



REPORTE DE PERIODO SABÁTICO

AUTORIZACIÓN No. AS-2-290/2023

Periodo sabático

01 de septiembre de 2023 al 31 de agosto de 2024

**Programa No. 5 Elaboración de apuntes, libros, objetivos
educacionales y reactivos de evaluación, bajo la modalidad:**

B.1 Apuntes para Estadística para la Administración I

Clave de la asignatura LAD-1016

CARRERA: LICENCIATURA EN ADMINISTRACION

PRESENTADO POR: M.C. MARIO ALBERTO SANTIESTEBAN POLANCO

Cd. Cuahtémoc Chihuahua

Contribución académica: Análisis de la contribución del tema a los objetivos educativos y perfil de egreso del programa educativo al cual contribuye la asignatura.

Esta asignatura aporta al perfil del Licenciado en Administración la capacidad para tomar decisiones en base a los análisis resultantes de la aplicación de herramientas estadísticas tanto descriptiva como inferencial. La probabilidad, el muestreo, la estimación y el control estadístico de procesos son conocimientos auxiliares para el desarrollo de habilidades en la función administrativa que el administrador necesita aprender para implementar procesos de mejora, proyectos de innovación y solución de problemas en las organizaciones.



Temario

Unidad 1: Principios generales de la estadística en las organizaciones

Subtemas:

- 1.1. La estadística en las actividades empresariales con un enfoque administrativo.
- 1.2. Su importancia y aplicaciones.
- 1.3. Conceptos básicos.
- 1.4. Aplicación del proceso administrativo en los estudios estadísticos.
- 1.5. Aplicación de la estadística descriptiva en las actividades del administrador.

Unidad 2: Estadística descriptiva

Subtemas:

- 2.1. Tablas de distribuciones de frecuencia (una, dos o múltiples entradas).
- 2.2. Gráficas (Histogramas, ojivas, pictogramas, etc.).
- 2.3. Diagramas de caja.

Unidad 3: Medidas de posición y variación para datos agrupados y no agrupados.

Subtemas:

- 3.1. Media aritmética, mediana y moda.
- 3.2. Cuartiles, deciles y percentiles.
- 3.3. Rango, varianza, desviación estándar, coeficiente de variación y de Pearson.

Unidad 4: Probabilidad y teoría de conjuntos

Subtemas:

- 4.1. Aspectos generales de la probabilidad (Conceptos, tipos de probabilidad y enfoques).
- 4.2. Leyes de la probabilidad.
- 4.3. Aplicaciones de la probabilidad en la administración.
- 4.4. Árboles de probabilidad.
- 4.5. Teorema de Bayes.
- 4.6. Teoría de conjuntos aplicados en la administración.

Unidad 5: Distribuciones de probabilidad aplicadas en la administración

Subtemas:

- 5.1. Distribuciones para variables discretas.
- 5.2. Distribuciones para variables continuas.

Unidad 6: Muestreo y estimación aplicado al control estadístico de procesos

Subtemas:

- 6.1. Fundamentos teóricos del muestreo y estimación.
- 6.2. Distribución del muestreo; Características y aplicación en el área administrativa.
- 6.3. Teorema del límite central.
- 6.4. Tipos de estimación y características.
- 6.5. Determinación del tamaño de la muestra.
- 6.6. Intervalos de confianza aplicados al control estadístico de procesos.
- 6.7. Gráficas de control y tipos de variaciones en los procesos.

Introducción

El objetivo en la elaboración de estos apuntes de la asignatura de “Estadística para la Administración I” de la carrera: Licenciatura en Administración, es para facilitar la comprensión y aplicación de los conceptos de probabilidad y estadística básica de un egresado(a) de dicha carrera y acompañarlo(a) a lo largo de su trayectoria académica; Dichos apuntes proporcionaran al estudiante los conocimientos y técnicas básicas para coleccionar, describir, analizar e interpretar datos numéricos para la toma de decisiones acertadas, partiendo de una muestra para hacer inferencias de una población.

Lo anterior partiendo de una forma simple y sencilla e ir incrementando el nivel de complejidad en cada uno de los subtemas que conforman dicho programa.

Dado que la mayoría de los libros existentes que abordan los temas considerados en este programa rebasan los límites de comprensión por parte de los estudiantes, es necesario contar con material de apoyo más simple y fácil de entender, es muy común que los estudiantes cuestionen la utilidad y la aplicación de ésta importante asignatura, por ello el objetivo de realizar apuntes más enfocados al programa académico.

ÍNDICE

Unidad 1: Principios generales de la estadística en las organizaciones **págs. 6 - 19**

Subtemas:

- 1.1. La estadística en las actividades empresariales con un enfoque administrativo.
- 1.2. Su importancia y aplicaciones.
- 1.3. Conceptos básicos.
- 1.4. Aplicación del proceso administrativo en los estudios estadísticos.
- 1.5. Aplicación de la estadística descriptiva en las actividades del administrador.

Unidad 2: Estadística descriptiva **págs. 20 - 35**

Subtemas:

- 2.1. Tablas de distribuciones de frecuencia (una, dos o múltiples entradas).
- 2.2. Gráficas (Histogramas, ojivas, pictogramas, etc.).
- 2.3. Diagramas de caja.

Unidad 3: Medidas de posición y variación para datos agrupados y no agrupados.

Subtemas:

págs. 36 - 47

- 3.1. Media aritmética, mediana y moda.
- 3.2. Cuartiles, deciles y percentiles.
- 3.3. Rango, varianza, desviación estándar, coeficiente de variación y de Pearson.

Unidad 4: Probabilidad y teoría de conjuntos

págs. 48 - 61

Subtemas:

- 4.1. Aspectos generales de la probabilidad (Conceptos, tipos de probabilidad y enfoques).
- 4.2. Leyes de la probabilidad.
- 4.3. Aplicaciones de la probabilidad en la administración.
- 4.4. Árboles de probabilidad.
- 4.5. Teorema de Bayes.
- 4.6. Teoría de conjuntos aplicados en la administración.

Unidad 5: Distribuciones de probabilidad aplicadas en la administración **págs. 62 - 78**

Subtemas:

- 5.1. Distribuciones para variables discretas.
- 5.2. Distribuciones para variables continuas.

Unidad 6: Muestreo y estimación aplicado al control estadístico de procesos

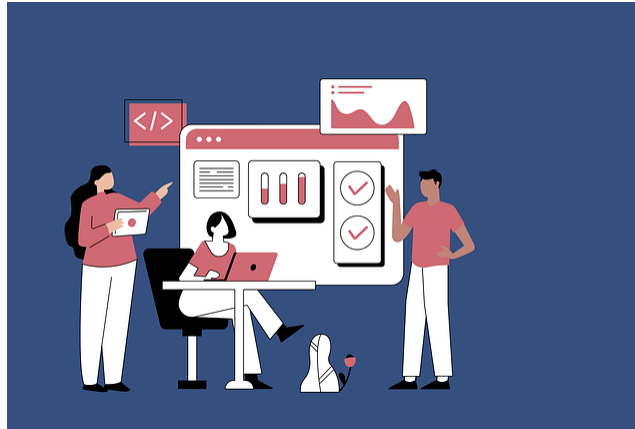
Subtemas:

págs. 79 - 93

- 6.1. Fundamentos teóricos del muestreo y estimación.
- 6.2. Distribución del muestreo; Características y aplicación en el área administrativa.
- 6.3. Teorema del límite central.
- 6.4. Tipos de estimación y características.
- 6.5. Determinación del tamaño de la muestra.
- 6.6. Intervalos de confianza aplicados al control estadístico de procesos.
- 6.7. Gráficas de control y tipos de variaciones en los procesos.

Unidad 1

Principios generales de la estadística en las organizaciones



La relevancia de la Estadística en nuestra sociedad, según los siguientes autores:

“Los datos se presentan en el mundo real sin un orden visible y corresponde al estadístico hallarle sentido a fin de aprender sobre las conclusiones que se obtienen. Veamos algunos ejemplos de situaciones que requieren el uso de La Estadística:

- Alimentación de animales. Hallar dietas balanceadas, de acuerdo con su edad y al propósito de la crianza.
- Control de calidad en la producción industrial: Estimar el número de piezas o artículos defectuosos.
- Aplicación de nuevos métodos de enseñanza.
- Estimación del pago de las pólizas de seguro contra un cierto riesgo.
- Pronosticar la tasa de inflación de un país después de un determinado número de años.
- Determinar la marca de un producto preferida por un sector de una población.
- Pronosticar el comportamiento de los valores de las acciones de un inmueble en la bolsa de valores. MARTÍNEZ, E. (2020)

“La estadística se usa en los negocios para ayudar a tomar mejores decisiones mediante la comprensión de las fuentes *de variación* y el descubrimiento de *patrones y relaciones* en los datos de negocios.” Kazmier, L. (2006)

“La estadística ofrece los procedimientos para recolectar y transformar los datos de manera que sean útiles a quienes toman decisiones en los negocios”. Levine, David M., Krehbiel, Timothy C., Berenson, Mark L. (2006)

Como se puede observar la gran relevancia que tiene la estadística en nuestra vida personal, la sociedad y en el sector productivo; es por ello el entenderla y poder aplicarla de una forma adecuada.

1.1 La estadística en las actividades empresariales con un enfoque administrativo

Hay ocasiones que dentro de algunas organizaciones se llegan a tomar decisiones de forma impulsiva (considerada toma de decisión no estructurada). Por ello, la recomendación de emplear la estadística para decisiones más acertadas.

La estadística desempeña un papel importante en las actividades empresariales en el aspecto administrativo. Se utiliza para recopilar, organizar, analizar e interpretar datos numéricos con el fin de tomar decisiones informadas y respaldar la gestión eficiente de cualquier organización. Habiendo varias formas en las que la estadística se encuentra inmersa en el ámbito empresarial, algunas de ellas son:

- *Toma de decisiones informadas*
- *Análisis de mercado*
- *Gestión de operaciones*
- *Control de calidad*
- *Toma de decisiones financieras*
- *Recopilación y análisis de datos de recursos humanos*
- *Investigación de mercado*
- *Planificación estratégica*
- *Optimización de precios*

1.2 Su importancia y aplicaciones

La estadística, proporciona herramientas y técnicas para recopilar, analizar e interpretar datos, lo que permite a las organizaciones tomar decisiones informadas y mejorar sus operaciones ya sea en el ámbito financiero y/o procesos productivos. Aquí hay algunas de las razones por las cuales la estadística es importante en el ámbito empresarial y algunas de sus aplicaciones más comunes:

- **Toma de decisiones informadas:** La estadística permite a las empresas tomar decisiones basadas en evidencia y datos concretos en lugar de depender únicamente de la intuición. Esto es especialmente relevante en situaciones donde las decisiones pueden tener un impacto significativo en la rentabilidad y el éxito de la organización.
- **Análisis de mercado:** Las empresas utilizan métodos y técnicas estadísticas para analizar el mercado, proveedores y comprender mejor a sus clientes, sus preferencias y sus necesidades, ofreciendo un precio más justo. Esto puede incluir la segmentación del mercado, el análisis de tendencias y la evaluación de la oferta y la demanda.
- **Gestión de operaciones:** La estadística se utiliza para mejorar la eficiencia y la productividad en la cadena de suministro y las operaciones. Por ejemplo, se pueden emplear técnicas de pronóstico para prever la demanda de productos y así planificar la producción y el inventario de manera adecuada, con una adecuada selección de proveedores de los recursos a emplear.
- **Control de calidad:** Las herramientas estadísticas, como el control estadístico de procesos (CEP) permiten a las empresas supervisar y mantener la calidad de sus productos y servicios. Esto ayuda a prevenir defectos y garantiza que los productos cumplan con los estándares establecidos para la satisfacción de los clientes.
- **Toma de decisiones financieras:** La estadística se aplica en el análisis financiero, incluyendo la evaluación de riesgos, el cálculo de razones financieras y la predicción de flujos de efectivo. Estos análisis son fundamentales para la toma de decisiones de inversión y financiamiento, y contar con una organización más estable y sólida.
- **Recopilación y análisis de datos de recursos humanos:** Las empresas pueden utilizar la estadística para evaluar el desempeño de los empleados, analizar la rotación de personal y diseñar estrategias de recursos humanos más efectivas en beneficio de un buen ambiente laboral y mejor desempeño productivo.

