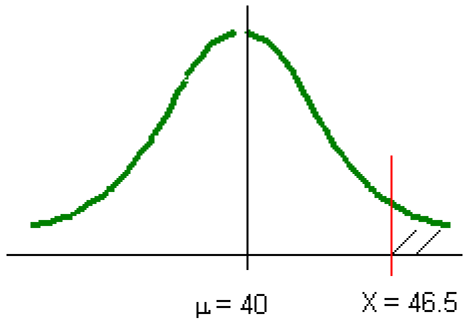


X = 29.5 □ = 40

$$z = \frac{(x - 1/2) - \mu}{\sigma} = \frac{(30 - 1/2) - 40}{4.899} = \frac{29.5 - 40}{4.899} = -2.1433 \cong -2.14$$

p(z = -2.14) =0.4838
p(x □ 30) = p(z = -2.14) +0.5 = 0.4838 + 0.5 = 0.9838

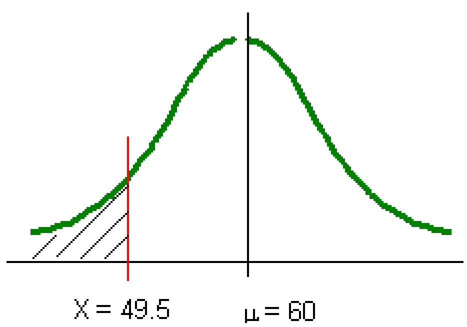
a)



$$z = \frac{(x + 1/2) - 40}{4.899} = \frac{(46 + 1/2) - 40}{4.899} = \frac{46.5 - 40}{4.899} = 1.33$$

p(z = 1.33) = 0.4082
p(x □ 46) = 0.5 – p(z = 1.33) = 0.5 – 0.4082 = 0.0918

b) n = 100
p = p(paciente no sobreviva) = 0.60
q = p(paciente sobreviva) = 1 – p = 0.40
$\mu = np = (100)(0.60) = 60$ pacientes que no se recuperan
$\sigma = \sqrt{npq} = \sqrt{100(0.60)(0.40)} = 4.899$ pacientes que no se recuperan
x = variable que nos define el número de pacientes que no sobreviven
x = 0, 1, 2, ...,100



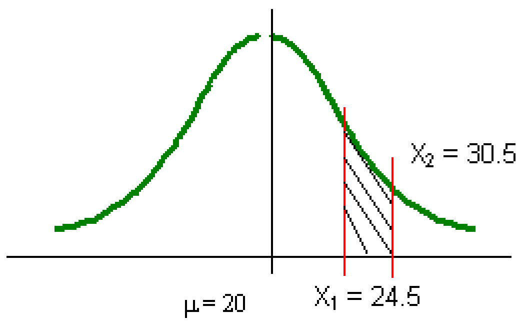
$$z = \frac{(50 - 1/2) - 60}{4.899} = \frac{49.5 - 60}{4.899} = -2.14$$

p(z = -2.14) = 0.4838

p(x □ 50) = 0.5 – p(z = -2.14) = 0.5 – 0.4838 = 0.0162

2. Una prueba de opción múltiple tiene 200 preguntas, cada una con 4 posibles respuestas, de las cuáles solo una es la correcta ¿cuál es la probabilidad de que al azar se den de 25 a 30 respuestas correctas para 80 de las 200 preguntas acerca de los cuales el estudiante no tiene conocimientos?

Solución:
n = 80
p = p(dar una contestación correcta) = 0.25
q = p(dar una contestación incorrecta) = 1 – p = 0.75
$\mu = np = 80x0.25 = 20$ preguntas contestadas correctamente
$\sigma = \sqrt{npq} = \sqrt{(80)(0.25)0.75)} = 3.8729$ preguntas contestadas correctamente
x = número de preguntas que son contestadas correctamente = 0, 1, 2,...,80



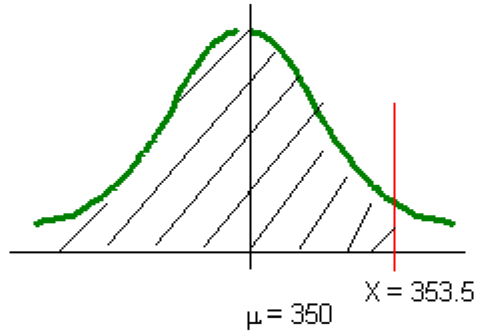
$$z_1 = \frac{(x_1 - 1/2) - \mu}{\sigma} = \frac{(25 - 1/2) - 20}{3.8729} = 1.1619 \cong 1.16$$, p(z_1 = 1.16) = 0.377

$$z_2 = \frac{(x_2 + 1/2) - \mu}{\sigma} = \frac{(30 + 1/2) - 20}{3.8729} = 2.7111 \cong 2.71$$, p(z_2 = 2.71) = 0.4966

p(25 $\leq$ x $\leq$ 30) = p(z_2) – p(z_1) = 0.4966 – 0.377 = 0.1196

3. Si 35% de los productos manufacturados en cierta línea de producción es defectuoso, ¿cuál es la probabilidad de que entre los siguientes 1000 productos manufacturados en esa línea a) menos de 354 productos sean defectuosos?, b) entre 342 y 364 productos sean defectuosos?, c)exactamente 354 productos sean defectuosos?

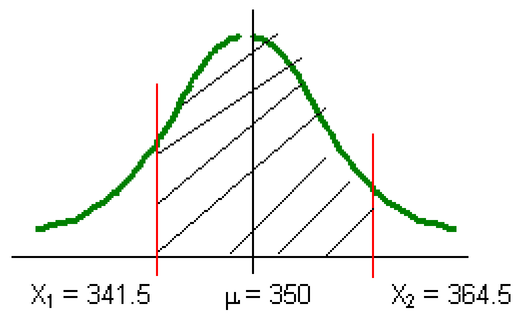
Solución:
a)n = 1000
p = p(un producto sea defectuoso) = 0.35
q = p(un producto no sea defectuoso) = 1- p = 0.65
$\mu = np = 1000(0.35) = 350$ productos defectuosos
$\sigma = \sqrt{npq} = \sqrt{(1000)(0.35(0.65)} =$ 15.0831 productos defectuosos
x = número de productos defectuosos que se manufacturan en la línea = 0, 1, 2,..., 1000



$$z=\frac{(354-1/2)-\mu}{\sigma}=\frac{(354-1/2)-350}{15.0831}=0.2320\cong 0.23$$, $p(z = 0.23) = 0.091$

$p(x \leq 354) = 0.5 + p(z = 0.23) = 0.5 + 0.091 = 0.5091$

b)

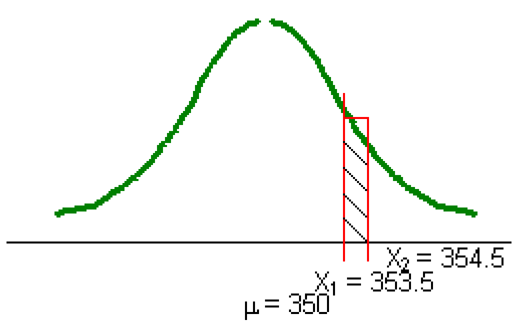


$$z_1=\frac{(342-1/2)-350}{15.0831}=-0.5635\cong -0.56$$, $p(z_1 = -0.56) = 0.2123$

$$z_2=\frac{(364+1/2)-\mu}{\sigma}=\frac{(364+1/2)-350}{15.0831}=0.9613$$ $\cong 0.96$, $p(z_2 = 0.96) = 0.3315$

$p(342 \leq x \leq 364) = p(z_1) + p(z_2) = 0.2123 + 0.3315 = 0.5438$

c)



$$z_1 = \frac{(354 - 1/2) - \mu}{\sigma} = \frac{(354 - 1/2) - 350}{15.0831} = 0.2320$$ □ 0.23, p(z_1 = 0.23) = 0.091

$$z_2 = \frac{(354 + 1/2) - \mu}{\sigma} = \frac{(354 + 1/2) - 350}{15.0831} = 0.2983 \cong 0.30$$, p(z_2= 0.30) = 0.1179

p(x = 354) = p(z_2) - p(z_1) = 0.1179 – 0.091 = 0.0269

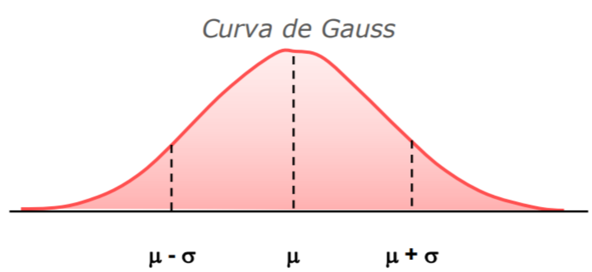
3.12 Propiedades: Media, Varianza y Desviación Estándar

Media $\mu = n \cdot p$ Varianza $\sigma^2 = n \cdot p \cdot q$ Desviación típica $\sigma = \sqrt{\sigma^2} = \sqrt{n \,.\, p \,.\, q}$

Distribución Normal Una variable aleatoria continua, x, sigue una distribución normal de media μ y desviación típica σ, y se designa por N (μ, σ), si se cumplen las siguientes condiciones:

1. La variable puede tomar cualquier valor: (-∞, +∞)

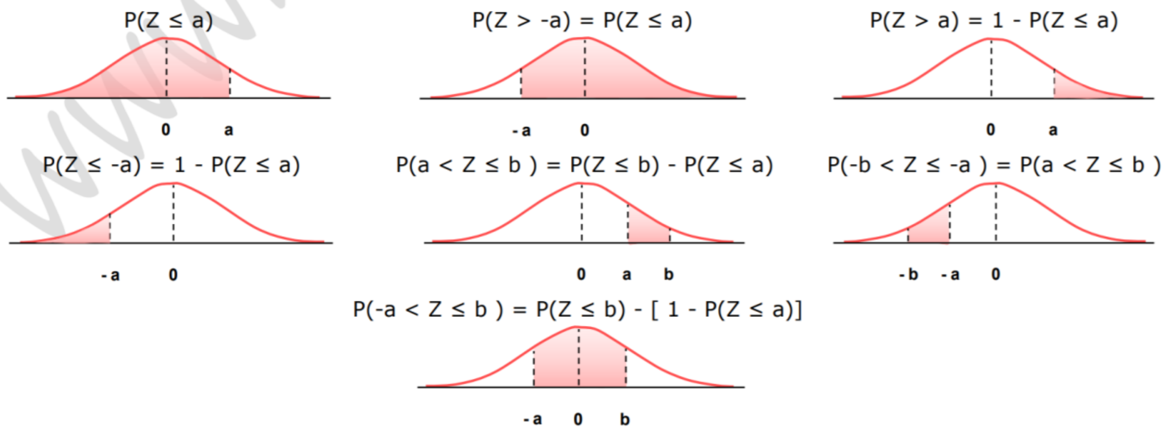
2. La función de densidad, es la expresión en términos de ecuación matemática de la curva de Gauss.