- **Planificación estratégica:** Las empresas pueden utilizar análisis estadísticos para evaluar la viabilidad de nuevos proyectos, servicios y/o productos, identificar oportunidades de mejora y diseñar estrategias de expansión.
- **Optimización de precios:** Las técnicas estadísticas permiten a las empresas establecer precios competitivos y justos al analizar los datos de costos, demanda y competencia.

Como podemos observar, la estadística en las actividades empresariales con un enfoque administrativo proporciona las herramientas y estrategias necesarias para tomar decisiones basadas en datos sólidos, concretos, para mejorar la eficiencia operativa y aumentar la competitividad en el mercado.

1.3 Conceptos básicos

Es importante considerar algunos conceptos básicos esenciales en el aspecto de estadística, para la generación de información y conocimiento en general, y así poder tomar una mejor decisión dentro de un proceso productivo y/o administrativo.

- **Datos y Variables:** La estadística se basa en la recopilación y el análisis de datos. Los datos son valores numéricos que pueden ser discretos o continuos o categorías que representan características o medidas de interés. Las variables son las características que se están midiendo, y pueden ser de dos tipos: variables cualitativas (categorías) y variables cuantitativas (números).

Clasificación de datos y variables:

Los datos pueden ser:

Datos Cuantitativos: Estos son datos numéricos que representan cantidades medibles. Pueden ser datos discretos (como el número de productos vendidos, el número de personas, unidades aprobadas de alguna materia, etc.) o continuos (valores en un rango, como la altura de las personas, distancias, peso o masa de algún objeto, etc.).

Datos Cualitativos: También conocidos como datos categóricos, representan categorías o etiquetas. No tienen una magnitud numérica intrínseca. Ejemplos, incluyen el estado civil (soltero, casado, divorciado), el color de ojos, preferencia partidista en la política o de algún deporte, etc.

Las variables pueden ser:

Variables Independientes: Son aquellas que se manipulan o controlan en un experimento o estudio para observar su efecto en otras variables. En estadística inferencial, por lo general se llaman variables predictoras.

Variables Dependientes: Son las variables que se miden o registran en respuesta a las manipulaciones de las variables independientes. También se conocen como variables de respuesta.

Ejemplos de variables independientes y dependientes:

- 1) En el estudio llevado a cabo por economistas sobre la relación que existe entre: *la demanda de un producto representa la variable independiente y el precio de éste es la variable dependiente.*
- 2) La relación que hay entre *las inversiones que se realizan en un rubro es la variable independiente y el posicionamiento que tiene ese rubro en el mercado es la variable dependiente.*
- 3) La influencia que tiene *la cantidad de habitantes vacunados contra cierta enfermedad es la variable independiente, en la cantidad de personas que contraen dicha enfermedad que es la variable dependiente.*
- 4) El estudio sobre la repercusión que el suelo tiene en los cultivos de maíz, *el tipo de suelo representa la variable independiente y el crecimiento del maíz es la variable dependiente, ya que el suelo, al proporcionar los nutrientes, e incide en el desarrollo del maíz.*
- 5) El estudio sobre las propiedades de los metales, *la variable independiente es la temperatura y la variable dependiente es el estado de los metales. Una variación en la temperatura produce un cambio en el estado de los metales.*

Variables Continuas: Son variables cuantitativas que pueden tomar cualquier valor en un intervalo determinado. Ejemplos son la altura, el peso y el tiempo.

Variables Discretas: Son variables cuantitativas que solo pueden tomar valores enteros. Por ejemplo, el número de clientes que ingresan a una tienda en un día, número de empresas, etc.

Variables Nominales: Son variables cualitativas que representan categorías sin ningún orden inherente. Ejemplos son el género, la nacionalidad, el color favorito, etc.

Variables Ordinales: También son variables cualitativas, pero tienen un orden natural entre las categorías. Ejemplos son la clasificación de satisfacción del cliente (baja, media, alta) o el nivel educativo (primaria, secundaria, universitario).

Variables Aleatorias: En probabilidad y estadística, una variable aleatoria es una variable cuyos valores se determinan por el azar. Puede ser discreta o continua, siendo las recomendadas para llevar a cabo cualquier tipo de estudio.

Variables Controladas: En un experimento dichas variables se mantienen constantes, no afectando los resultados obtenidos, y ayudan para aislar el efecto de la variable independiente.

Los datos y/o variables pueden ser recopiladas mediante fuentes primarias o secundarias; siendo fuentes primarias: Entrevistas, observación directa, encuestas y cuestionarios. Y fuentes secundarias: Análisis de documentos, sensores y dispositivos tecnológicos, redes sociales y análisis de texto, etc.

- **Población:** Consiste en la totalidad de elementos de un grupo acerca de los cuales se desea obtener cierta información. La población puede ser finita o infinita. Una población finita puede ser por ejemplo el número de productos elaborados en un turno; Una población infinita es por ejemplo todos los resultados (cara o cruz) que se pueden obtener lanzando una moneda una y otra vez.
- **Muestra:** En ocasiones es difícil recopilar datos de una población completa (todos los elementos de interés). Por lo que se toma una muestra (de forma aleatoria), que es un subconjunto representativo de la población. Existen algunos tipos de muestreo aleatorio.
 - **Muestreo aleatorio:** Cada elemento de la población que se estudia, tiene una posibilidad conocida, y generalmente la misma de ser seleccionada para formar parte de la muestra. Esto garantiza que los elementos de la muestra se elijan sin sesgo.
 - **Muestreo aleatorio simple:** En este tipo de muestreo se escogen al azar elementos individuales de la población de estudio, utilizando una tabla de números aleatorios o cierto programa de computadora e inclusive la mayoría de las calculadoras generan números aleatorios. Ver anexo 1

- **Muestreo aleatorio sistemático:** En este tipo de muestreo los elementos de la población se eligen de una lista a intervalos uniformes (por ejemplo, elegir cada quinto elemento de un proceso productivo). El primero de los cinco elementos se elige de forma aleatoria.
- **Muestreo aleatorio estratificado:** Se deben clasificar primero los elementos de la población en estratos o subgrupos con base en una o más características relevantes. Como ejemplo, sea el caso de la inconformidad entre los docentes de dos academias “A” y “B” con respecto a la asignación de sus horarios, por el tipo de asignatura, así como entre hombres y mujeres. Debemos considerar un plan de muestreo estratificado en el cual se tome una muestra aleatoria simple por separado de cada uno de los cuatro estratos: hombre academia “A”, mujer academia “A”, hombre academia “B” y mujer academia “B”.
- **Muestreo aleatorio por conglomerados:** Aquí los elementos de la población se presentan en forma natural en subgrupos, eligiendo de forma aleatoria los subgrupos.

➤ **Medidas de Tendencia Central:** Ayudan a entender el valor típico o central de un conjunto de cualquier tipo de datos. Son medidas, como:

- La media o promedio aritméticos (suma de los valores en el grupo de datos dividida entre el número total de ellos.

Media poblacional
$$\mu = \frac{\sum_{i=1}^N X_i}{N}$$

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

Donde: x_i , cada valor

N , número total de valores

- La mediana (una vez ordenados los valores, se toma el valor central o el promedio de los valores centrales).

Mediana $= \frac{n+1}{2}$, se determina la posición

Donde: n, número total de valores

- La moda (valor que más se repite), que ayuda a resumir dónde se encuentran concentrados los datos en una distribución.

➤ **Medidas de Dispersión:** Estas medidas, ayudan a comprender la variabilidad de los datos, indican cuánto se extienden los datos alrededor de la medida de tendencia central. Algunas de estas medidas son:

- Desviación estándar, siendo la raíz cuadrada de la sumatoria de la diferencia de cada valor con respecto a su media aritmética, esta diferencia se eleva al cuadrado antes de sumar, entre el total de los valores para el caso de una población y en el caso de una muestra sería entre el total de valores menos uno.

$$\sigma = \sqrt{\frac{\sum_{i=1}^N (X_i - \mu)^2}{N}}$$

Desviación estándar poblacional

$$S = \sqrt{\frac{\sum_{i=1}^N (X_i - \bar{X})^2}{N - 1}}$$

Desviación estándar muestral

Dónde: x_i = cada valor

N= El total de los valores

μ = El promedio total de los valores

- El rango, **R**, es la diferencia entre los valores mayor y menor de un conjunto de valores.
- Rango modificado, es aquel en el cual algunos de los valores extremos, a ambos lados de la distribución, se eliminan y no se toman en consideración.
- Desviación media absoluta o desviación promedio, basada en el valor absoluto de la diferencia entre cada valor del conjunto de datos y la media del grupo

$$DAM \text{ poblacional} = \frac{\sum |X - \mu|}{N}$$

Desviación media absoluta poblacional

$$DAM \text{ muestral} = \frac{\sum |X - \bar{X}|}{n}$$

Desviación media absoluta muestral

Donde: X, cada valor

μ , media poblacional

$\bar{X}$, media muestral

N, tamaño de la población

n, tamaño de la muestra

- Varianza, es la suma de la diferencia de cada valor del conjunto de datos y la media del grupo

varianza poblacional

$$\sigma^2 = \frac{\sum_{i=1}^N (X_i - \mu)^2}{N}$$

varianza muestral

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1}$$

- **Distribuciones de Probabilidad:** En administración, se pueden usar distribuciones de probabilidad para modelar y predecir eventos inciertos, como la demanda de productos, los tiempos de entrega, etc. La distribución normal es especialmente relevante debido a su frecuente aparición en fenómenos naturales y cualquier tipo de procesos.
- **Estadística Inferencial:** Se utiliza para hacer inferencias sobre una población basándose en los datos de muestra. Incluye conceptos como la estimación de parámetros y la prueba de hipótesis. Por ejemplo, una empresa podría usar

estadística inferencial para determinar si una nueva estrategia de marketing es efectiva.

- **Regresión y Correlación:** Estas técnicas permiten analizar la relación entre dos o más variables y predecir valores futuros. Por ejemplo, se puede utilizar la regresión para prever las ventas futuras, lanzamientos de nuevos productos, aspectos relacionados con el clima, etc.
- **Análisis de Series Temporales:** En muchos negocios, es crucial analizar datos en función del tiempo, como las ventas mensuales o los ingresos trimestrales. El análisis de series temporales ayuda a identificar patrones, tendencias y estacionalidades.
- **Estadísticas Descriptivas:** Son herramientas que resumen y presentan datos de manera concisa, como tablas, gráficos y medidas resumen. Ayudan a visualizar y comunicar información clave de manera efectiva e incide en la toma de decisiones.

Estos son solo algunos de los conceptos básicos de estadística que tienen aplicaciones directas en la administración empresarial. Utilizar estas herramientas de manera efectiva puede ayudar a las empresas a tomar decisiones informadas, optimizar operaciones y mejorar el rendimiento en general.

1.4 Aplicación del proceso administrativo en los estudios estadísticos

El proceso administrativo se refiere a las etapas fundamentales que se siguen para planificar, organizar, dirigir y controlar las actividades en una organización con el fin de alcanzar sus objetivos de manera eficiente y efectiva. Estas etapas son interdependientes y se aplican en una variedad de contextos, incluidos los estudios estadísticos. Más a detalle se puede observar en que consiste cada etapa:

Planificación:

Aquí se definen los objetivos y metas del estudio estadístico. Se determina qué datos se recopilarán, cómo se recopilarán, cuál será el tamaño de la muestra, y cuáles serán los métodos de análisis estadístico-utilizados. También se establece un cronograma para cada fase del estudio.

Organización:

Se determinan y asignan responsabilidades y se asignan los recursos necesarios para llevar a cabo el estudio. Se forman equipos para la recopilación de datos, se

designan roles específicos y se establecen procedimientos para garantizar que el proceso se realice de manera ordenada y eficiente.

Dirección:

En esta fase, se supervisa y guía el proceso de recopilación y análisis de datos. Los líderes del estudio estadístico deben asegurarse de que los equipos estén trabajando según lo planeado, resolver cualquier duda o problema que surja y motivar al equipo para alcanzar los objetivos establecidos.

Control:

El control implica monitorear el progreso del estudio estadístico y compararlo con los objetivos y metas establecidos en la etapa de planificación. En caso de algún desvío de lo programado, se toman medidas correctivas para corregir el rumbo. Esto puede incluir ajustes en el cronograma, cambios en los métodos de recopilación de datos o en el enfoque del análisis estadístico utilizado.

1.5 Aplicación de la estadística descriptiva en las actividades del administrador

La estadística descriptiva es una herramienta fundamental en la toma de decisiones y la gestión de datos en actividades administrativas. Los administradores utilizan la estadística descriptiva para analizar, resumir y presentar datos de manera que puedan comprender mejor una situación, identificar tendencias, patrones y así poder tomar decisiones informadas con mayor certidumbre. A continuación, podemos observar algunas aplicaciones comunes de la estadística descriptiva en las actividades del administrador:

Resumen de datos: Los administradores a frecuentemente deben resumir grandes cantidades de datos para comprender rápidamente la situación. La estadística descriptiva les permite calcular medidas de resumen como la media, la mediana, la moda y la desviación estándar para obtener una idea de la tendencia central y la dispersión de los datos.

Análisis de tendencias: Los administradores pueden utilizar gráficos y diagramas, como gráficos de barras, gráficos de líneas, gráficos tipo pastel o diagramas de dispersión, para visualizar tendencias a lo largo del tiempo o entre diferentes grupos. Ya que por lo general un dibujo dice más que mil palabras. Esto les ayuda a identificar patrones y tomar decisiones basadas en datos.

Segmentación de datos: La estadística descriptiva aporta a los administradores a segmentar datos en diferentes categorías o grupos. Esto es útil para identificar segmentos de mercado, grupos de clientes o áreas de negocio que requieren atención inmediata y específica.

Evaluación de desempeño: Los administradores pueden utilizar medidas de tendencia central y de dispersión para evaluar el desempeño de los empleados, equipos o departamentos. Por ejemplo, pueden calcular la media y variación promedio de las ventas mensuales de un equipo de ventas para evaluar su rendimiento promedio y su variabilidad.

Control de calidad: En la gestión de la calidad, la estadística descriptiva se utiliza para analizar datos de calidad, como la variabilidad en la producción o la satisfacción del cliente hacia un producto. Esto ayuda a identificar problemas y tomar medidas correctivas.

Pronóstico: Los administradores pueden utilizar técnicas de pronóstico basadas en datos históricos, como el promedio móvil o el análisis de series temporales, para prever tendencias futuras y así poder planificar.

Toma de decisiones: La estadística descriptiva proporciona una base sólida para la toma de decisiones informadas. Al comprender la distribución de datos y las relaciones entre variables, los administradores pueden tomar decisiones estratégicas que maximicen los resultados y tener clientes satisfechos.

Evaluación de riesgos: Los administradores pueden analizar datos relacionados con riesgos financieros, operativos o de mercado para comprender la probabilidad y el impacto de eventos no favorables.

Ejercicios propuestos:

1.- Relacione correctamente ambas columnas:

1.- Gestión de operaciones	() Conocidas también como variables de respuesta.
2.- Control de calidad	() Métodos y técnicas estadísticas para comprender mejor a sus clientes, preferencias y sus necesidades, ofreciendo un precio justo
3.- Variables independientes	() Representan cantidades medibles pueden ser discretos o continuos.
4.- Estadística	() Ayuda a prevenir defectos y garantiza que los productos cumplan con los estándares establecidos.
5.- Toma de decisiones financieras	() Ayudan a comprender la variabilidad de los datos.
6.- Medidas de dispersión	() Mejora la eficiencia y la productividad en la cadena de suministro y las operaciones.
7.- Variables dependientes	() Nos ayudan a entender el valor típico de un conjunto de datos.
8.- Análisis de mercado	() Se manipulan o controlan en un experimento o estudio para observar su efecto en otras variables.
9.- Medidas de tendencia central	() Procedimientos para recolectar y transformar los datos para la toma de decisiones.
10.- Datos cuantitativos	() Análisis para la toma de decisiones de inversión y financiamiento.

Respuestas:

7
8
10
2
6
1
9
3
4
5

2.- Complete las oraciones con el siguiente listado de palabras: *Muestra, Población, Análisis de series temporales, Control, Pronóstico, Estadística descriptiva, Estadística inferencial, Muestreo aleatorio sistemático, Muestreo aleatorio estratificado, Variables controladas, Variables nominales.*

1.- Primero se clasifican los elementos de una población en subgrupos con base en una o más características relevantes _____
2.- En un experimento dichas variables se mantienen constantes _____
3.- En este tipo de muestreo los elementos de la población se eligen de una lista a intervalos uniformes _____
4.- Ayuda a identificar patrones, tendencias y estacionalidades _____
5.- Variables cualitativas, representan categorías sin ningún orden inherente _____
6.- Implica monitorear el progreso del estudio estadístico y lo compara con los objetivos y metas establecidas en la planificación _____

7.- Parte representativa de una población _____
8.- Técnicas basadas en datos históricos, como el análisis de series temporales, para prever tendencias futuras y poder planificar _____

Respuestas:

- 1.- Muestreo aleatorio estratificado
- 2.- Variables controladas
- 3.- Muestreo aleatorio sistemático
- 4.- Análisis de series temporales
- 5.- Variables nominales
- 6.- Control
- 7.- Muestra
- 8.- Pronóstico

Generar números aleatorios en la calculadora:

<https://www.youtube.com/watch?v=Kf6nTDaIFnU>

Generar números aleatorios en Excel:

<https://www.youtube.com/watch?v=PuvIGYz5Ifk>

Unidad 2

Estadística descriptiva



2.1 Tablas de distribuciones de frecuencia (una, dos o múltiples entradas)

Las tablas de distribución de frecuencia son una forma común de resumir y organizar datos en estadística descriptiva. Estas tablas se utilizan para mostrar la distribución de datos en categorías o intervalos, y se pueden aplicar tanto a una sola variable (distribución univariante) como a múltiples variables (distribución multivariante). A continuación. Ejemplos de tablas de distribución de frecuencia para ambas situaciones:

Distribución de frecuencia de una sola variable (univariante):

Supongamos que se tiene una muestra de las edades de 20 personas en una comunidad y se desea crear una tabla de distribución de frecuencia para estas edades. Aquí está un ejemplo de cómo podría verse:

Edad	Frecuencia
20 - 29	4
30 - 39	6
40 - 49	6
50 - 59	2
60 - 69	2

Distribución de frecuencia de una sola variable.

En esta tabla, las edades se han agrupado en intervalos de 10 años (por ejemplo, 20-29, 30-39, etc.), y la columna "Frecuencia" muestra cuántas personas están en cada grupo de edades.

Otro ejemplo para distribución de frecuencia de una sola variable:

La siguiente tabla resumen está basada en una encuesta en la que se preguntó por qué la gente compra regalos en línea en cierta época del año:

Razón	Porcentaje
Comparación de compra	23
Conveniencia	33
Gastos de envío sin costo	34
Variedad en la selección	6
Rapidez	4
Total	100

Distribución de frecuencia de una sola variable.

Distribución de frecuencia de múltiples variables (multivariante):

Cuando se tienen múltiples variables, se elaboran tablas de distribución de frecuencia cruzada para ver cómo se relacionan dos variables diferentes. Por ejemplo, si se desea analizar la relación entre género y nivel de educación en una comunidad, la tabla de distribución de frecuencia podría ser así:

Género	Nivel de Educación	Frecuencia
Masculino	Secundaria	10
Masculino	Bachillerato	16
Femenino	Secundaria	14
Femenino	Bachillerato	19

Distribución de frecuencia de dos variables.

Esta tabla muestra cómo se distribuyen las personas por género y nivel de educación. Puedes ver cuántas personas en cada categoría de género tienen cierto nivel de educación.

Otro ejemplo de este tipo, donde intervienen dos o más variables está representado por:

Nivel de riesgo

Objetivo	Alto	Promedio	Bajo	Total
Crecimiento	14	23	12	49
Valor	3	23	46	72
Total	17	46	58	121

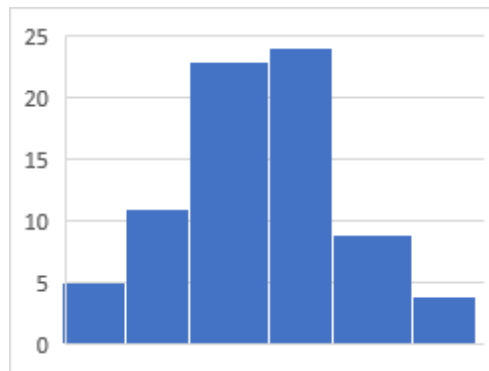
Distribución de frecuencias de dos variables.

Para examinar si hay o no una relación entre el nivel de riesgo y el objetivo del fondo de inversión (crecimiento contra valor), veamos la siguiente tabla para los 121 fondos de inversión

2.2 Gráficas (Histogramas, ojivas, pictogramas, etc.)

En estadística descriptiva, se utilizan varios tipos de gráficas para representar y visualizar datos de manera efectiva, así como la relación entre las variables. Ayudan a resumir, interpretar y comunicar información sobre la distribución y las características de un conjunto de datos. Algunos de los tipos más comunes de gráficas en estadística descriptiva son:

Histograma: Un histograma es una gráfica que muestra la distribución de frecuencias de una variable continua. Los datos se dividen en intervalos y se representa la frecuencia o la densidad de probabilidad en el eje vertical frente a los intervalos en el eje horizontal. Es útil para ver la forma de la distribución, que puede ser simétrica o asimétrica (con algún sesgo, a la derecha o a la izquierda)



Histograma.

Gráfico de barras: Los gráficos de barras se utilizan para representar datos categóricos o discretos. Cada categoría se muestra en el eje horizontal, y la altura de las barras representa la frecuencia o el recuento de esa categoría. Puede ser un gráfico de barras simples o un gráfico de barras apiladas para comparar categorías en diferentes grupos.

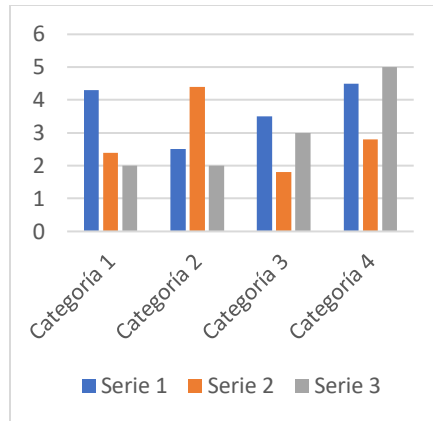


Gráfico de barras.

Diagrama de sectores (o gráfico de pastel): Este tipo de gráfico se utiliza para representar proporciones o porcentajes de un conjunto de datos. Un círculo se divide en sectores, y cada sector representa una categoría o una parte proporcional de un todo.

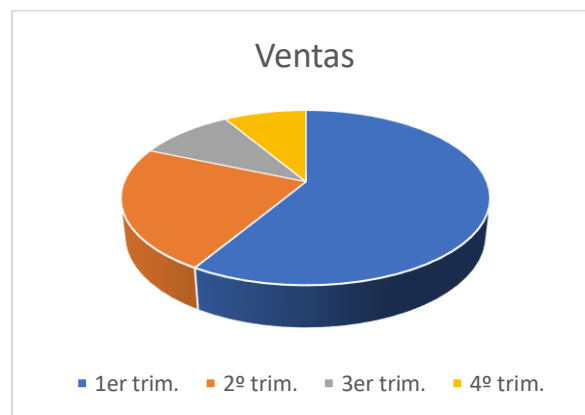


Gráfico de pastel.

Gráfico de líneas: Los gráficos de líneas se utilizan para representar datos continuos a lo largo del tiempo o de una secuencia ordenada. Son especialmente útiles para mostrar tendencias y cambios en los datos a lo largo de un período.

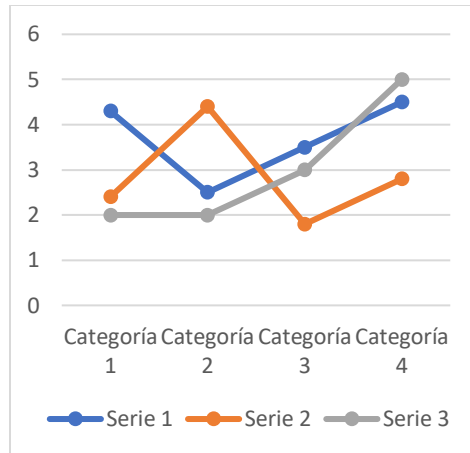


Gráfico de líneas.