Determina la forma de cada distribución de probabilidad

- Es una curva positiva continua
- El campo de existencia es cualquier valor real, es decir, (-∞, +∞).
- Es simétrica respecto a la media.
- Tiene un máximo en la media: $(\mu, \frac{1}{\sigma\sqrt{2\pi}})$
- Crece hasta la media y decrece a partir de ella.
- En los puntos (μ - σ) y (μ - σ) presenta puntos de inflexión.
- El eje de abscisas es una asíntota de la curva.
- El **área** encerrada bajo la curva normal siempre es **1** y equivale a la **probabilidad**

3.13 Grafica



4. Muestreo

En ocasiones en que no es posible o conveniente realizar un censo (analizar a todos los elementos de una población), se selecciona una muestra, entendiendo por tal una parte representativa de la población.

El muestreo es por lo tanto una herramienta de la investigación científica, cuya función básica es determinar que parte de una población debe examinarse, con la finalidad de hacer inferencias sobre dicha población.

La muestra debe lograr una representación adecuada de la población, en la que se reproduzca de la mejor manera los rasgos esenciales de dicha población que son importantes para la investigación. Para que una muestra sea representativa, y por lo tanto útil, debe de reflejar las similitudes y diferencias encontradas en la población, es decir ejemplificar las características de ésta.

Los errores más comunes que se pueden cometer son:

1.- Hacer conclusiones muy generales a partir de la observación de sólo una parte de la Población, se denomina error de muestreo.

2.- Hacer conclusiones hacia una Población mucho más grandes de la que originalmente se tomó la muestra. Error de Inferencia.

En la estadística se usa la palabra población para referirse no sólo a personas sino a todos los elementos que han sido escogidos para su estudio y el término muestra se usa para describir una porción escogida de la población

En las investigaciones llevadas por empresarios y de la medicina se usa muestreo extensivamente en recoger información sobre poblaciones.

Cabe mencionar que para que el muestreo sea válido y se pueda realizar un estudio adecuado (que consienta no solo hacer estimaciones de la población sino estimar también los márgenes de error correspondientes a dichas estimaciones), debe cumplir ciertos requisitos. Nunca podremos estar enteramente seguros de que el resultado sea una muestra representativa, pero sí podemos actuar de manera que esta condición se alcance con una probabilidad alta.

En el muestreo, si el tamaño de la muestra es más pequeño que el tamaño de la población, se puede extraer dos o más muestras de la misma población. Al conjunto de muestras que se pueden obtener de la población se denomina espacio muestral. La variable que asocia a cada muestra su probabilidad de extracción, sigue la llamada distribución muestral.

El muestreo es un término mayormente utilizado en el campo de la estadística, la cual para poder realizar estudios a una población (que es el conjunto de elementos físicos, que presentan alguna característica en común, situados en un espacio geográfico determinado en un lapso de tiempo específico, y sobre los cuales se desea investigar), es necesario tomar una muestra de esa población dada, debido a que estas pueden ser finitas o infinitas, y aún en el caso en el que sean finitas estas pueden estar formadas por una gran cantidad de elementos lo que hace imposible un análisis completo, un ejemplo de una población tomando en cuenta todas sus características puede ser “la cantidad de maíz cosechado en la granja –La Esperanza- durante el año 2010”. Para poder estudiarlos es necesario obtener una muestra. FRECUENCIAS RELATIVAS (a posteriori o Un muestreo es el procedimiento a través del cual es seleccionada una muestra (que es un subconjunto de elementos de una población, es decir, una porción de elementos extraídos de una población previamente definida) a partir de una población. El muestreo se refiere a esa reducción de elementos que componen a un universo o población, para así poder cumplir con la investigación correspondiente.

Para poder realizar el estudio que se desea a la población (a partir de la observación de unos pocos de sus componentes), la muestra extraída debe ser representativa de ella, para que las conclusiones a las que se llegue o los resultados que se obtengan del análisis sean válidos e imparciales, esto es gracias a la técnica del muestreo.

4.1 Definición de Muestreo

El muestreo es el proceso de seleccionar un conjunto de individuos de una población con el fin de estudiarlos y poder caracterizar el total de la población.

En la referencia estadística se conoce como muestreo a la técnica para la selección de una muestra a partir de una población estadística.

Al elegir una muestra aleatoria se espera conseguir que sus propiedades sean extrapolables a la población. Este proceso permite ahorrar recursos, y a la vez obtener resultados parecidos a los que se alcanzarían si se realizase un estudio de toda la población.

4.1.1 Tipos de muestreo aleatorio, sistematizado y conglomerado

Existen diferentes criterios de clasificación de los diferentes tipos de muestreo, aunque en general pueden dividirse en dos grandes grupos: métodos de muestreo probabilísticos y métodos de muestreo no probabilísticos.

a). El Muestreo Probabilístico. Se conoce o puede calcularse la probabilidad de cada elemento, por tanto, de cada muestra posible.

El muestreo probabilístico**.** Es una técnica de muestreo en virtud de la cual las muestras son recogidas en un proceso que brinda a todos los individuos de la población las mismas oportunidades de ser seleccionados.

En esta técnica de muestreo, el investigador debe garantizar que cada individuo tenga las mismas oportunidades de ser seleccionado y esto se puede lograr si el investigador utiliza la aleatorización.

La ventaja de utilizar una muestra aleatoria es la ausencia de sesgos de muestreo y sistemáticos. Si la selección aleatoria se hace correctamente, la muestra será representativa de toda la población.

El efecto de esto es un sesgo sistemático ausente o mínimo que es la diferencia entre los resultados de la muestra y los resultados de la población. El sesgo de muestreo también se elimina ya que los sujetos son elegidos al azar.

Los métodos de muestreo probabilísticos son aquellos que se basan en el principio de equi-probabilidad. Es decir, aquellos en los que todos los individuos tienen la misma probabilidad de ser elegidos para formar parte de una muestra y, consiguientemente, todas las posibles muestras de tamaño n tienen la misma probabilidad de ser seleccionadas. Sólo estos métodos de muestreo probabilísticos nos aseguran la representatividad de la muestra extraída y son, por tanto, los más recomendables.

Es aquella rama de la estadística que apoyándose en el cálculo de probabilidades

Y a partir de estos datos muéstrales, efectúa estimaciones, decisiones, predicciones u otras generalizaciones sobre un conjunto mayor de datos. Puede definirse como aquella rama de la estadística que hace posible la estimación de una característica de una población o la toma de una decisión referente a una población, fundamentándose solo en los resultados de la muestra.

b). El Muestreo No probabilístico**.** Se desconoce o no interesa la probabilidad de cada elemento; el investigador selecciona aquella muestra que considera más representativa o que le resulte más fácil.

En la muestra no probabilística la selección de las unidades de análisis dependen de las características, criterios personales, etc., del investigador por lo que no son muy confiables en una investigación con fines científicos o tecnológicos. Este tipo de muestra adolece de fundamentación probabilística, es decir, no se tiene la seguridad de que cada unidad muestral integre a la población total en el proceso de selección de la muestra. El muestreo no probabilístico comprende los procedimientos de muestreo intencional y occidental.

a). Muestreo intencional. Es un procedimiento que permite seleccionar los casos característicos de la población limitando la muestra a estos casos. Se utiliza en situaciones en las que la población es muy variable y consecuentemente la muestra es muy pequeña.

b). El muestreo occidental consiste en tomar casos hasta que se completa el número de unidades de análisis que indica el tamaño de muestra deseado. Los anteriores procedimientos de muestreo no son recomendables para una investigación científica.

El muestreo probabilístico permite conocer la probabilidad de cada unidad de análisis, tiene que ser integrada a la muestra mediante la selección al azar. Este tipo de muestreo comprende los procedimientos de muestreo simple o al azar.

Dentro de los métodos de muestreo probabilísticos encontramos los siguientes tipos:

1. Muestreo aleatorio simple
2. Muestreo aleatorio estratificado
3. Muestreo aleatorio sistemático
4. Muestreo aleatorio conglomerados

Muestreo aleatorio simple

El muestreo aleatorio simple es la forma más fácil de muestreo probabilístico. Lo único que el investigador tiene que hacer es asegurarse de que todos los miembros de la población sean incluidos en la lista y luego seleccionar al azar el número deseado de sujetos.

Existen muchos métodos para hacer esto. Puede ser tan mecánico como sacar tiras de papel de un sombrero con nombres escritos mientras el investigador tiene los ojos vendados o puede ser tan fácil como usar un software de computadora para hacer la selección aleatoria.

Dentro de los métodos de muestreo probabilísticos encontramos los siguientes tipos: Muestreo aleatorio simple:

El procedimiento empleado es el siguiente: 1) se asigna un número a cada individuo de la población y 2) a través de algún medio mecánico (bolas dentro de una bolsa, tablas de números aleatorios, números aleatorios generados con una calculadora u ordenador, etc.) se eligen tantos sujetos como sea necesario para completar el tamaño de muestra requerido.

Este procedimiento, atractivo por su simpleza, tiene poca o nula utilidad práctica cuando la población que estamos manejando es muy grande.

Muestreo aleatorio estratificado
El muestreo aleatorio estratificado también es conocido como muestreo aleatorio proporcional. Ésta es una técnica de muestreo probabilístico en donde los sujetos son inicialmente agrupados en diferentes categorías, tales como la edad, el nivel socioeconómico o el género.

Luego, el investigador selecciona aleatoriamente la lista final de sujetos de los distintos estratos. Es importante tener en cuenta que los estratos no se superpongan.

Generalmente, los investigadores utilizan un muestreo aleatorio estratificado si quieren estudiar un determinado subgrupo dentro de la población. También es preferible el muestreo aleatorio simple porque garantiza resultados estadísticos más precisos.

El Muestreo aleatorio estratificado: Trata de obviar las dificultades que presentan los anteriores ya que simplifican los procesos y suelen reducir el error muestral para un tamaño dado de la muestra. Consiste en considerar categorías típicas diferentes entre sí (estratos) que poseen gran homogeneidad respecto a alguna característica (se puede estratificar, por ejemplo, según la profesión, el municipio de residencia, el sexo, el estado civil, etc.). Lo que se pretende con este tipo de muestreo es asegurarse de que todos los estratos de interés estarán representados adecuadamente en la muestra. Cada estrato funciona independientemente, pudiendo aplicarse dentro de ellos el muestreo aleatorio simple o el estratificado para elegir los elementos concretos que formarán parte de la muestra. En ocasiones las dificultades que plantean son demasiado grandes, pues exige un conocimiento detallado de la población. (Tamaño geográfico, sexos, edades,...).

La distribución de la muestra en función de los diferentes estratos se denomina afijación, y puede ser de diferentes tipos:

Afijación Simple: A cada estrato le corresponde igual número de elementos muéstrales.

Afijación Proporcional: La distribución se hace de acuerdo con el peso (tamaño) de la población en cada estrato.

Afijación Óptima: Se tiene en cuenta la previsible dispersión de los resultados, de modo que se considera la proporción y la desviación típica. Tiene poca aplicación ya que no se suele conocer la desviación.