Diagrama de dispersión (scatter plot): Un diagrama de dispersión se utiliza para representar la relación entre dos variables numéricas. Cada punto en el gráfico representa un par de valores de las dos variables, lo que permite visualizar la relación entre ellas, como la correlación o la dispersión.

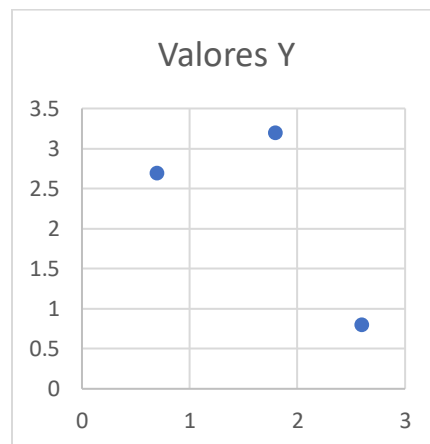


Diagrama de dispersión.

Gráfico de frecuencia acumulada: Este gráfico muestra cómo se acumulan las frecuencias en una distribución. Puede ser un gráfico de frecuencia acumulada ascendente o descendente, lo que permite ver cómo se distribuyen los datos en función de sus valores.

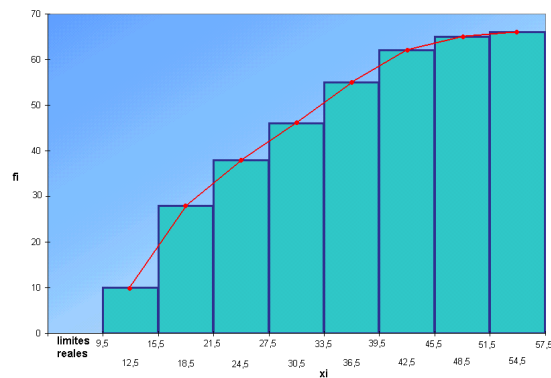


Gráfico de frecuencia acumulada.

Gráfico de Pareto: Este gráfico se utiliza para identificar y priorizar problemas o categorías en función de su importancia relativa; para una variable cualitativa. Se basa en el principio de Pareto, que sugiere que un pequeño número de causas suele ser responsable de la mayoría de los problemas.

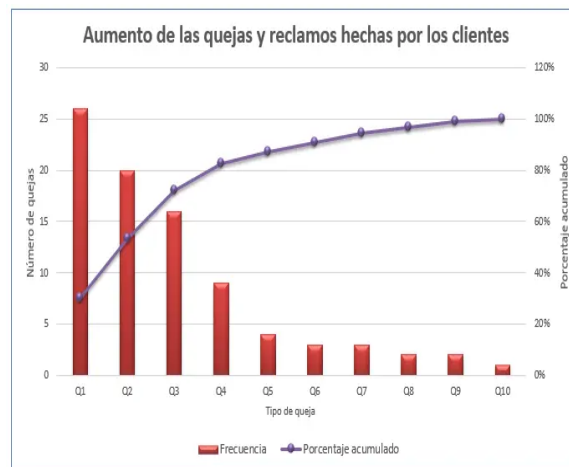


Gráfico de Pareto.

Gráfico de Gantt: Se utiliza para representar la planificación y programación de actividades a lo largo del tiempo. Es comúnmente utilizado en la gestión de proyectos.

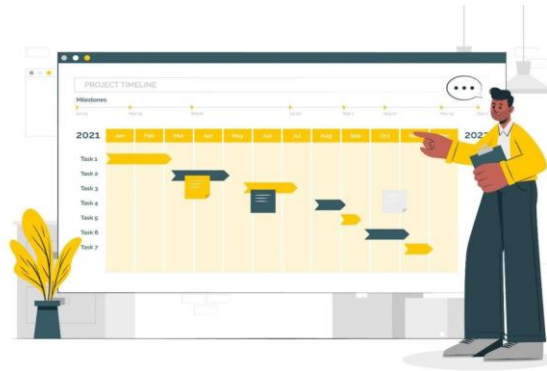


Gráfico de Gantt.

Es importante señalar que la elección de la gráfica adecuada depende del tipo de datos que se esté analizando, así como el objetivo del análisis. Cada tipo de gráfica tiene sus propias ventajas y limitaciones, por lo que es de gran relevancia seleccionar la que mejor se adapte a las necesidades de representación visual de datos.

2.3 Diagramas de caja

Diagrama de caja y bigotes (box plot): Esta gráfica se utiliza para representar la distribución de datos numéricos y mostrar medidas resumen como la mediana, los cuartiles como medidas de posición y el valor del rango intercuartílico como medida de variabilidad. Ayuda a identificar la dispersión y la presencia de valores extremos en los datos.

Lo componen los siguientes elementos:

Caja (box): La caja representa el rango intercuartílico (RIC O IQR), que es la distancia entre el primer cuartil (Q1) y el tercer cuartil (Q3), los cuartiles son medidas de posición. El 50% central de los datos se encuentra dentro de esta caja. La longitud de la caja es proporcional al IQR.

Línea en la caja (línea mediana): En el centro de la caja, hay una línea que representa la mediana (o segundo cuartil, Q2), que es el valor que divide el conjunto de datos en dos mitades iguales. No debemos de olvidar que la mediana es una medida de tendencia central.

Bigotes (whiskers): Los bigotes se extienden desde la caja hacia los valores mínimos y máximos dentro de ciertos límites. Estos límites suelen estar definidos por la "regla

de los 1.5 veces el RIC (IQR)". Los valores fuera de estos límites se consideran valores atípicos y se representan como puntos individuales.

Valores atípicos (outliers): Los valores atípicos son aquellos que están por encima o por debajo de los límites de los bigotes. Se representan como puntos individuales en el gráfico. Los valores atípicos son importantes para identificar datos inusuales o extremos.

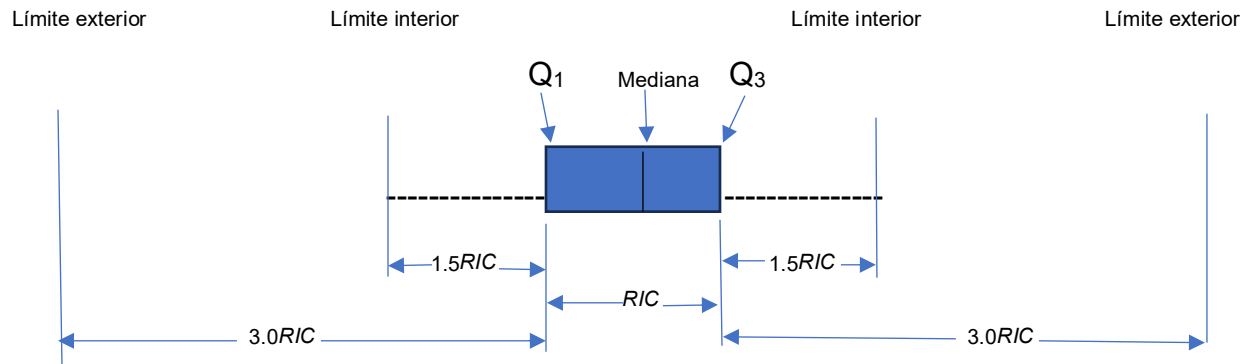
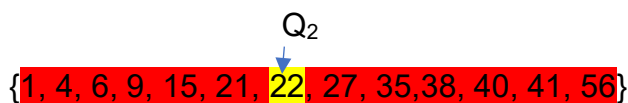


Diagrama de caja y bigotes.

El diagrama de caja es especialmente útil para comparar la distribución de datos entre diferentes grupos o categorías y para identificar valores atípicos en un conjunto de datos. También proporciona una representación visual de la simetría o asimetría de la distribución y muestra la concentración de datos alrededor de la mediana.

Ejemplo:

Encuentre el RIC (IQR) para el conjunto:



Las siguientes fórmulas se pueden utilizar para determinar la posición de los cuartiles en el conjunto

$$Q1=0.25 (n + 1)$$

$$Q3=0.75 (n + 1)$$

donde n es el número de términos del conjunto. Por lo tanto:

$$Q1=0.25 (13 + 1) = 3.5, \text{ posición}$$

$$Q3=0.75 (13 + 1) = 10.5, \text{ posición}$$

Los valores decimales indican que el cuartil se encuentra entre los elementos más cercanos al valor. Por lo tanto, Q1 se encuentra entre el tercer y cuarto elemento del conjunto, y Q3 se encuentra entre el décimo y onceavo elemento. Al promediar los términos en esas posiciones, se obtienen los resultados del primer y tercer cuartil:

$$Q_1 = \frac{6 + 9}{2} = 7.5$$

$$Q_3 = \frac{38 + 40}{2} = 39$$

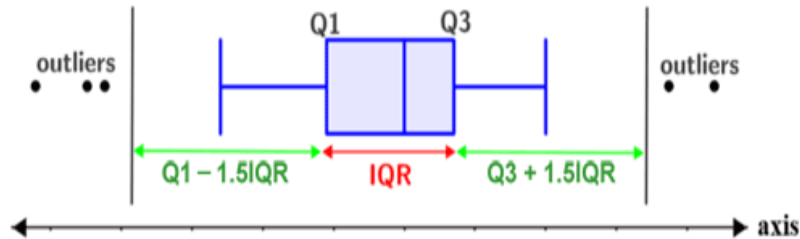
Por lo tanto, el rango intercuartílico se puede calcular como:

$$\text{RIC (IQR)} = Q_3 - Q_1$$

$$\text{RIC (IQR)} = 39 - 7.5 = 31.5$$

Uso del RIC (IQR) para encontrar valores atípicos

El RIC(IQR) se puede utilizar para encontrar valores atípicos (valores en el conjunto que se encuentran significativamente fuera del valor esperado). Los valores que se encuentran a más de 1,5 veces el RIC(IQR) de cualquiera de los extremos del RIC(IQR) (Q1 o Q3) se consideran valores atípicos, como se muestra en la siguiente figura:



Por lo tanto, el rango de valores esperado es:

$$[Q1 - 1.5 (IQR), Q3 + 1.5 (IQR)] = [7.5 - 1.5 (31.5), 39 + 1.5 (31.5)] = [-39.75, 86.25]$$

Todo lo que esté fuera del rango de valores anterior es un valor atípico. En este ejemplo no se cuenta con algún valor atípico.

Ejemplo

Busque valores atípicos para el siguiente conjunto de datos:

Q₂
↓
{1, 3, 4, 6, 13, 20, 25, 26, 28, 62, 95}

Q₂ es la mediana y, dado que hay 11 elementos en el conjunto, Q₂ es el valor medio. Q₁ es la mediana del primer 25% de los valores (primer elemento hasta Q₁) y Q₃ es la mediana del último 25% de los valores (Q₃ hasta el elemento final).

$$Q_1 = 0.25 (11 + 1) = 3, \text{ posición: } Q_1 = 4$$

$$Q_3 = 0.75 (11 + 1) = 9, \text{ posición: } Q_3 = 28$$

{1, 3, 4, 6, 13, 20, 25, 26, 28, 62, 95}

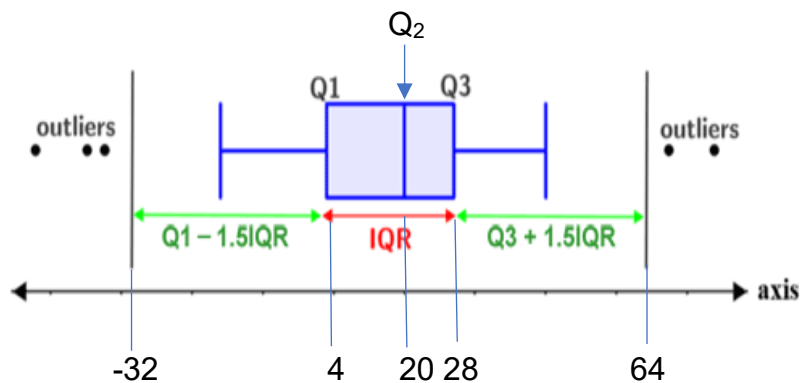
RIC(IQR) es la diferencia entre Q3 y Q1:

$$\text{RIC(IQR)} = 28 - 4 = 24$$

Por lo tanto, cualquier valor fuera del rango

$$[4 - 1,5 (24), 28 + 1,5 (24)] = [-32, 64]$$

es un valor atípico. Todos los datos, excepto el 95, se encuentran dentro del rango anterior. Por lo tanto, 95 es el único valor atípico del conjunto.