El muestreo aleatorio sistemático
Este procedimiento exige, como el anterior, numerar todos los elementos de la población, pero en lugar de extraer n números aleatorios sólo se extrae uno. Se

parte de ese número aleatorio i, que es un número elegido al azar, y los elementos que integran la muestra son los que ocupa los lugares i, i+k, i+2k, i+3k,...,i+(n-1)k, es decir se toman los individuos de k en k, siendo k el resultado de dividir el tamaño de la población entre el tamaño de la muestra: k= N/n. El número i que empleamos como punto de partida será un número al azar entre 1 y k.

El riesgo este tipo de muestreo está en los casos en que se dan periodicidades en la población ya que al elegir a los miembros de la muestra con una periodicidad constante (k) podemos introducir una homogeneidad que no se da en la población. Imaginemos que estamos seleccionando una muestra sobre listas de 10 individuos en los que los 5 primeros son varones y las 5 últimos mujeres, si empleamos un muestreo aleatorio sistemático con k=10 siempre seleccionaríamos o sólo hombres o sólo mujeres, no podría haber una representación de los dos sexos.

Se puede comparar con una progresión aritmética en donde la diferencia entre dos números consecutivos es la misma. Por ejemplo, supongamos que estás en una clínica y tienes 100 pacientes.

Lo primero que tienes que hacer es elegir un número entero que sea menor que el número total de la población. Éste será tu primer sujeto, por ejemplo:

Selecciona otro número entero que será el número de individuos entre los sujetos, por ejemplo.

Tus sujetos serán los pacientes 3, 8, 13, 18, 23 y así sucesivamente.

No existe una ventaja clara en la utilización de esta técnica.

Muestreo aleatorio por conglomerados

El muestreo aleatorio por conglomerados se realiza cuando es imposible el muestreo aleatorio simple debido al tamaño de la población. Imagínate hacer un muestreo aleatorio simple cuando la población en cuestión es toda la población de Asia.

En el muestreo por conglomerados, la investigación identifica primero las fronteras, en el caso de nuestro ejemplo. Pueden ser los países de Asia.

El investigador selecciona aleatoriamente un número de áreas identificadas. Es importante que todas las áreas (países) dentro de la población tengan las mismas posibilidades de ser seleccionadas.

El investigador puede incluir todos los individuos dentro de las áreas seleccionadas o seleccionar aleatoriamente a los sujetos de las áreas identificadas.

El Muestreo aleatorio por conglomerados: Son los métodos presentados hasta ahora están pensados para seleccionar directamente los elementos de la población, es decir, que las unidades muéstrales son los elementos de la población.

En el muestreo por conglomerados la unidad muestral es un grupo de elementos de la población que forman una unidad, a la que llamamos conglomerado. Las unidades hospitalarias, los departamentos universitarios, una caja de determinado producto, etc., son conglomerados naturales. En otras ocasiones se pueden utilizar conglomerados no naturales como, por ejemplo, las urnas electorales. Cuando los conglomerados son áreas geográficas suele hablarse de "muestreo por áreas".

El muestreo por conglomerados consiste en seleccionar aleatoriamente un cierto número de conglomerados (el necesario para alcanzar el tamaño muestral establecido) y en investigar después todos los elementos pertenecientes a los conglomerados elegidos.

El muestreo no probabilístico en general se selecciona a los sujetos siguiendo determinados criterios procurando en la medida de lo ´posible, que la muestra sea representativa.

Entre los métodos de muestreo no probabilísticos más utilizados en investigación encontramos:

1. Muestreo por cuotas
2. Muestreo intencional o de conveniencia
3. Bola de nieve
4. Muestreo discrecional

4.2 CONCEPTO DE DISTRIBUCIÓN DE MUESTREO DE LA MEDIA

Distribución de Muestreo de la Media**.**

Un estadístico está distribuido normalmente cuando la muestra que se toma es grande, conocido como el teorema del límite central. Cuando el tamaño de la muestra es grande y la varianza de la población es conocida se toma la distribución normal estándar como estadístico de prueba. Pero cuando el tamaño de la muestra no es grande y a su vez se desconoce la varianza de la población, es aconsejable aplicar la Distribución de students. Estas condiciones se conocen como el TEOREMA DEL LIMITE CENTRAL.

4.2.1 DISTRIBUCIÓN MUESTRAL DE LA MEDIA CON VARIANZA CONOCIDA Y DESCONOCIDA

Una muestra de tamaño n, extraída de una población cuya media es μ y varianza σ^2, constituida por un conjunto de variables aleatorias independientes X1, X2,…, Xn es una muestra aleatoria y los n valores que toma X serán los datos que conforman la muestra.

Tomando en cuenta lo dicho en la introducción, supongamos que se han extraído k muestras aleatorias de la población de tamaño N. Si $\bar{X}$1, $\bar{X}$2 ,…, $\bar{X}$k son las medias muéstrales de cada una de las muestras, entonces podemos afirmar que $\bar{X}$, media aritmética de las medias muéstrales, es una variable aleatoria definida como Media muestral.

Siendo $\bar{X}$ una variable aleatoria, entonces debe tener una distribución de probabilidad la cual estará definida por su media $\mu_{\bar{X}}$ y su varianza $\sigma^2_{\bar{X}}$, donde

$$\mu_{\overline{X}} = E(\overline{X}) = E\left(\frac{\sum x_i}{n}\right) = \frac{1}{n}E\left(\sum x_i\right) = \frac{1}{n}n\mu = \mu$$

$$\sigma_{\overline{X}}^2 = V(\overline{X}) = V\left(\frac{\sum x_i}{n}\right) = \frac{1}{n^2}V\left(\sum x_i\right) = \frac{1}{n^2}n\sigma^2 = \frac{\sigma^2}{n}$$

Nota 1

Siendo $\bar{X}$ una variable aleatoria y tiene una distribución de probabilidad, es natural preguntarnos: ¿Se puede calcular P($\bar{X}$ ≤ k)?

El siguiente teorema nos autorizará el uso de la distribución normal para resolver problemas como se plantea en la pregunta, bajo ciertas condiciones.

Teorema del Límite central

Sea X1, X2, …, Xn una muestra aleatoria extraída de una población de parámetros μ y σ2. Si $\bar{X}$ es la media muestral, entonces Z= ($\bar{X}$ - μ)/(σ/ √n) es una variable normal estándar, siempre que n sea suficientemente grande (n ≥ 30).

Nota 2:

$$\text{Según esto } P(\overline{X} \leq k) = P\left(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \leq \frac{k-\mu}{\sigma/\sqrt{n}}\right) = P\left(Z \leq \frac{k-\mu}{\sigma/\sqrt{n}}\right), \text{ cuando } n \geq 30.$$

Cuando la varianza poblacional es conocida

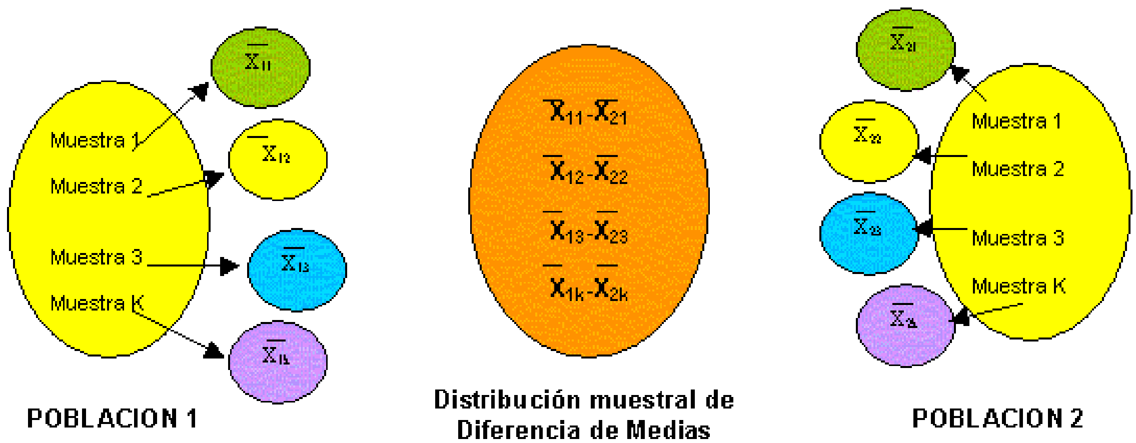
La Distribución muestral de la media muestral ¯X, cuando la varianza poblacional σ^2 es conocida, aplicando el Teorema del Límite Central, será normal con $\mu_{\bar{X}}$ = μ y $\sigma^2_{\bar{X}}$ = σ2/n, según la deducción realizada líneas arriba. Luego $\bar{X}$ → N(μ , σ2/n).

Cuando la varianza poblacional es desconocida

Si la varianza poblacional es desconocida y la población desde donde se extrae la muestra es normal, entonces la variable T=($\bar{X}$ - μ)/(s/√n) → t(n-1). Esto es, cuando la varianza poblacional no sea conocida, usaremos la distribución t de Student para resolver el problema.

4.2.2 DISTRIBUCIÓN MUESTRAL DE LA DIFERENCIA ENTRE DOS MEDIAS CON VARIANZA CONOCIDA Y DESCONOCIDA

Distribución muestral de la diferencia entre dos medias con σ, conocida y desconocida. Distribución Muestral de Diferencia de Medias Suponga que se tienen dos poblaciones distintas, la primera con media μ1 y desviación estándar σ1, y la segunda con media μ2 y desviación estándar σ2. Más aún, se elige una muestra aleatoria de tamaño n1 de la primera población y una muestra independiente aleatoria de tamaño n2 de la segunda población; se calcula la media muestral para cada muestra y la diferencia entre dichas medias. La colección de todas esas diferencias se llama distribución muestral de las diferencias entre medias o la distribución muestral del estadístico $\bar{x}_1 - \bar{x}_2$.



La distribución es aproximadamente normal para n1$\geq$30 y n2$\geq$30. Si las poblaciones son normales, entonces la distribución muestral de medias es normal sin importar los tamaños de las muestras.

En ejercicios anteriores se había demostrado que $\mu = \mu_{\bar{x}}$ y que $\sigma_{\bar{x}} = \sigma / \sqrt{n}$, por lo que no es difícil deducir que $\mu_{\bar{x}1} - \mu_{\bar{x}2} = \mu_1 - \mu_2$ y que $\sigma_{\bar{x}1-\bar{x}2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$.

La fórmula que se utilizará para el cálculo de probabilidad del estadístico de diferencia de medias 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.- En un estudio para comparar los pesos promedio de niños y niñas de sexto grado en una escuela primaria se usará una muestra aleatoria de 20 niños y otra de 25 niñas. Se sabe que tanto para niños como para niñas los pesos siguen una distribución normal. El promedio de los pesos de todos los niños de sexto grado de esa escuela es de 100 libras y su desviación estándar es de 14.142, mientras que el promedio de los pesos de todas las niñas del sexto grado de esa escuela es de 85 libras y su desviación estándar es de 12.247 libras. Si $\bar{x}_1$ representa el promedio de los pesos de 20 niños y $\bar{x}_2$ es el promedio de los pesos de una muestra de 25 niñas, encuentre la probabilidad de que el promedio de los pesos de los 20 niños sea al menos 20 libras más grande que el de las 25 niñas.