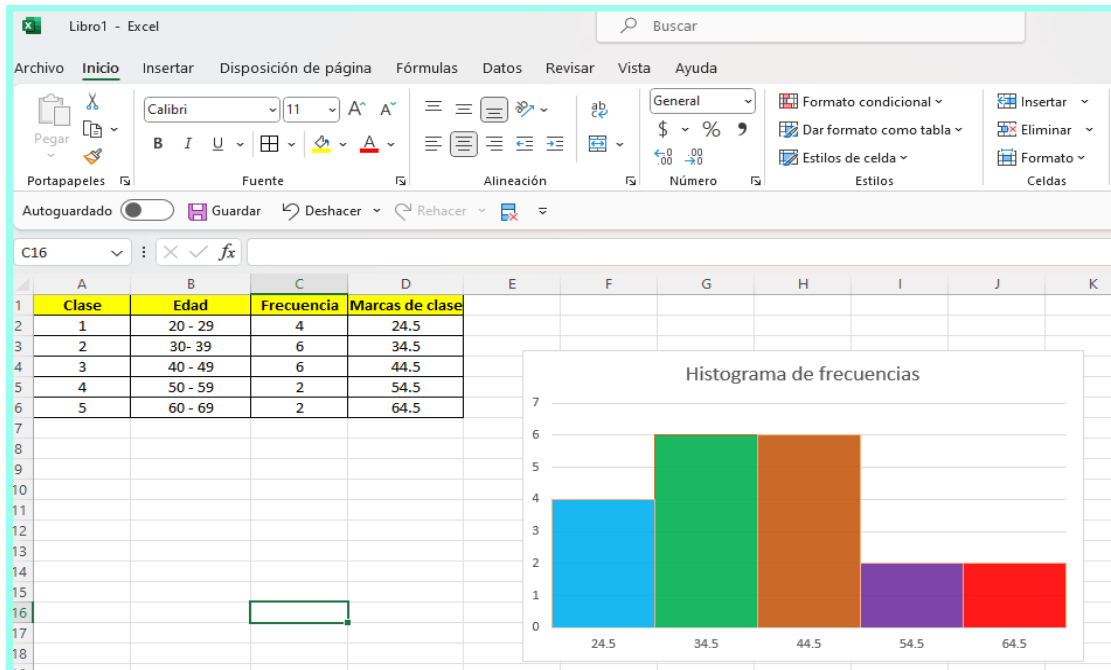


Ejercicios:

1.- Muestra de las edades de 20 personas en una comunidad y se desea obtener las siguientes gráficas:

Edad	Frecuencia
20 - 29	4
30 - 39	6
40 - 49	6
50 - 59	2
60 - 69	2

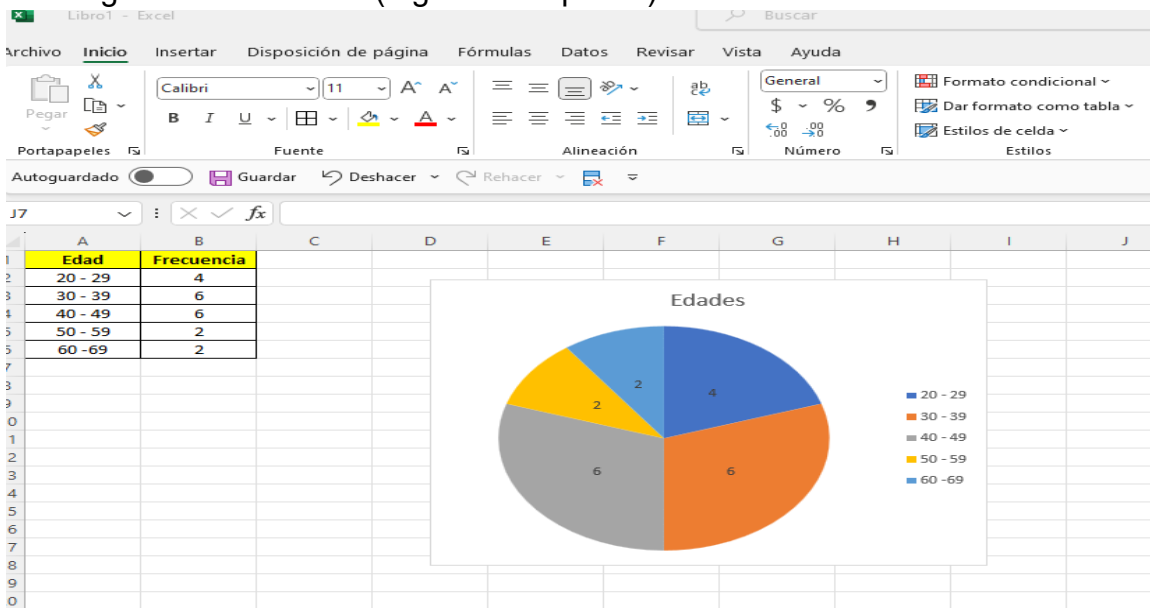
1.- Histograma



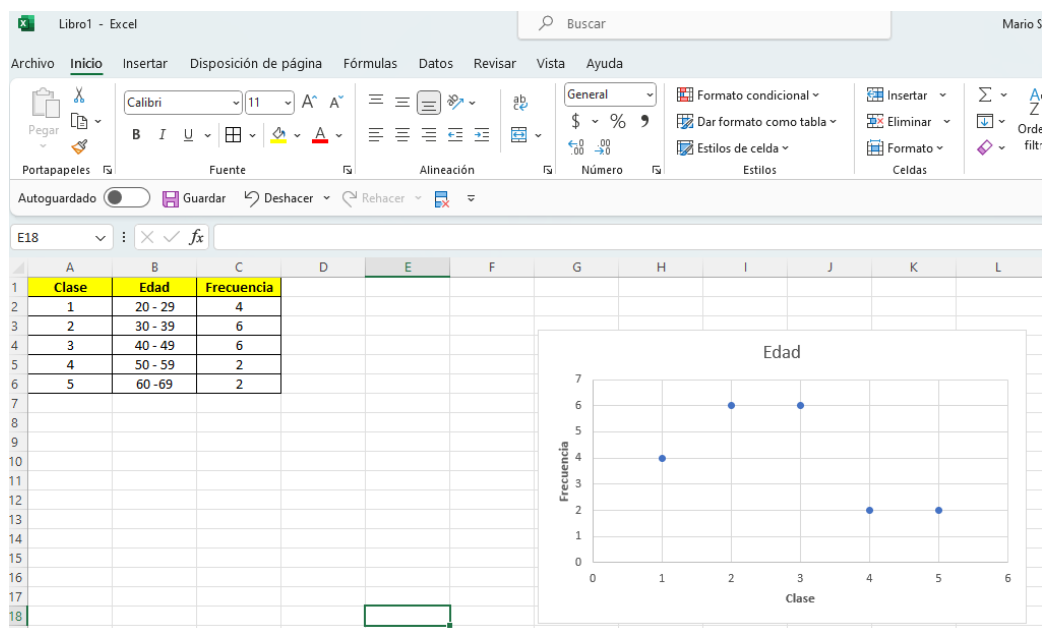
<https://www.youtube.com/watch?v=uZ3Q6Nth7-E>

Tips para ayuda en youtube, de elaborar un histograma.

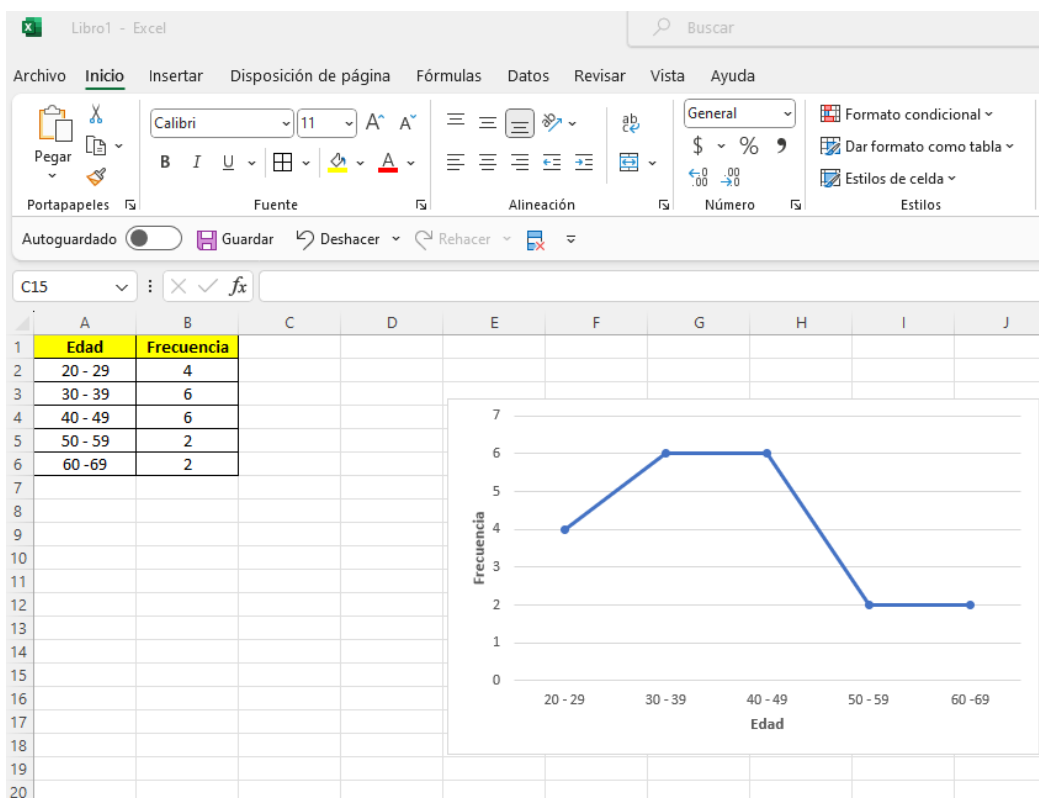
2.- Diagrama de sectores (o gráfico de pastel)



3.- Gráfica de dispersión



4.- Gráfico de líneas



Estos son algunos ejemplos de gráficos que se pueden obtener directamente de Excel de una manera sencilla, una vez que capturas los datos en una tabla de Excel,

en la opción insertar, seleccionas ambas columnas de la información y en gráficos eliges el tipo de gráfico que desees.

Ejercicios propuestos:

I.- 30 administradores realizaron los siguientes tiempos en minutos en el inventario de un producto en una fábrica de celulares:

10	9
14	9
15	10
13	10
17	11
16	11
12	11
14	11
11	12
13	12
15	12
18	12
9	13
14	13
14	13
9	14
15	14
11	14
13	14
11	14
12	15
10	15
17	15
16	15
12	16
11	16
16	16
12	17
14	17
15	18

La primera columna son datos sin ordenar, la segunda columna ya ordenados.

- 1.- Realice una distribución de frecuencias de 5 clases que contenga las siguientes columnas: Tiempo de inventario, frecuencia, marca de clase, límites exactos,
- 2.- ¿Cuáles son los límites establecidos inferior y superior de la segunda clase?
- 3.- ¿Cuáles son los límites exactos inferior y superior de la segunda clase?
- 4.- ¿Cuál es el tamaño o anchura del intervalo?
- 5.- ¿Cuál es el punto medio de la tercera clase?
- 6.- ¿Cuáles son los límites exactos inferior y superior de la clase en la que se tabuló la menor cantidad de tiempo de inventario?
- 7.- Elabore para estos datos: a) Histograma y polígono de frecuencia, b) Gráfico tipo pastel.
- 8.- Elabore el diagrama de caja

Respuestas:

1.-

Clase	Tiempo de inventario	Frecuencia	Marca de clase	Límites exactos
1	9 - 10	4	9.5	8.5 – 10.5
2	11 - 12	8	11.5	10.5 – 12.5
3	13 - 14	8	13.5	12.5 – 14.5
4	15 - 16	7	15.5	14.5 – 16.5
5	17 - 18	3	17.5	16.5 – 18.5
	Total	30		

Para la elaboración de la distribución de frecuencia, se consideró:

Rango = Dato mayor – dato menor = $18 - 9 = 9$

Anchura de la clase = Rango / Número de clases = $9 / 5 = 1.8 \sim 2$

Marca de clase = Promedio de los límites de cada clase, ej: $(9 + 10) / 2 = 9.5$

Límites exactos = Promedio del límite superior de una clase con el límite inferior de la siguiente clase, ej: $(10 + 11) / 2 = 10.5$, representa el límite exacto superior de la primera clase, así como el límite exacto inferior de la segunda clase. Para los límites exactos inferior de la primera clase y el límite exacto superior de la última clase, seguimos la tendencia que siguen ambas columnas.

2.- 11 y 12

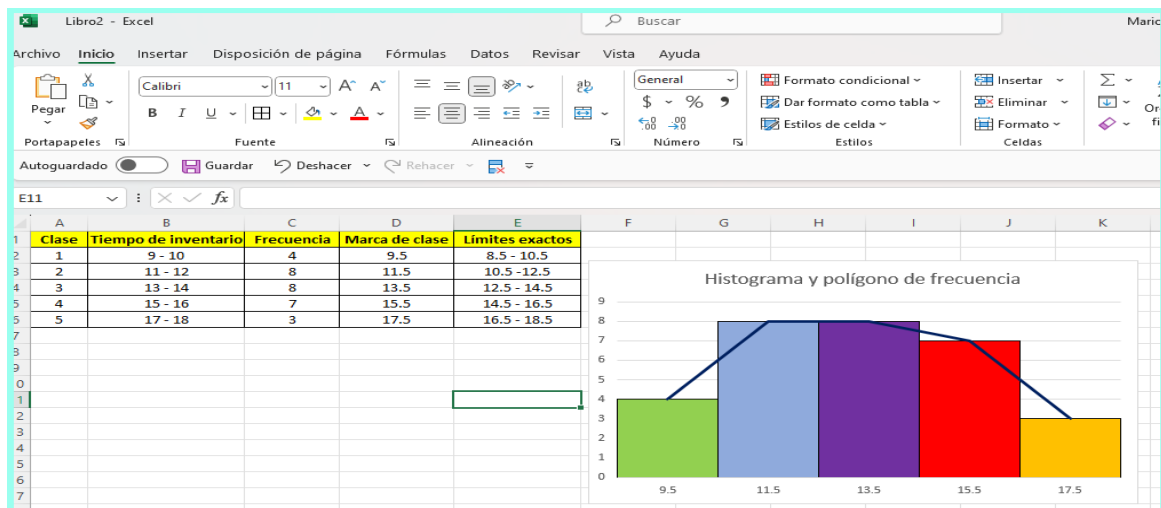
3.- 10.5 y 12.5

4.- 2

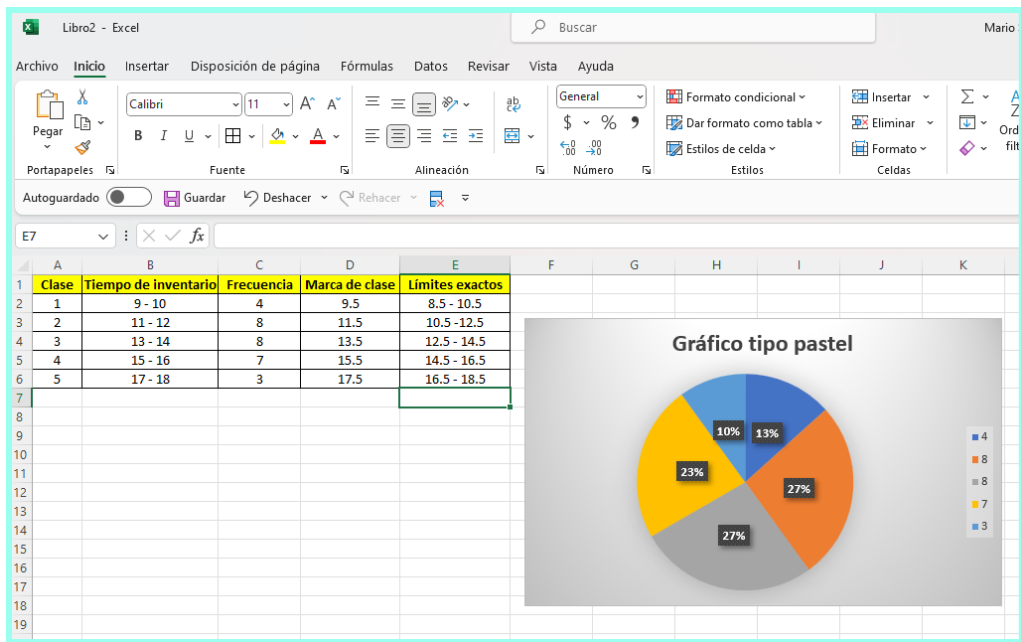
5.- 13.5

6.- 16.5 y 18.5

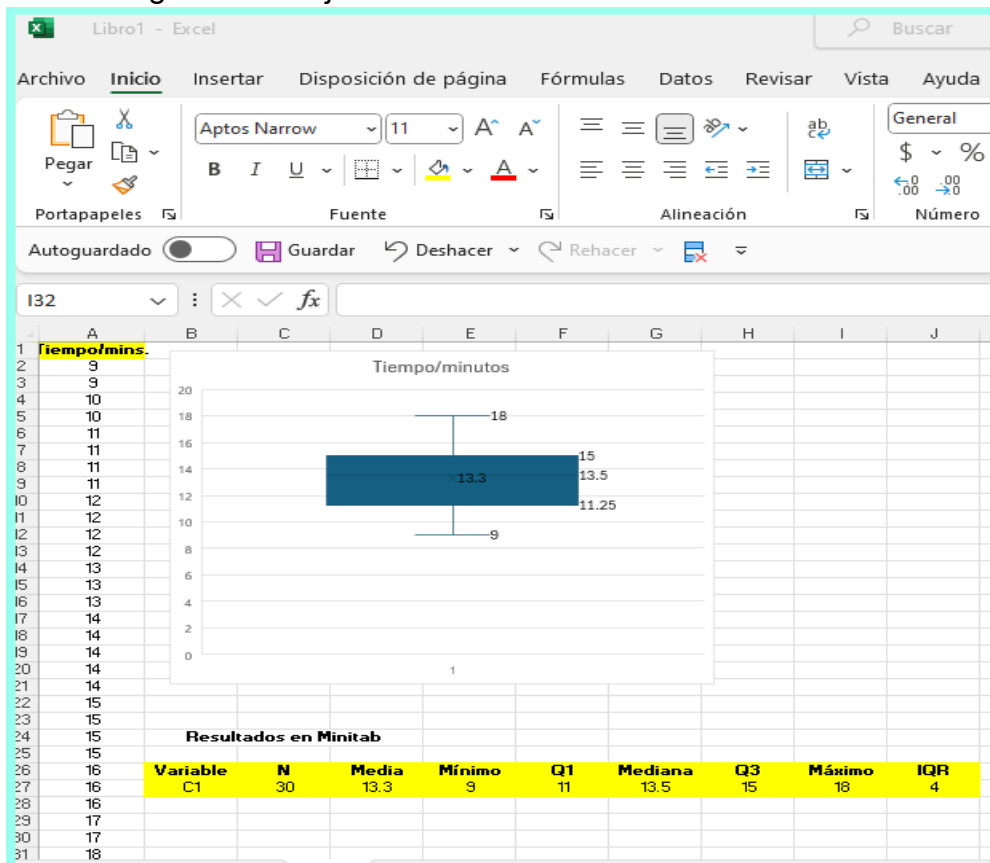
7.- a) Histograma y polígono de frecuencia



b) Gráfico tipo pastel



8.- Diagrama de caja



Unidad 3

Medidas de posición y variación para datos agrupados y no agrupados



3.1 Media aritmética, mediana y moda

Estas medidas denominadas de posición son valores que se calculan a partir de un grupo de datos y nos ayudan a describirlos. También conocidas como medidas de tendencia central.

I.- Media aritmética (o promedio aritmético): Se define como la suma de los valores del grupo de datos dividida entre el número total de valores. En el caso de la media aritmética poblacional (*parámetro poblacional, fórmula 3.1*), se representa con la letra griega μ (léase “mu”), para el caso de la media aritmética de una muestra (*estadístico muestral, fórmula 3.2*), se representa con una letra latina $\bar{X}$ (léase “X” barra).

$$\text{Fórmula 3.1 } \mu = \frac{\sum_{i=1}^N fX_i}{N}$$

$$\text{Fórmula 3.2 } \bar{X} = \frac{\sum_{i=1}^n fX_i}{n}$$

Ejemplo 1: Durante el año 2022 en el mes de junio, en la localidad de Cd. Cuauhtémoc se registraron las siguientes temperaturas en grados $^{\circ}\text{C}$: 18,17, 20,16,15,19,15,18, 21, 20, 22, 23,19,18,14,15,16,18, 22, 21, 20,19, 20, 20,16,17, 18, 16, 18, 24. Considerando este mes como la población estadística de interés, la media aritmética resulta ser:

$$\mu = \frac{\sum_{i=1}^N X_i}{N} = \frac{555}{30} = 18.5^{\circ}\text{C}$$

II.- La mediana: Para un grupo de valores se determina de la siguiente manera, deberán ordenarse los valores ya sea de forma ascendente o descendente. Si el número de elementos es par, la mediana se obtiene promediando los dos valores de en medio. Cuando el número de elementos es impar, el resultado de la mediana es el valor que ocupa la posición central.

La siguiente fórmula nos es útil para determinar la posición de la mediana, una vez que los valores han sido ordenados:

$$\text{Med} = X_{[(n/2) + (1/2)]}$$

Ejemplo 2: De la siguiente serie de calificaciones de un grupo de estudiantes de segundo semestre de la carrera de Licenciatura en Administración, obtener la mediana: 62, 73, 85, 85, 86, 88, 90, 91, 92, 95. Obsérvese que dicha serie se encuentra ordenada.

$\text{Med} = X_{[(10/2) + (1/2)]} = X_{[5.5]}$, la mediana se localiza promediando el valor de la posición 5 con el valor que se encuentra en la posición 6, por lo tanto:

$$\text{Med} = (86 + 88) / 2 = 87$$

III.- La moda: Para un número de elementos, la moda es el valor que se presenta con mayor frecuencia y se denomina *unimodal*, cuando en una serie de elementos existen dos valores que se repiten el mismo número de veces se dice que es *bimodal*; y cuando se presentan tres o más valores que se repiten el mismo número de veces se denomina *multimodal* (varias modas).

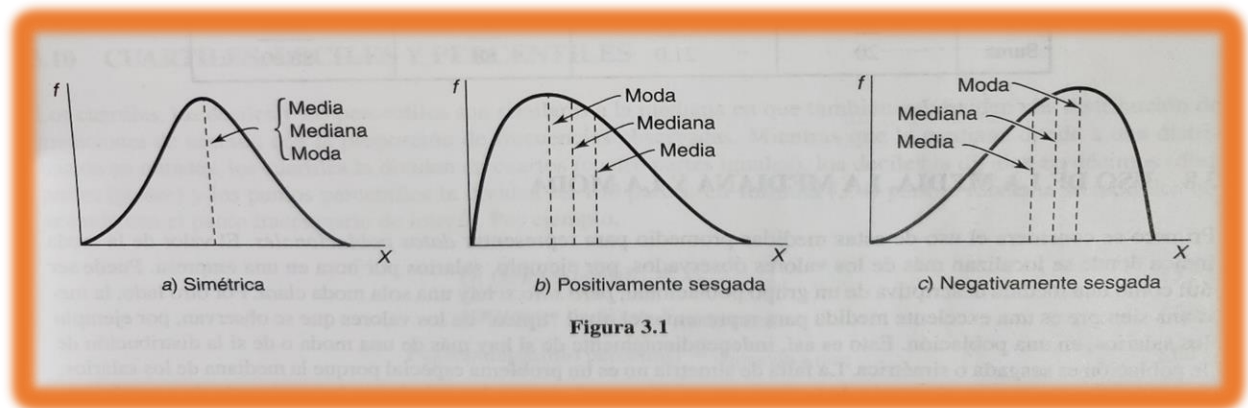
Ejemplo 3: Tomando como referencia el ejemplo 2, el de las calificaciones 62, 73, 85, 85, 86, 88, 90, 91, 92, 95. Su moda es el valor **85**.