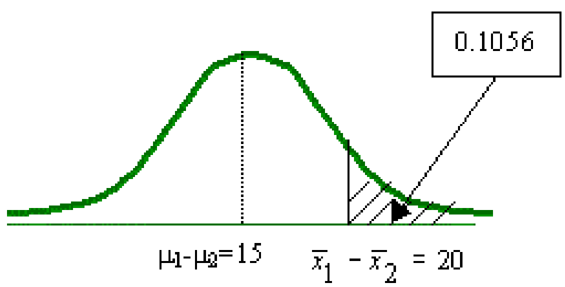
Solución:

Datos:
μ_1 = 100 libras
μ_2 = 85 libras
σ_1 = 14.142 libras
σ_2 = 12.247 libras

n_1 = 20 niños
n_2 = 25 niñas
$p(\bar{x}_1 - \bar{x}_2 \rangle 20)$ = ?

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{20 - (100 - 85)}{\sqrt{\frac{(14.142)^2}{20} + \frac{(12.247)^2}{25}}} = 1.25$$

Por lo tanto, la probabilidad de que el promedio de los pesos de la muestra de niños sea al menos 20 libras más grande que el de la muestra de las niñas es 0.1056.



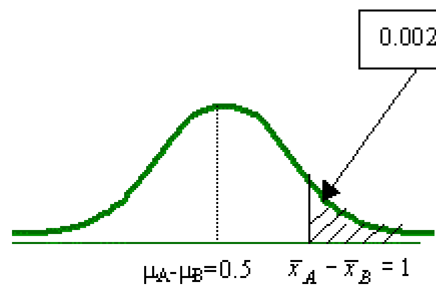
Ejemplo:

2.- Uno de los principales fabricantes de televisores compra los tubos de rayos catódicos a dos compañías. Los tubos de la compañía A tienen una vida media de 7.2 años con una desviación estándar de 0.8 años, mientras que los de la B tienen una vida media de 6.7 años con una desviación estándar de 0.7. Determine la probabilidad de que una muestra aleatoria de 34 tubos de la compañía A tenga una vida promedio de al menos un año más que la de una muestra aleatoria de 40 tubos de la compañía B.

Solución:
Datos:
μ_A = 7.2 años
μ_B = 6.7 años
σ_A = 0.8 años

σ_B = 0.7 años
n_A = 34 tubos
n_B = 40 tubos
$p(\bar{x}_A - \bar{x}_B \rangle 1)$ = ?

$$z = \frac{(\bar{x}_A - \bar{x}_B) - (\mu_A - \mu_B)}{\sqrt{\frac{\sigma_A^{\,2}}{n_A} + \frac{\sigma_B^{\,2}}{n_B}}} = \frac{1 - (7.2 - 6.7)}{\sqrt{\frac{(0.8)^2}{34} + \frac{(0.7)^2}{40}}} = 2.84$$



Ejemplo:

3.- Se prueba el rendimiento en km/L de 2 tipos de gasolina, encontrándose una desviación estándar de 1.23km/L para la primera gasolina y una desviación estándar de 1.37km/L para la segunda gasolina; se prueba la primera gasolina en 35 autos y la segunda en 42 autos.

¿Cuál es la probabilidad de que la primera gasolina, de un rendimiento promedio mayor de 0.45km/L, que la segunda gasolina?

¿Cuál es la probabilidad de que la diferencia en rendimientos promedio se encuentre entre 0.65 y 0.83km/L a favor de la gasolina 1?

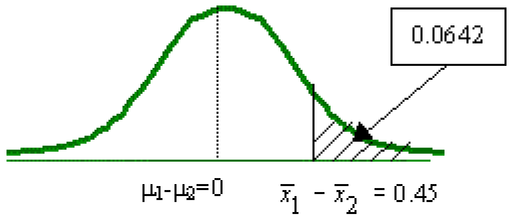
Solución:

En este ejercicio no se cuenta con los parámetros de las medias en ninguna de las dos poblaciones, por lo que se supondrán que son iguales.

Datos:

σ_1 = 1.23 Km/Lto
σ_2 = 1.37 Km/Lto
n_1 = 35 autos
n_2 = 42 autos
$p(\bar{x}_1 - \bar{x}_2 \rangle 0.45)$ = ?

$$z=\frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}=\frac{0.45-0}{\sqrt{\frac{(1.23)^2}{35}+\frac{(1.37)^2}{42}}}=1.52$$

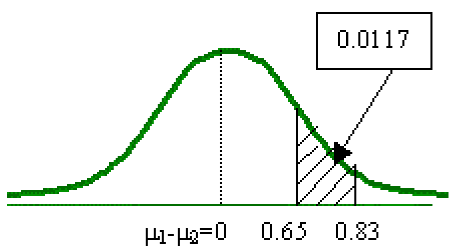


$$p(0.65\leq\bar{x}_1-\bar{x}_2\leq 0.83)=$$

?

$$z=\frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}=\frac{0.65-0}{\sqrt{\frac{(1.23)^2}{35}+\frac{(1.37)^2}{42}}}=2.19$$

$$z=\frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}=\frac{0.83-0}{\sqrt{\frac{(1.23)^2}{35}+\frac{(1.37)^2}{42}}}=2.80$$



La probabilidad de que la diferencia en rendimientos promedio en las muestras se encuentre entre 0.65 y 0.83 Km/Lto a favor de la gasolina 1 es de 0.0117.

4.2.3 DISTRIBUCIÓN MUESTRAL DE LA PROPORCIÓN

Un estimador puntual de la proporción P en un experimento binomial está dado por la estadística P=X/N, donde x representa de la muestra p =x/n se utilizará como estimador puntual del parámetro P.

Si no se espera que la proporción P desconocida esté demasiado cerca de 0 ó de 1, se puede establecer un intervalo de confianza para P al considerar la distribución muestral de proporciones.

$$z = \frac{p - P}{\sqrt{\frac{Pq}{n}}}$$

Al despejar P de esta ecuación nos queda:

$$P = p \pm z\sqrt{\frac{Pq}{n}}$$

En este despeje podemos observar que se necesita el valor del parámetro P y es precisamente lo que queremos estimar, por lo que lo sustituiremos por la proporción de la muestra *p* siempre y cuando el tamaño de muestra no sea pequeño.

$$P = p \pm z\sqrt{\frac{pq}{n}}$$

Cuando n es pequeña y la proporción desconocida P se considera cercana a 0 ó a 1, el procedimiento del intervalo de confianza que se establece aquí no es confiable, por tanto, no se debe utilizar. Para estar seguro, se debe requerir que np ó nq sea mayor o igual a 5.

El error de estimación será la diferencia absoluta entre p y P, y podemos tener el nivel de confianza de que esta diferencia no excederá $z\sqrt{\frac{pq}{n}}$.

Ejemplos:

Un fabricante de reproductores de discos compactos utiliza un conjunto de pruebas amplias para evaluar la función eléctrica de su producto. Todos los reproductores de discos compactos deben pasar todas las pruebas antes de venderse. Una muestra aleatoria de 500 reproductores tiene como resultado 15 que fallan en una o más pruebas. Encuentre un intervalo de confianza de 90% para la proporción de los reproductores de discos compactos de la población que no pasan todas las pruebas.

Solución:
n=500
p = 15/500 = 0.03
z(0.90) = 1.645

$$P = p \pm z\sqrt{\frac{pq}{n}} = 0.03 \pm (1.645)\sqrt{\frac{(0.03)(0.97)}{500}}$$

0.0237<P<0.0376

Se sabe con un nivel de confianza del 90% que la proporción de discos defectuosos que no pasan la prueba en esa población está entre 0.0237 y 0.0376.

En una muestra de 400 pilas tipo B fabricadas por la Everlast Company, se encontraron 20 defectuosas. Si la proporción p de pilas defectuosas en esa muestra se usa para estimar P, que vendrá a ser la proporción verdadera de todas las pilas defectuosas tipo B fabricadas por la Everlast Company, encuentre el máximo error de estimación ε tal que se pueda tener un 95% de confianza en que *P* dista menos de ε de *p.*

Solución:
p=x/n = 20/400=0.05
z(0.95)=1.96

$$\varepsilon = z\sqrt{\frac{pq}{n}} = 1.96\sqrt{\frac{(0.05)(0.95)}{400}} = 0.021$$

Si p=0.05 se usa para estimar P, podemos tener un 95% de confianza en que P dista menos de 0.021 de p. En otras palabras, si p=0.05 se usa para estimar P, el error máximo de estimación será aproximadamente 0.021 con un nivel de confianza del 95%.

Para calcular el intervalo de confianza se tendría:

$$p \pm \varepsilon = 0.05 \pm 0.021$$

Esto da por resultado dos valores, (0.029, 0.071). Con un nivel de confianza del 95% se sabe que la proporción de pulas defectuosas de esta compañía está entre 0.029 y 0.071.

Si se requiere un menor error con un mismo nivel de confianza sólo se necesita aumentar el tamaño de la muestra.

En un estudio de 300 accidentes de automóvil en una ciudad específica, 60 tuvieron consecuencias fatales. Con base en esta muestra, construya un intervalo del 90% de confianza para aproximar la proporción de todos los accidentes automovilísticos que en esa ciudad tienen consecuencias fatales.

Solución:
P= 60/300 = 0.20
Z(0.90) = 1.645

$$P = p \pm z\sqrt{\frac{pq}{n}} = 0.20 \pm (1.645)\sqrt{\frac{(0.20)(0.80)}{300}} = 0.20 \pm 0.038$$

0.162<P<0.238

4.2.4 DISTRIBUCIÓN MUESTRAL DE LA DIFERENCIA DE DOS PROPORCIONES

Muchas aplicaciones involucran poblaciones de datos cualitativos que deben compararse utilizando proporciones o porcentajes. A continuación, se citan algunos ejemplos:

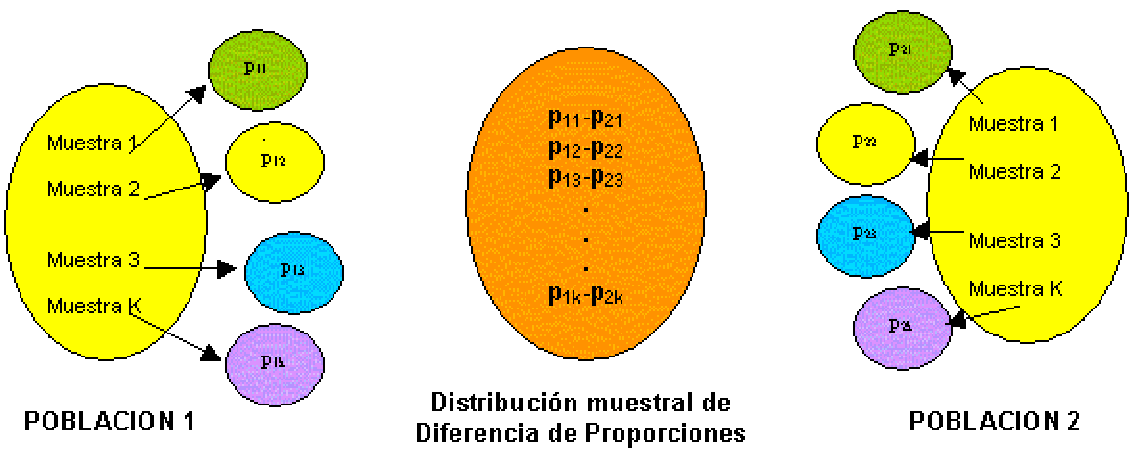
Educación. - ¿Es mayor la proporción de los estudiantes que aprueban matemáticas que las de los que aprueban inglés?

Medicina. - ¿Es menor el porcentaje de los usuarios del medicamento A que presentan una reacción adversa que el de los usuarios del fármaco B que también presentan una reacción de ese tipo?

Administración. - ¿Hay diferencia entre los porcentajes de hombres y mujeres en posiciones gerenciales.

Ingeniería. - ¿Existe diferencia entre la proporción de artículos defectuosos que genera la máquina A a los que genera la máquina B?

Cuando el muestreo procede de dos poblaciones binomiales y se trabaja con dos proporciones muéstrales, la distribución muestral de diferencia de proporciones es aproximadamente normal para tamaños de muestra grande ($n1p1 \geq 5$, $n1q1 \geq 5$, $n2p2 \geq 5$ y $n2q2 \geq 5$). Entonces p1 y p2 tienen distribuciones muéstrales aproximadamente normales, así que su diferencia p1-p2 también tiene una distribución muestral aproximadamente normal.



Cuando se estudió a la distribución muestral de proporciones se comprobó que $P = \mu_p$ y que $\sigma_p = \sqrt{\frac{Pq}{n}}$, por lo que no es difícil deducir que $\mu_{p1} - \mu_{p2} = P_1 - P_2$ y que $\sigma_{p1-p2} = \sqrt{\frac{P_1 q_1}{n_1} + \frac{P_2 q_2}{n_2}}$.

La fórmula que se utilizará para el cálculo de probabilidad del estadístico de diferencia de proporciones es:

$$z = \frac{(p_1 - p_2) - (P_1 - P_2)}{\sqrt{\frac{P_1 q_1}{n_1} + \frac{P_2 q_2}{n_2}}}$$

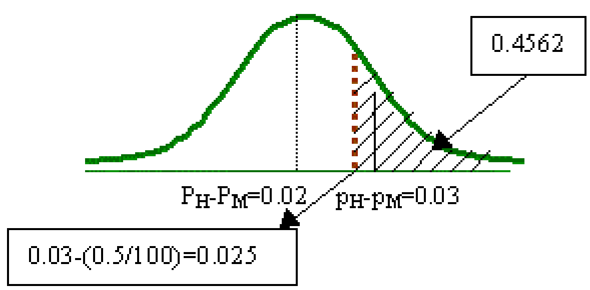
Ejemplos:

1.- Los hombres y mujeres adultos radicados en una ciudad grande del norte difieren en sus opiniones sobre la promulgación de la pena de muerte para personas culpables de asesinato. Se cree que el 12% de los hombres adultos están a favor de la pena de muerte, mientras que sólo 10% de las mujeres adultas lo están. Si se pregunta a dos muestras aleatorias de 100 hombres y 100 mujeres su opinión sobre la promulgación de la pena de muerte, determine la probabilidad de que el porcentaje de hombres a favor sea al menos 3% mayor que el de las mujeres.

Solución:
Datos:
$P_H = 0.12$
$P_M = 0.10$
$n_H = 100$
$n_M = 100$
$p(p_H - p_M \geq 0.03) = ?$

Se recuerda que se está incluyendo el factor de corrección de 0.5 por ser una distribución binomial y se está utilizando la distribución normal.

$$z = \frac{(p_H - p_M) - (P_H - P_M)}{\sqrt{\frac{P_H q_H}{n_H} + \frac{P_M q_M}{n_M}}} = \frac{0.025 - (0.12 - 0.10)}{\sqrt{\frac{(0.12)(0.88)}{100} + \frac{(0.10)(0.90)}{100}}} = 0.11$$



Se concluye que la probabilidad de que el porcentaje de hombres a favor de la pena de muerte, al menos 3% mayor que el de mujeres es de 0.4562.

2.- Una encuesta del Boston College constó de 320 trabajadores de Michigan que fueron despedidos entre 1979 y 1984, encontró que 20% habían estado sin trabajo durante por lo menos dos años. Supóngase que tuviera que seleccionar otra muestra aleatoria de 320 trabajadores de entre todos los empleados despedidos entre 1979 y 1984. ¿Cuál sería la probabilidad de que su porcentaje muestral de trabajadores sin empleo durante por lo menos dos años, difiera del porcentaje obtenido en la encuesta de Boston College, en 5% o más?

Solución:
En este ejercicio se cuenta únicamente con una población, de la cual se están extrayendo dos muestras y se quiere saber la probabilidad de la diferencia de los porcentajes en esas dos muestras, por lo que se debe de utilizar la distribución muestral de proporciones con P1= P2, ya que es una misma población.

Otra de las situaciones con la cual nos topamos es que desconocemos la proporción de trabajadores despedidos entre 1979 y 1984 que estuvieron desempleados por un período de por lo menos dos años, sólo se conoce la p1= 0.20 ya que al tomar una muestra de 320 trabajadores se observó esa proporción.

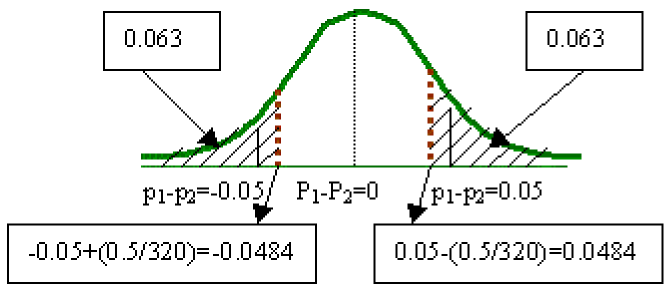
3.-En la fórmula de la distribución muestral de proporciones para el cálculo de probabilidad se necesita saber las proporciones de las poblaciones, las cuales en este ejercicio las desconocemos, por lo que se utilizará el valor de 0.20 como una estimación puntual de P. En el siguiente tema se abordará el tema de estimación estadística y se comprenderá por qué estamos utilizando de esa manera el dato.

También debe de comprenderse la pregunta que nos hace este problema, ¿cuál sería la probabilidad de que su porcentaje muestral de trabajadores sin empleo durante por lo menos dos años, difiera del porcentaje obtenido en la encuesta de Boston College, en 5% o más?, la palabra difiera quiere decir que puede existir una diferencia a favor de la muestra uno, o a favor de la muestra dos, por lo que se tendrán que calcular dos áreas en la distribución y al final sumarlas.

Datos:
$p_1 = 0.20$
$n_1 = 320$ trabajadores
$n_2 = 320$ trabajadores
$P_1 = P_2$

$$z = \frac{(p_1 - p_2) - (P_1 - P_2)}{\sqrt{\frac{P_1 q_1}{n_1} + \frac{P_2 q_2}{n_2}}} = \frac{0.0484 - 0}{\sqrt{\frac{(0.20)(0.80)}{320} + \frac{(0.20)(0.80)}{320}}} = 1.53$$

$$z=\frac{(p_1-p_2)-(P_1-P_2)}{\sqrt{\frac{P_1q_1}{n_1}+\frac{P_2q_2}{n_2}}}=\frac{-0.0484-0}{\sqrt{\frac{(0.20)(0.80)}{320}+\frac{(0.20)(0.80)}{320}}}=-1.53$$



$$p(-0.05\geq p_1-p_2\geq 0.05)=0.063+0.063=0.1260$$

La probabilidad de que su proporción muestral de trabajadores sin empleo durante por lo menos dos años, difiera del porcentaje obtenido en la encuesta de Boston College, en 0.05 o más es de 0.1260.

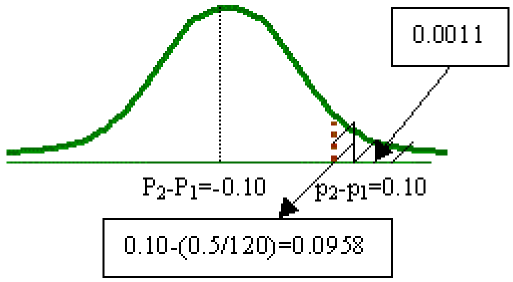
4.- Se sabe que 3 de cada 6 productos fabricados por la máquina 1 son defectuosos y que 2 de cada 5 objetos fabricados por la máquina 2 son defectuosos; se toman muestras de 120 objetos de cada máquina:

¿cuál es la probabilidad de que la proporción de artículos defectuosos de la máquina 2 rebase a la máquina 1 en por lo menos 0.10?

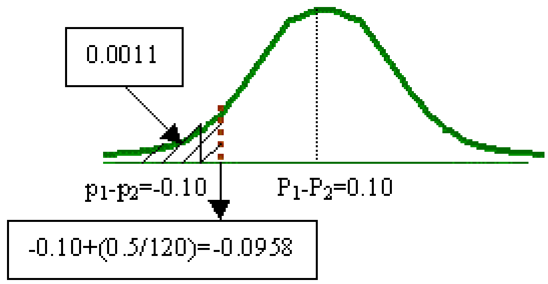
¿cuál es la probabilidad de que la proporción de artículos defectuosos de la máquina 1 rebase a la máquina 2 en por lo menos 0.15?

Solución:
Datos:
$P_1 = 3/6 = 0.5$
$P_2 = 2/5 = 0.4$
$n_1 = 120$ objetos
$n_2 = 120$ objetos
$p(p_2-p_1\geq 0.10) = ?$

$$z=\frac{(p_2-p_1)-(P_2-P_1)}{\sqrt{\frac{P_1q_1}{n_1}+\frac{P_2q_2}{n_2}}}=\frac{0.0958-(-0.10)}{\sqrt{\frac{(0.50)(0.50)}{120}+\frac{(0.40)(0.60)}{120}}}=3.06$$



Otra manera de hacer este ejercicio es poner P_1-P_2:

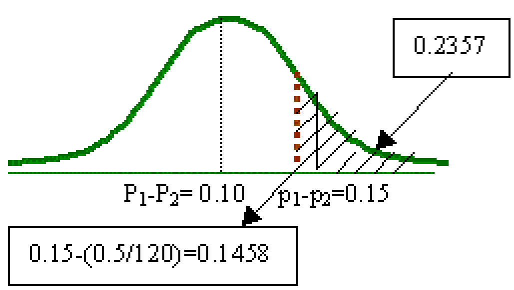


$$z=\frac{(p_1-p_2)-(P_1-P_2)}{\sqrt{\frac{P_1q_1}{n_1}+\frac{P_2q_2}{n_2}}}=\frac{-0.0958-0.10}{\sqrt{\frac{(0.50)(0.50)}{120}+\frac{(0.40)(0.60)}{120}}}=-3.06$$

La probabilidad de que exista una diferencia de proporciones de artículos defectuosos de por lo menos 10% a favor de la máquina 2 es de 0.0011.