Aunque no es necesario ordenar los valores se recomienda ordenarlos, sobre todo cuando se tiene un gran número de datos para evitar confusiones. Como el caso del ejemplo 1, el de las temperaturas registradas durante el mes de junio del año 2022 en Cd. Cuauhtémoc, que ya ordenadas quedarían de la siguiente manera:

14, 15, 15, 15, 16, 16, 16, 16, 17, 17, 18, 18, 18, 18, 18, 18, 19, 19, 19, 20, 20, 20, 20, 20, 21, 21, 22, 22, 23, 24.

Las medidas de tendencia central que se determinaron fue para el caso de datos sin agrupar, para la situación de datos agrupados cambia la manera de calcular dichos valores, ya que se debe contar con una tabla de distribución de frecuencias.

Es importante señalar la relación que existe entre las medidas de tendencia central: Media, Mediana y Moda; En cualquier distribución simétrica, éstas tres medidas coinciden en valor (véase figura 3.1 a), $\bar{X} = m_e = m_o$. Cuando se nos presenta una distribución positivamente sesgada, es decir, con sesgo a la derecha (véase figura 3.1 b), los valores de dicha medida tendrán el siguiente comportamiento; $m_o < m_e < \bar{X}$. En el caso de una distribución negativamente sesgada, es decir, con sesgo a la izquierda (véase figura 3.1 c), tendremos en valores, $\bar{X} < m_e < m_o$.



Para una mejor comprensión de esta relación se recomienda ver el siguiente video disponible en youtube.

<https://www.youtube.com/watch?v=her7uXE91CQ>

Se debe destacar el señalamiento que hace Kazmier, Leonard (2006), en relación con los criterios matemáticos que satisfacen la mediana y la media “la mediana minimiza la suma de las *desviaciones absolutas* respecto de los valores individuales representados. En cambio, la media aritmética minimiza la suma de las *desviaciones al cuadrado* respecto de los valores individuales del grupo. Al criterio mediante el cual el objetivo es la minimización de la suma de las desviaciones al cuadrado correspondientes a un valor representativo se le denomina *criterio de los mínimos cuadrados*”. Considerado el criterio más importante en la inferencia estadística fundamentada en datos muestrales.

3.2 Cuartiles, deciles y percentiles

Este tipo de medidas resultan ser muy similares a la mediana en que también subdividen a un conjunto de valores de acuerdo con la proporción de frecuencias observadas, recordemos que la mediana divide a una distribución de mediciones en mitades, ahora

los *cuartiles* la dividen en cuartos (cuatro partes iguales), los *deciles* la dividen en décimos (diez partes iguales) y los *percentiles* la dividen en cien partes.

En el caso de los cuartiles: el *primer cuartil* Q_1 separa al 25% que incluye a los valores más pequeños, del 75% restante, integrados por los que son mayores. El *segundo cuartil* Q_2 es la mediana: 50% de sus valores son menores que la mediana y 50% son mayores. El *tercer cuartil* Q_3 separa al 25% que abarca a los valores más grandes, del 75% restante constituido por los que son menores.

Tips a considerar para determinar el Q_1 y Q_3 :

$Q_1 = (n + 1) / 4$, encontramos la posición de dicho valor,

$Q_3 = 3(n + 1) / 4$, localizamos la posición de dicho valor.

Criterios que considerar para calcular los cuartiles:

1. Si el resultado es un número entero, entonces el cuartil es igual al valor clasificado.
2. Si el resultado es una fracción de mitad (2.5, 3.5, etcétera), entonces el cuartil es igual al promedio de los valores clasificados correspondientes.
3. Si el resultado no es un número entero ni una fracción de mitad, se redondea al entero más cercano y se selecciona ese valor clasificado.

Ejemplo 4, una vez ordenados en forma ascendente una serie de valores:

Valores ordenados	28	35	36	36	40	43	50	51
Clasificación	1	2	3	4	5	6	7	8

$Q_1 = (n + 1) / 4 = (8 + 1) / 4 = 2.25$, al aplicar el tercer criterio de los cuartiles, redondeamos al segundo valor clasificado, resultando el número 35, que representa hasta ese valor el 25% de los valores.

$Q_2 = 2(n + 1) / 4 = 2(8 + 1) / 4 = 4.5$, al aplicar el segundo criterio de los cuartiles, promediamos los valores que están en las clasificaciones 4 y 5, $(36 + 40) / 2 = 38$, que representa el 50% del total de los valores.

$Q_3 = 3(n + 1) / 4 = 3(8 + 1) / 4 = 6.75$, de igual manera aplicamos el tercer criterio de los cuartiles, redondeamos al séptimo valor clasificado, resultando el número 50, que representa el 75% de los valores.

En el caso de los deciles, los cuales divide la cantidad de los valores en diez grupos, empleamos las siguientes relaciones: $D_1 = (n + 1) / 10$, $D_2 = 2(n + 1) / 10$, $D_3 = 3(n + 1) / 10$, $D_4 = 4(n + 1) / 10$, $D_{10} = 10(n + 1) / 10$.

Para el ejemplo 4, calcular los siguientes deciles: D_4 , D_6 y D_8 . *Nota: considerar los mismos criterios que en el caso de los cuartiles para determinar los resultados.

$D_4 = 4(8 + 1) / 10 = 3.6$; Por lo tanto, redondeamos a la clasificación, 4; $D_4 = 36$, representa el 40% de los valores.

$D_6 = 6(8 + 1) / 10 = 5.4$; al redondear quedaría en la clasificación, 5; $D_6 = 43$, representa el 60% de los valores.

$D_8 = 8(8 + 1) / 10 = 7.2$; al redondear quedaría en la clasificación, 7; $D_8 = 50$, representa el 80% de los valores.

Sea el caso de los percentiles, la cantidad de valores se divide en cien partes, y se representa de la siguiente manera: $P_1 = (n + 1) / 100$, $P_2 = 2(n + 1) / 100$, $P_3 = 3(n + 1) / 100$, $P_{100} = 100(n + 1) / 100$

Del ejemplo 4, determinar, P_{30} , P_{52} , P_{84} . *Nota: considerar los mismos criterios que en el caso de los cuartiles para determinar los resultados.

$P_{30} = 30(8 + 1) / 100 = 2.7$; al redondear quedaría en la clasificación, 3; $P_{30} = 36$, representa el 30% del total de los valores.

$P_{52} = 52(8 + 1) / 100 = 4.68$; al redondear quedaría en la clasificación, 5; $P_{52} = 40$, representa el 52% del total de los valores.

$P_{84} = 84(8 + 1) / 100 = 7.56$; al redondear quedaría en la clasificación, 8; $P_{84} = 51$, representa el 84% del total de los valores.

3.3 Rango, varianza, desviación estándar, coeficiente de variación y de Pearson

El *rango* es la medida numérica más sencilla de la variación en un conjunto de datos, y se obtiene restando al valor mayor el valor más pequeño; considerando el ejemplo 4, para obtener su *rango*; quedaría:

$$\text{Rango} = X_{\text{mayor}} - X_{\text{menor}} = 51 - 28 = 23$$

Es importante señalar que el *rango* no indica si los valores están distribuidos de manera uniforme en todo el conjunto de datos, agrupados cerca de la parte media, o agrupados cerca de uno o ambos extremos. De tal manera que no es confiable utilizar el *rango* como medida de la variación cuando al menos uno de los valores es extremo. Por lo tanto, es recomendable utilizar, **el rango intercuartil** (llamado también *dispersión media*) el cual se obtiene con la diferencia entre el *tercer* y *primer cuartil* de un conjunto de datos. Veamos lo anterior con el siguiente ejemplo.

Ejemplo 5, registro de 10 tiempos en minutos, necesarios para revisar diez estados de cuentas: 29, 31, 35, 39, 39, 40, 43, 44, 44, 52. Obtener el *rango* y el *rango intercuartil*.

$$\text{Rango} = 52 - 31 = 21 \text{ min.}$$

Rango intercuartil = $Q_3 - Q_1 = 44 - 35 = 9$ min. (a este intervalo se le denomina *mitad media*)

$$Q_3 = 3(n + 1) / 4 = 3(10 + 1) / 4 = 8.25; Q_3 = 44$$

$$Q_1 = (n + 1) / 4 = (10 + 1) / 4 = 2.75; Q_1 = 35$$

El rango intercuartil del tiempo necesario para revisar los estados de cuentas es de 9 min.

Se denominan *medidas resistentes*, a las medidas de resumen como la mediana, Q_1 , Q_3 , y el rango intercuartil, ya que no se ven influenciadas por valores extremos.

La varianza y la desviación estándar son dos medidas de la variación muy utilizadas para tomar en cuenta cómo se distribuyen los datos. Miden la dispersión “promedio” alrededor de la media, es decir, qué tanto varían los valores más grandes que están por encima de ella y cómo se distribuyen los valores menores que están por debajo de ella. Importante, *ni la varianza ni la desviación estándar podrán ser negativas*.

Varianza poblacional

$$\sigma^2 = \frac{\sum_{i=1}^N (X_i - \mu)^2}{N}$$

Varianza muestral

$$s^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1}$$

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2}$$

Desviación estándar poblacional

$$s = \sqrt{\frac{1}{N - 1} \sum_{i=1}^N (x_i - \mu)^2}$$

Desviación estándar muestral

Dónde: x_i = cada valor

N = El total de los valores

μ = El promedio total de los valores

Ejemplo 6: Considerando el ejemplo 2, de las calificaciones de un grupo de estudiantes de segundo semestre de la carrera de Licenciatura en Administración: 62, 73, 85, 85, 86, 88, 90, 91, 92, 95; obtener la varianza y la desviación estándar.

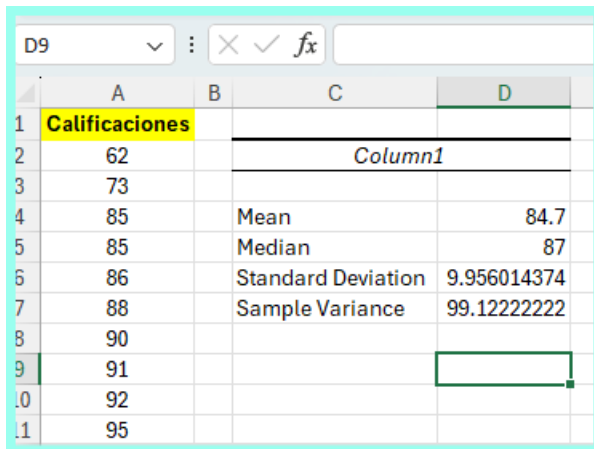
Para ello primero debemos calcular la media: $X = \frac{\sum X}{n} = (62 + 73 + 85 + 85 + 86 + 88 + 90 + 91 + 92 + 95) / 10 = 84.7$

El cálculo de la varianza se determina;

$$s^2 = \frac{\sum_{i=1}^n (X_i - X)^2}{n - 1}$$
$$s^2 = \frac{(62 - 84.7)^2 + (73 - 84.7)^2 + \dots + (95 - 84.7)^2}{10 - 1} = 99.12$$

Para obtener la desviación estándar, solamente obtenemos la raíz cuadrada de la varianza: $s = \sqrt{99.12} = 9.95$

Con la ayuda de Excel, utilizando Análisis de datos, quedaría:



	A	B	C	D
1	Calificaciones			
2	62		Column1	
3	73			
4	85	Mean		84.7
5	85	Median		87
6	86	Standard Deviation		9.956014374
7	88	Sample Variance		99.12222222
8	90			
9	91			
10	92			
11	95			

Utilizando Minitab:

↓	C1	C2	C3	C4	C5	C6	C7
	Calificaciones	Media1	Desv.Est1	Varianza1	Mediana1	N1	
1	62	84.7	9.95601	99.1222	87	10	
2	73						
3	85						
4	85						
5	86						
6	88						
7	90						
8	91						
9	92						
10	95						
11							

El coeficiente de variación, CV, nos ayuda a determinar la magnitud relativa de la desviación estándar en comparación con la media del comportamiento de las mediciones en porcentaje, es muy útil al comparar la variabilidad de dos o más conjuntos de datos en relación con el nivel general de los valores (y por tanto en relación con la media) de cada conjunto; fórmulas a emplear;

En el caso de una población; $CV = \left(\frac{\sigma}{\mu}\right) \times 100$

Para una muestra; $CV = \left(\frac{s}{\bar{x}}\right) \times 100$

Ejemplo: Para dos emisiones de acciones de la industria farmacéutica, el precio de cierre medio diario en el mercado diario durante un periodo de un mes fue para la acción A de \$150 con una desviación estándar de \$5. Para la acción B, el precio medio fue de \$50 con una desviación estándar de \$3. Basándose en una comparación absoluta, la variabilidad en el precio de la acción A fue mayor, debido a que la desviación estándar es mayor. Pero en relación con el nivel de precios, deben compararse cada coeficiente de variación:

$$CV(A) = \left(\frac{\sigma}{\mu}\right) \times 100 = (5/150) \times 100 = 3.3\%$$

$$CV(B) = \left(\frac{\sigma}{\mu}\right) \times 100 = (3/50) \times 100 = 6.0 \%$$

En relación con el nivel promedio de precios de cada emisión de acciones, se concluye que la acción B, ha tenido una variación casi del doble que la acción A.