$p(p_1-p_2\geq$
$0.15)=?$

$$z=\frac{(p_1-p_2)-(P_1-P_2)}{\sqrt{\frac{P_1q_1}{n_1}+\frac{P_2q_2}{n_2}}}=\frac{0.1458-0.10}{\sqrt{\frac{(0.50)(0.50)}{120}+\frac{(0.40)(0.60)}{120}}}=0.72$$



La probabilidad de que exista una diferencia de proporciones de artículos defectuosos de por lo menos 15% a favor de la máquina 1 es de 0.2357.

5.- Si se tienen dos poblaciones con medias μ1 y μ2 y varianzas σ 12 y σ 22, respectivamente, un estimador puntual de la diferencia entre μ1 y μ2 está dado por la estadística $\bar{x}_1 - \bar{x}_2$. Por tanto. Para obtener una estimación puntual de μ1- μ2, se seleccionan dos muestras aleatorias independientes, una de cada población, de tamaño n1 y n2, se calcula la diferencia $\bar{x}_1 - \bar{x}_2$, de las medias muéstrales.

Recordando a la distribución muestral de diferencia de medias:

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Al despejar de esta ecuación μ_1- μ_2 se tiene:

$$\mu_1 - \mu_2 = (\bar{x}_1 - \bar{x}_2) \pm z\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

En el caso en que se desconozcan las varianzas de la población y los tamaños de muestra sean mayores a 30 se podrá utilizar la varianza de la muestra como una estimación puntual.

6. Se lleva a cabo un experimento en que se comparan dos tipos de motores, A y B. Se mide el rendimiento en millas por galón de gasolina. Se realizan 50 experimentos con el motor tipo A y 75 con el motor tipo B. La gasolina que se utiliza y las demás condiciones se mantienen constantes. El rendimiento promedio de gasolina para el motor A es de 36 millas por galón y el promedio para el motor B es 24 millas por galón. Encuentre un intervalo de confianza de 96% sobre la diferencia promedio real para los motores A y B. Suponga que las desviaciones estándar poblacionales son 6 y 8 para los motores A y B respectivamente.

Solución:

Es deseable que la diferencia de medias sea positiva por lo que se recomienda restar la media mayor menos la media menor. En este caso será la media del motor B menos la media del motor A.

El valor de z para un nivel de confianza del 96% es de 2.05.

$$\mu_B - \mu_A = (\bar{x}_B - \bar{x}_A) \pm z\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}} = (42 - 36) \pm 2.05\sqrt{\frac{36}{50} + \frac{64}{75}}$$

$$3.43 < \mu_B - \mu_A < 8.57$$

La interpretación de este ejemplo sería que con un nivel de confianza del 96% la diferencia del rendimiento promedio esta entre 3.43 y 8.57 millas por galón a favor del motor B. Esto quiere decir que el motor B da más rendimiento promedio que el motor A, ya que los dos valores del intervalo son positivos.

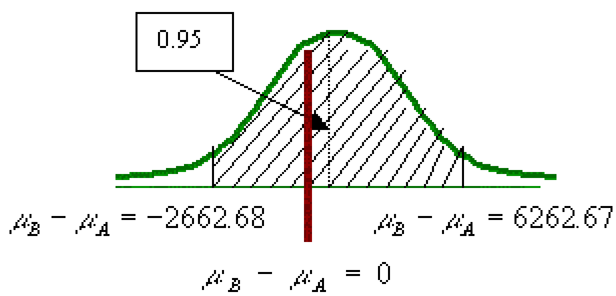
7. Una compañía de taxis trata de decidir si comprar neumáticos de la marca A o de la B para su flotilla de taxis. Para estimar la diferencia de las dos marcas, se lleva a cabo un experimento utilizando 12 de cada marca. Los neumáticos se utilizan hasta que se desgastan, dando como resultado promedio para la marca A 36,300 kilómetros y para la marca B 38,100 kilómetros. Calcule un intervalo de confianza de 95% para la diferencia promedio de las dos marcas, si se sabe que las poblaciones se distribuyen de forma aproximadamente normal con desviación estándar de 5000 kilómetros para la marca A y 6100 kilómetros para la marca B.

Solución:

$$\mu_B - \mu_A = (\bar{x}_B - \bar{x}_A) \pm z\sqrt{\frac{\sigma_A^{\ 2}}{n_A} + \frac{\sigma_B^{\ 2}}{n_B}} = (38100 - 36300) \pm 1.96\sqrt{\frac{5000^2}{12} + \frac{6100^2}{12}}$$

$$-2662.68 < \mu_B - \mu_A < 6262.67$$

Gráficamente:



Como el intervalo contiene el valor "cero", no hay razón para creer que el promedio de duración del neumático de la marca B es mayor al de la marca A, pues el cero nos está indicando que pueden tener la misma duración promedio.

4.3 TEOREMA DE LÍMITE CENTRAL

Sea X1, X2, …, Xn una muestra aleatoria extraída de una población de parámetros μ y σ2. Si $\bar{X}$ es la media muestral, entonces Z= ($\bar{X}$ - μ)/(σ/ √n) es una variable normal estándar, siempre que n sea suficientemente grande (n ≥ 30).

Nota 2:

$$\text{Según esto } P(\overline{X} \leq k) = P\left(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \leq \frac{k-\mu}{\sigma/\sqrt{n}}\right) = P\left(Z \leq \frac{k-\mu}{\sigma/\sqrt{n}}\right), \text{ cuando } n \geq 30.$$

Cuando la varianza poblacional es conocida

La Distribución muestral de la media muestral ¯X, cuando la varianza poblacional σ^2 es conocida, aplicando el Teorema del Límite Central, será normal con $\mu_{\overline{X}} = \mu$ y $\sigma^2_{\overline{X}} = \sigma^2/n$, según la deducción realizada líneas arriba. Luego $\overline{X} \rightarrow N(\mu, \sigma^2/n)$.

El Teorema del límite central establece que bajo ciertas condiciones (como pueden ser independientes e idénticamente distribuidas con varianza finita), la suma de un gran número de variables aleatorias se distribuye aproximadamente como una normal. La importancia práctica del Teorema del límite central es que la función de distribución de la normal puede usarse como aproximación de algunas otras funciones de distribución. Por ejemplo:

Ejercicios:

1. La renta media de los habitantes de un país se distribuye uniformemente entre 4,0 millones ptas. y 10,0 millones ptas. Calcular la probabilidad de que al seleccionar al azar a 100 personas la suma de sus rentas supere los 725 millones ptas.

Cada renta personal es una variable independiente que se distribuye según una función uniforme. Por ello, a la suma de las rentas de 100 personas se le puede aplicar el Teorema Central del Límite.

La media y varianza de cada variable individual es:

m = (4 + 10) / 2 = 7
s^2 = (10 - 4)^2 / 12 = 3

Por tanto, la suma de las 100 variables se distribuye según una normal cuya media y varianza son:

Media: n * m = 100 * 7 = 700
Varianza: n * s^2 = 100 * 3 = 300

Para calcular la probabilidad de que la suma de las rentas sea superior a 725 millones ptas, comenzamos por calcular el valor equivalente de la variable normal tipificada:

$$Y = \frac{725 - 700}{17{,}3} = 1{,}44$$

Luego:
P (X > 725) = P (Y > 1,44) = 1 - P (Y < 1,44) = 1 - 0,9251 = 0,0749

Es decir, la probabilidad de que la suma de las rentas de 100 personas seleccionadas al azar supere los 725 millones de pesetas es tan sólo del 7,49%

2. En una asignatura del colegio la probabilidad de que te saquen a la pizarra en cada clase es del 10%. A lo largo del año tienes 100 clases de esa asignatura. ¿Cuál es la probabilidad de tener que salir a la pizarra más de 15 veces?

Se vuelve a aplicar el Teorema Central del Límite.

Salir a la pizarra es una variable independiente que sigue el modelo de distribución de Bernoulli:

"Salir a la pizarra", le damos el valor 1 y tiene una probabilidad del 0,10

"No salir a la pizarra", le damos el valor 0 y tiene una probabilidad del 0,9

La media y la varianza de cada variable independiente es:

m = 0,10
s^2 = 0,10 * 0,90 = 0,09

Por tanto, la suma de las 100 variables se distribuye según una normal cuya media y varianza son:

Media: n * m = 100 * 0,10 = 10
Varianza: n * s^2 = 100 * 0,09 = 9

Para calcular la probabilidad de salir a la pizarra más de 15 veces, calculamos el valor equivalente de la variable normal tipificada:

$$Y = \frac{15 - 10}{3,0} = 1,67$$

Luego:

P (X > 15) = P (Y > 1,67) = 1 - P (Y < 1,67) = 1 - 0,9525 = 0,0475

Es decir, la probabilidad de tener que salir más de 15 veces a la pizarra a lo largo del curso es tan sólo del 4,75% (¡¡¡ ánimo !!!, no es tan grave)

Si una población tiene media μ y desviación típica σ, y tomamos muestras de tamaño n (n > 30, ó cualquier tamaño si la población es "normal"), las medias de estas muestras siguen aproximadamente la distribución:

$$N\left(\mu,\ \frac{\sigma}{\sqrt{n}}\right)$$

Consecuencias
1. Permite averiguar la probabilidad de que la media de una muestra concreta esté en un cierto intervalo.

2. Permite calcular la probabilidad de que la suma de los elementos de una muestra esté, a priori, en un cierto intervalo.

$$\sum_{i=1}^{n} x_i \rightarrow \quad N\left(n\mu,\ \sigma\sqrt{n}\right)$$

3. Las bolsas de sal envasadas por una máquina tienen μ = 500 g y σ = 35 g. Las bolsas se empaquetaron en cajas de 100 unidades.

Calcular la probabilidad de que la media de los pesos de las bolsas de un paquete sea menor que 495 g.

$$N\left(500,\ \frac{35}{\sqrt{100}}\right) \qquad\qquad N(500,\ 3.5)$$

$$p\left(\bar{x} < 495\right) = p\left(z < \frac{495-500}{3.5}\right) = p\left(z < -1.43\right) = p\left(z > 1.43\right) =$$

$$= 1 - p\left(z \leq 1.43\right) = 0.0764$$

4. Calcular la probabilidad de que una caja 100 de bolsas pese más de 51 kg.

$$N\left(500 \cdot 100,\ 35\sqrt{100}\right) \qquad\qquad N(50000,\ 350)$$

$$p\left(\sum \bar{X} > 51000\right) = p\left(z > \frac{51\,000 - 50\,000}{350}\right) = p\left(Z > 2.86\right) =$$

$$= 1 - p\left(Z \leq 2.86\right) = 0.0021$$

TAREA:
1. Las bolsas de azúcar rellenadas por una máquina tienen μ = 500 g y σ = 35 g. Las bolsas se empaquetaron en cajas de 100 unidades. Calcular la probabilidad de que una caja 100 de bolsas pese más de 51 kg.

2) De una variable aleatoria x de distribución desconocida, media 14 y desviación típica 1,7 se extraen muestras de tamaño n.
¿Qué se puede decir de la distribución de las medias muéstrales $\bar{x}$:

a) en el caso de que $n = 40$?
b) en el caso de que $n = 18$?
3) Una variable aleatoria x se distribuye según una normal N (110; 17). ¿Qué se puede afirmar de la distribución de las medias $\bar{x}$ de las muestras de tamaño n:

a) si $n = 39$?
b) si $n = 12$?

4) Una variable aleatoria sigue una distribución normal de parámetros $\mu = 17$ y $\sigma = 3$. Halla las distribuciones de las medias muéstrales de muestras de tamaño 15, 25, 40, 100 y 135. ¿Se podría hacer de no seguir la población una distribución normal?

5) Una población de un tipo de plantas tiene una talla media de 17 centímetros y la desviación típica es de 1,8 centímetros. Se toma al azar una muestra de 30 plantas. ¿Cuál es la probabilidad de que la media de las tallas de la muestra sea superior a 15 centímetros?

6) Un dado cúbico está numerado del 1 al 6.

a) Calcula la media y la desviación típica de la población formada por los seis números.

b) Forma todas las muestras posibles de tamaño 2 que podamos obtener con repetición de esta población.

c) Halla la media y la desviación típica de las medias muéstrales.

7) De una variable aleatoria x de distribución desconocida, media 14 y desviación típica 1,7 se extraen muestras de tamaño n.

¿Qué se puede decir de la distribución de las medias muéstrales $\bar{x}$:

a) en el caso de que $n = 40$?

b) en el caso de que $n = 18$?

8) Una variable aleatoria x se distribuye según una normal N (110; 17). ¿Qué se puede afirmar de la distribución de las medias $\bar{x}$ de las muestras de tamaño n:

a) si $n = 39$?

b) si $n = 12$?

9) Una variable aleatoria sigue una distribución normal de parámetros $\mu=17$ y $\sigma=3$ Halla las distribuciones de las medias muéstrales de muestras de tamaño 15, 25, 40, 100 y 135. ¿Se podría hacer de no seguir la población una distribución normal?

10) Una población de un tipo de plantas tiene una talla media de 17 centímetros y la desviación típica es de 1,8 centímetros. Se toma al azar una muestra de 30 plantas. ¿Cuál es la probabilidad de que la media de las tallas de la muestra sea superior a 15 centímetros?

4.4 Tipos de estimaciones y características

Estimar qué va a ocurrir respecto a algo (o qué está ocurriendo, o qué ocurrió), a pesar de ser un elemento muy claramente estadístico, está muy enraizado en nuestra cotidianidad. Dentro de ello, además hacemos estimaciones dentro de un intervalo de posibilidades. Por ejemplo: “creo que terminaré la tarea en unos 5-6 días”. Lo que hacemos en el terreno del análisis de datos es aplicar matizaciones técnicas a este hábito. Vamos a dedicar este documento al concepto de estimación, comenzando con la estimación puntual. Después nos ocuparemos de desarrollar un modelo de estimación por intervalo donde identificaremos los elementos fundamentales, con su significado y símbolo. Y, por último, habrá que desarrollar cómo se calculan esos elementos.

Existen tres tipos de estimación estadística:

<u>Estimación</u>

El objetivo principal de la estadística inferencial es la estimación, esto es que mediante el estudio de una muestra de una población se quiere generalizar las conclusiones al total de la misma. Como vimos en la sección anterior, los estadísticos varían mucho dentro de sus distribuciones muéstrales, y mientras menor sea el error estándar de un estadístico, más cercanos serán unos de otros sus valores.

Existen dos tipos de estimaciones para parámetros; puntuales y por intervalo. Una estimación puntual es un único valor estadístico y se usa para estimar un parámetro. El estadístico usado se denomina estimador.

Una estimación por intervalo es un rango, generalmente de ancho finito, que se espera que contenga el parámetro.

<u>Estimación Puntual</u>

Una estimación es puntual cuando se usa un solo valor extraído de la muestra para estimar el parámetro desconocido de la población. Al valor usado se le llama estimador

Una muestra aleatoria de 3 baterías para calculadora podría presentar duraciones observadas en horas de x1=5.0, x2=6.4 y x3=5.9. El valor calculado de la duración media muestral es $\bar{x}$ = 5.77, y es razonable considerar 5.77 como el valor más adecuado de μ.

Una estimación puntual de un parámetro θ es un sólo número que se puede considerar como el valor más razonable de θ. La estimación puntual se obtiene al seleccionar una estadística apropiada y calcular su valor a partir de datos de la muestra dada. La estadística seleccionada se llama estimador puntual de θ.

Estimación por intervalo

Una estimación por intervalo de un parámetro θ es algún par de funciones de la muestra que satisfacen L(x) ≤ U(x) para todo x ∈ X . El intervalo aleatorio [L(X), U(X)] es llamado un estimador por intervalo.

Estimación bayesiana

El enfoque bayesiano se basa en la interpretación subjetiva de la probabilidad, el cual considera a ésta como un grado de creencia con respecto a la incertidumbre.

Un parámetro es visto como una variable aleatoria a la que, antes de la evidencia muestral, se le asigna una distribución a priori de probabilidad, con base en un cierto grado de creencia con respecto al comportamiento aleatorio. Cuando se obtiene la evidencia muestral, la distribución a priori es modificada y entonces surge una distribución a posteriori de probabilidad.

Ejemplo:

En el futuro habrá cada vez más interés en desarrollar aleaciones de Mg de bajo costo, para varios procesos de fundición. En consecuencia, es importante contar con métodos prácticos para determinar varias propiedades mecánicas de esas aleaciones. Examine la siguiente muestra de mediciones del módulo de elasticidad obtenidos de un proceso de fundición a presión:

44.2 43.9 44.7 44.2 44.0 43.8 44.6 43.1

Suponga que esas observaciones son el resultado de una muestra aleatoria. Se desea estimar la varianza poblacional σ^2. Un estimador natural es la varianza muestral:

$$\hat{\sigma}^2 = s^2 = \frac{\Sigma(x_i - \bar{x})^2}{n-1} = \frac{(44.2-44.0625)^2 + (43.9-44.0625)^2 + \ldots + (43.1-44.0625)^2}{8-1} = 0.251$$

En el mejor de los casos, se encontrará un estimador $\hat{\theta}$ para el cual $\hat{\theta} = \theta$ siempre. Sin embargo, $\hat{\theta}$ es una función de las Xi muéstrales, por lo que en sí misma una variable aleatoria.

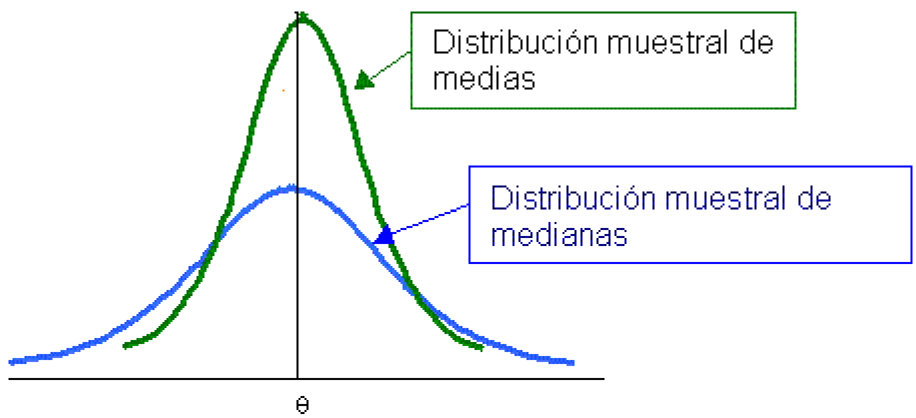
$\hat{\theta} = \theta$ + error de estimación entonces el estimador preciso sería uno que produzca sólo pequeñas diferencias de estimación, de modo que los valores estimados se acerquen al valor verdadero. Propiedades de un Buen Estimador

Insesgado.- Se dice que un estimador puntual $\hat{\theta}$ es un estimador insesgado de θ si $E(\hat{\theta}) = \theta$, para todo valor posible de θ. En otras palabras, un estimador insesgado es aquel para el cual la media de la distribución muestral es el parámetro estimado. Si se usa la media muestral $\bar{x}$ para estimar la media poblacional μ, se sabe que la $\mu_{\bar{x}} = \mu$, por lo tanto la media es un estimador insesgado.

Eficiente o con varianza mínima.- Suponga que $\hat{\theta}_1$ y $\hat{\theta}_2$ son dos estimadores insesgados de θ. Entonces, aun cuando la distribución de cada estimador esté centrada en el valor verdadero de θ, las dispersiones de las distribuciones alrededor del valor verdadero pueden ser diferentes.

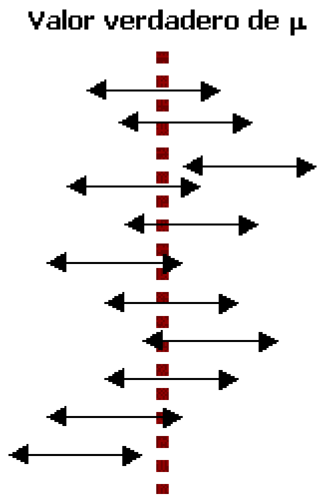
En otras palabras, la eficiencia se refiere al tamaño de error estándar de la estadística. Si comparamos dos estadísticas de una muestra del mismo tamaño y tratamos de decidir cuál de ellas es un estimador más eficiente, escogeríamos la estadística que tuviera el menor error estándar, o la menor desviación estándar de la distribución de muestreo.

Tiene sentido pensar que un estimador con un error estándar menor tendrá una mayor oportunidad de producir una estimación más cercana al parámetro de población que se está considerando.



Como se puede observar las dos distribuciones tienen un mismo valor en el parámetro sólo que la distribución muestral de medias tiene una menor varianza, por lo que la media se convierte en un estimador eficiente e insesgado.

Una interpretación correcta de la "confianza de 95%" radica en la interpretación frecuente de probabilidad a largo plazo: decir que un evento A tiene una probabilidad de 0.95, es decir que si el experimento donde A está definido re realiza una y otra vez, a largo plazo A ocurrirá 95% de las veces. Para este caso el 95% de los intervalos de confianza calculados contendrán a μ.



Esta es una construcción repetida de intervalos de confianza de 95% y se puede observar que de los 11 intervalos calculados sólo el tercero y el último no contienen el valor de μ

Estimación por intervalo

Una estimación por intervalo de un parámetro θ es algún par de funciones de la muestra que satisfacen $L(x) \leq U(x)$ para todo $x \in X$. El intervalo aleatorio $[L(X), U(X)]$ es llamado un estimador por intervalo.

Estimación bayesiana

El enfoque bayesiano se basa en la interpretación subjetiva de la probabilidad, el cual considera a ésta como un grado de creencia con respecto a la incertidumbre.

4.5 Determinación del tamaño de la muestra de una población

El tamaño de la muestra juega un papel importante para determinar la probabilidad de error, así como en la precisión de la estimación. Una vez que se ha seleccionado el nivel de confianza, dos factores importantes influyen en el tamaño muestral:

1. La varianza de la población.
2. El tamaño del error tolerable que el investigador está dispuesto a aceptar.

Mientras que el primer factor está más allá del control del investigador (no hay nada que se pueda hacer sobre la varianza de la población), sí es posible limitar el tamaño del error.

El tamaño del error que un investigador puede tolerar depende de qué tan crítico es el trabajo. Algunas tareas extremadamente delicadas requieren resultados exactos: los procedimientos médicos vitales de los cuales dependen vidas humanas, o la producción de piezas de una máquina que debe cumplir medidas precisas, pueden tolerar sólo un pequeño error. En otros casos, los errores más grandes pueden tener consecuencias menos graves.

Tamaño de la muestra

En Estadística el tamaño de la muestra es el número de sujetos que componen la muestra extraída de una población, necesarios para que los datos obtenidos sean representativos de la población.

Objetivos de la determinación del tamaño adecuado de una muestra

1.Estimar un parámetro determinado con el nivel de confianza deseado.

2.Detectar una determinada diferencia, si realmente existe, entre los grupos de estudio con un mínimo de garantía.

3. Reducir costes o aumentar la rapidez del estudio.

Por ejemplo, en un estudio de investigación epidemiológico la determinación de un tamaño adecuado de la muestra tendría como objetivo su factibilidad. Así:

1. Si el número de sujetos es insuficiente habría que modificar los criterios de selección, solicitar la colaboración de otros centros o ampliar el periodo de reclutamiento. Los estudios con tamaños muéstrales insuficientes, no son capaces de detectar diferencias entre grupos, llegando a la conclusión errónea de que no existe tal diferencia.

2. Si el número de sujetos es excesivo, el estudio se encarece desde el punto de vista económico y humano.

Además, es poco ético al someter a más individuos a una intervención que puede ser menos eficaz o incluso perjudicial.

El tamaño de una muestra es el número de individuos que contiene.

Cálculo del tamaño de la muestra

El tamaño de la muestra se determina para obtener una estimación apropiada de un determinado parámetro poblacional.

Estimación de parámetros
La estimación de parámetros consiste en el cálculo aproximado del valor de un parámetro en la población, utilizando la inferencia estadística, a partir de los valores observados en la muestra estudiada. Para el cálculo del tamaño de la muestra en una estimación de parámetros son necesarios los conceptos de Intervalo de confianza, variabilidad del parámetro, error, nivel de confianza, valor crítico y valor α (véase estimación por intervalos).

Estimación de una proporción

Los datos que tenemos que incluir en la fórmula para calcular el número de sujetos necesarios de la muestra (N) son:

1. Zα/2: valor de Z correspondiente al riesgo α fijad El riesgo α fijado suele ser 0,05 y Zα/2 de 1,96.

2. P: Valor de la proporción que se supone existe en la población.

3. Precisión con que se desea estimar el parámetro (2i es la amplitud del intervalo de confianza).

<u>Estimación de una media</u>
Los datos que tenemos que incluir en la fórmula para calcular el número de sujetos necesarios en la muestra (N) son:

1. Zα/2: valor de Z correspondiente al riesgo α fijado. El riesgo α fijado suele ser 0,05 y Zα/2 de 1,96.

2. s2: Varianza de la distribución de la variable cuantitativa que se supone que existe en la población.

3. i: Precisión con que se desea estimar el parámetro (2i es la amplitud del intervalo de confianza).

Normalmente de forma arbitraria se acepta un riesgo del 5%. Además, hay que establecer el riesgo que se acepta de cometer un error tipo II, que suele ser entre el 5 y el 20%.

• Si la hipótesis es unilateral o bilateral: El planteamiento de una hipótesis bilateral o "de dos colas" requiere mayor tamaño muestral.

• Definir la Magnitud de la diferencia efecto o asociación que se desea detectar: A mayores diferencias preestablecidas en el planteamiento de la hipótesis, menor tamaño muestral, y a menor diferencia, mayor tamaño muestral.

• Conocer la variabilidad del criterio de evaluación en la población. Comparación de dos proporciones

Para calcular el número de sujetos necesarios en cada una de las muestras (n), debemos prefijar:

• 1,96 = Valor Z correspondiente al riesgo deseado.
• 1,96 = Valor Z correspondiente al riesgo deseado, si es de dos colas.
• 0,13 = Valor de la proporción en el grupo de referencia, placebo, control o tratamiento habitual.

• 0,44 = Valor de la proporción en el grupo del nuevo tratamiento, intervención o técnica.
• 0,29 = Media de las dos proporciones p1 y p2.

4.6 INTERVALOS DE CONFIANZA PARA LA MEDIA, CON EL USO DE LA DISTRIBUCIÓN

Dada una variable aleatoria con distribución Normal N(μ, σ), el objetivo es la construcción de un intervalo de confianza para el parámetro μ, basado en una muestra de tamaño *n* de la variable.

Desde el punto de vista didáctico hemos de considerar dos posibilidades sobre la desviación típica de la variable: que sea conocida o que sea desconocida y tengamos que estimarla a partir de la muestra. El caso de σ conocida, ya comentado anteriormente, no pasa de ser un caso académico con poca aplicación en la práctica, sin embargo, es útil desde el punto de vista didáctico.

Caso de varianza conocida

Dada una muestra $X_1, ..., X_n$, el estadístico

$$z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

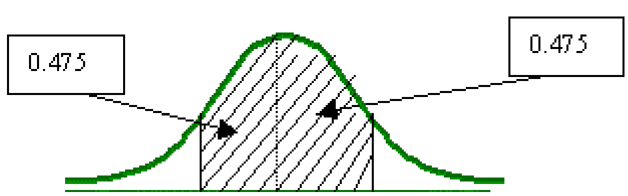
Existen varias tablas en las cuales podemos encontrar el valor de z, según sea el área proporcionada por la misma. En esta sección se realizará un ejemplo para encontrar el valor de z utilizando tres tablas diferentes.

Ejemplos:

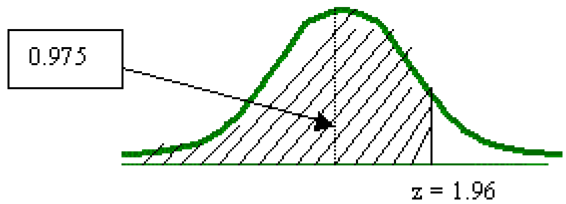
Encuentre el valor de z para un nivel de confianza del 95%.

Solución:

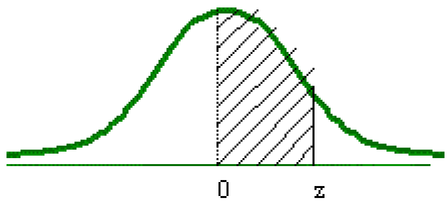
1, Se utilizará la tabla que tiene el área bajo la curva de $-\infty$ hasta z. Si lo vemos gráficamente sería:



El nivel de confianza bilateral está dividido en partes iguales bajo la curva:

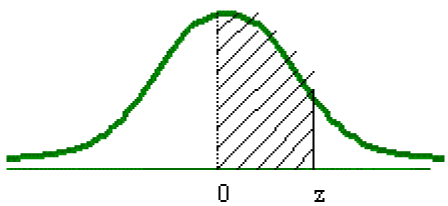


En base a la tabla que se está utilizando, se tendrá que buscar el área de 0.975, ya que cada extremo o cola de la curva tiene un valor de 0.025.



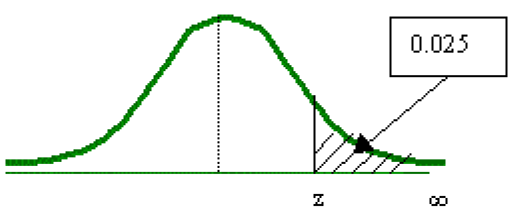
Por lo que el valor de z es de 1.96.

Solución*:*
2. Si se utiliza una tabla en donde el área bajo la curva es de 0 a z:



En este caso sólo se tendrá que buscar adentro de la tabla el área de 0.475 y el resultado del valor de z será el mismo, para este ejemplo 1.96.

Solución*:*
3. Para la tabla en donde el área bajo la curva va desde z hasta ∞:



Se busca el valor de 0.025 para encontrar z de 1.96.

Independientemente del valor del Nivel de Confianza este será el procedimiento a seguir para localizar a z. En el caso de que no se encuentre el valor exacto se tendrá que interpolar.

CONCLUSIONES

La Estadística responde a la actividad planificadora de la sociedad. Con la Revolución Industrial aparecen nuevos problemas, en este caso las emisiones atmosféricas. La Estadística es un instrumento para identificar causas e impactos que esta problemática genera en la sociedad.

La estadística es el conjunto de diversos métodos matemáticos que tienen como o obtener, presentar y analizar datos (ya sean números o cualidades).

La estadística nos permite realizar estudios reales, con poblaciones exactas; lo cual nos ayuda a mejorar nuestros proyectos.

Dentro de una planificación ambiental los datos estadísticos juegan un papel muy importante, pues nos van a determinar en primera medida gastos y nos garantizara la eficiencia.

Este trabajo evidencia todos y cada uno de los temas vistos dentro del plan semestral del programa ingeniería ambiental; lo aquí presentado permitió desarrollar el sentido de localización de cada uno de los estudiantes pues fijo datos reales a temas teóricos.

Llevar un buen registro de datos estadísticos nos permite conocer de mejor manera el problema, cuando nosotros conocemos la realidad de nuestras áreas afectadas; es más fácil dar soluciones.

Los diferentes tipos de distribuciones nos permiten prever eventos que puedan ocurrir, teniendo en cuenta lo que ha sucedido anteriormente (datos históricos).

Una de las técnicas más utilizadas dentro de la estadística es la medición de parámetros de tendencia central, la moda, mediana y media. Lo cual nos permite centrar el problema y plantear puntos de referencia.

Para desarrollar un buen proyecto ambiental siempre es necesario conocer las bases estadísticas del lugar donde vayamos a trabajar.

Conocer la teoría nos ayuda a enfocar soluciones y conocer la realidad nos ayuda a contextualizar y a diferenciar soluciones.

Bibliografía

Domínguez, F y Nieves, A. (2014). *Probabilidad y Estadística para Ingeniería, un enfoque moderno*. McGraw Hill, Novena Edición.

Irwin Miller, John E. Freund, Richard Johnson. (2000). *Probabilidad y estadística para ingenieros.* Prentice Hall inc.

Murray R. Spiegel (2001). *Estadística.* Schaum, Segunda Edición.

Murray, R. y Larry, D. (1992). *Estadística.* McGraw Hill, 3ª Edición.

Walpole, Myers. (1992). *Probabilidad y Estadística.* McGraw Hill, Cuarta Edición.

Walpole, Myers, Myers. (2012). *Probabilidad y Estadística para Ingeniería y Ciencias*. Person, Novena Edición.

William K. Hines., Douglas Montgomery. (2004). *Probabilidad y Estadística para ingeniería.* Grupo Patria Cultural.