El Coeficiente de asimetría de Pearson; mide el desvío de la simetría expresando la diferencia entre la media y la mediana en relación con la desviación estándar del grupo de mediciones; Para una distribución simétrica, el valor del coeficiente de asimetría siempre será cero, ya que la media y la mediana son de igual valor. En una distribución con sesgo positivo el valor del coeficiente es positivo; y en el caso de una distribución con sesgo negativo, el valor del coeficiente es negativo; fórmulas a utilizar:

$$\text{Asimetría de la población} = \frac{3(\mu - Med)}{\sigma}$$

$$\text{Asimetría de la muestra} = \frac{3(X - Med)}{s}$$

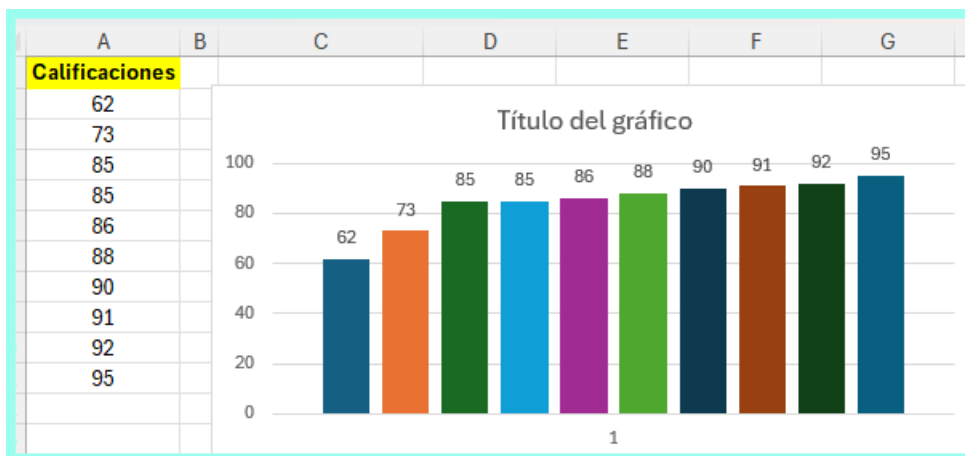
Ejemplo 7: Retomando del ejemplo 6, de las 10 calificaciones; donde:

Media muestral = 84.7

Mediana = 87

Desviación estándar = 9.95

El coeficiente de asimetría de Pearson = $\frac{3(84.7-87)}{9.95} = -0.693$; presentando una asimetría con sesgo a la izquierda.



Ejemplo de medidas de tendencia central media, mediana, moda; medidas de posición cuartiles, deciles y percentiles; medidas de dispersión rango, varianza, desviación estándar, coeficiente de variación y de Pearson. Para datos agrupados.

1.- Durante el año 2022 en el mes de junio, en la localidad de Cd. Cuauhtémoc se registraron las siguientes temperaturas en grados °C: 18,17, 20,16,15,19,15,18, 21, 20, 22, 23,19,18,14,15,16,18, 22, 21, 20,19, 20, 20,16,17, 18, 16, 18, 24. Considere los datos como una población; construya la distribución de frecuencias con seis intervalos de clase y obtenga las siguientes medidas:

a) Media, mediana, moda.

b) cuartil dos y cuartil tres (Q_2 y Q_3), decil cuatro y decil seis (D_4 y D_6), percentil treinta y cinco y percentil setenta (P_{35} y P_{70}).

c) rango, varianza, desviación estándar, coeficiente de variación y de Pearson.

Pasos:

1. Ordenar los datos:

A	B
Temperatura	
18	14
17	15
20	15
16	15
15	16
19	16
15	16
18	16
21	17
20	17
22	18
23	18
19	18
18	18
14	18
15	18
16	19
18	19
22	19
21	20
20	20
19	20
20	20
20	20
16	21
17	21
18	22
16	22
18	23
24	24

2. Obtener el Rango: Dato mayor – dato menor = $24 - 14 = 10$

3. Obtener la anchura o tamaño del intervalo: $c = \frac{\text{Rango}}{\text{Número de clases}} = \frac{10}{6} = 1.66 \cong 2$

4. Construir la tabla de distribución de frecuencias:

No. De clase	Int. De clase	Frec.	Marca de clase X_i	Límites reales	Frec. Acum.	$f(x_i - \bar{x})^2$	$f(x_i)$
1	14 – 15	4	14.5	13.5 – 15.5	4	68.23	58
2	16 – 17	6	16.5	15.5 – 17.5	10	27.22	99
3	18 – 19	9	18.5	17.5 – 19.5	19	0.12	166.5
4	20 – 21	7	20.5	19.5 – 21.5	26	24.48	143.5
5	22 – 23	3	22.5	21.5 – 23.5	29	44.93	67.5
6	24 – 25	1	24.5	23.5 – 25.5	30	34.46	24.5
		30				199.467	559

$$a) \text{ Media} = \mu = \frac{\sum_{i=1}^N fX_i}{N} = \frac{559}{30} = 18.63$$

$$\text{Mediana} = LRI_{kmed} + \left[\frac{\frac{n}{2} - (\sum f)_{k \text{ inf a } k \text{ med}}}{f_{k \text{ mediana}}} \right] c = 17.5 + \left[\frac{\frac{30}{2} - 10}{9} \right] 2 = 18.61$$

$$Moda = LRI_{k\ modal} + \left[\frac{\Delta_1}{\Delta_1 + \Delta_2} \right] c = 17.5 + \left[\frac{3}{3+2} \right] 2 = 18.70$$

$$b) Q_X = LRI_{Q_X} + \left(\frac{\frac{xn}{4} - (\sum f_{Q_X})}{f_{Q_X}} \right) c = Q_2 = 17.5 + \left(\frac{\frac{2*30}{4} - 10}{9} \right) 2 = 18.61$$

$$Q_3 = 19.5 + \left(\frac{\frac{3*30}{4} - 19}{7} \right) 2 = 20.5$$

$$D_X = LRI_{D_X} + \left[\frac{\frac{xn}{10} - (\sum f_{D_X})}{f_{D_X}} \right] c$$

$$D_4 = 17.5 + \left[\frac{\frac{4*30}{10} - 10}{9} \right] 2 = 17.94 \quad D_6 = 17.5 + \left[\frac{\frac{6*30}{10} - 10}{9} \right] 2 = 19.29$$

$$P_X = LRI_{P_X} + \left[\frac{\frac{xn}{100} - (\sum f_{P_X})}{f_{P_X}} \right] c$$

$$P_{35} = 15.5 + \left[\frac{\frac{35*30}{100} - 4}{6} \right] 2 = 17.67 \quad P_{70} = 19.5 + \left[\frac{\frac{70*30}{100} - 19}{10} \right] 2 = 19.9$$

$$c) Rango = 24 - 10 = 14$$

$$varianza\ poblacional\ \sigma^2 = \frac{\sum_{i=1}^N f(X_i - \mu)^2}{N}$$

$$\sigma^2 = \frac{\sum_{i=1}^N f(X_i - \mu)^2}{N} = \frac{199.467}{30} = 6.65$$

$$Desviación\ estándar\ \sigma = \sqrt{\sigma^2} = \sqrt{6.65} = 2.58$$

$$\text{Coeficiente de variación población } CV = \frac{\sigma}{\mu} \times 100 \quad CV = \frac{2.58}{18.7} * 100 = 13.80\%$$

$$\text{Coeficiente de asimetría de Pearson} = \frac{3(\mu - Mediana)}{\sigma} = \frac{3(18.7 - 18.61)}{2.58} = 0.105$$

Ejercicios:

1. En una muestra de 15 estudiantes en la cafetería de una escuela de nivel preparatoria, todos en dólares se tienen los siguientes montos de ventas en orden ascendente: 0.10, 0.10, 0.25, 0.25, 0.35, 0.40, 0.53, 0.90, 1.25, 1.35, 2.45, 2.71, 3.09, 4.10. Determine: a) la media, b) la mediana y c) la moda.

a) $\bar{X} = \$1.2$
b) *mediana* = \$0.53
c) *moda* = \$0.25
2. Para los datos del problema anterior determine los siguientes valores a) segundo cuartil, b) segundo decil y c) percentil 40.

a) $Q_2 = \$0.53$
b) $D_2 = \$0.25$
c) $P_{40} \cong \$0.38$
3. En una muestra de paquetes que salen de una oficina de correos, se encuentra que los pesos, dados hasta la onza más cercana, son: 21, 18, hoy 30, 12, 14, 17, 28, 10, 16, 25 onzas. Determine: a) la media, b) la mediana y c) la moda.

a) $\bar{X} = 19.1$
b) *mediana* = 17.5
c) *no presenta moda*
4. Con los datos del problema anterior determine los pesos correspondientes a) al tercer cuartil, b) al tercer decil y al c) percentil 70.

a) $Q_3 = 25 \text{ oz}$
b) $D_3 = 15 \text{ oz}$
c) $P_{70} = 23 \text{ oz}$
5. Con los datos del problema número uno. Determine a) el rango, b) rango intercuartílico, c) la desviación media absoluta y d) hoy la desviación estándar muestral.

a) $R = \$4.00$
b) $RIC \cong \$1.92$
c) $DMA = \$1.03$
d) $s \cong \$1.28$

Unidad 4



Probabilidad y teoría de conjuntos

4.1. Aspectos generales de la probabilidad (Conceptos, tipos de probabilidad y enfoques)

Es importante señalar la gran relevancia que implica el poder entender y aplicar algunos de los conceptos de la probabilidad en el campo de la estadística tanto descriptiva como inferencial, ambas ciencias van a la par.

Según L. Kazmier (4 ed. 2006)" se han desarrollado tres métodos conceptuales para definir la probabilidad y para determinar los valores de probabilidad: *el método clásico*, *el método de la frecuencia relativa* y *el método subjetivo*".

- **Método clásico de probabilidad**, si $N(A)$ son resultados elementales posibles favorables al evento A , $N(S)$ son los resultados posibles en el espacio muestral, siendo éstos igualmente probables y mutuamente excluyentes (cuando dos más eventos no pueden suceder al mismo tiempo), la probabilidad de que ocurra el evento A es; $P(A) = \frac{N(A)}{N(S)}$

También conocido este método como, *método a priori*, por la determinación de los valores de probabilidad antes de observar cualquier evento muestral; ejemplo:

Determinar la probabilidad al lanzar un dado una sola vez resulte un número par.

$$N(A) = \{2,4,6\} , N(S) = \{1,2,3,4,5,6\}$$

$$P(A) = \frac{N(A)}{N(S)} = \frac{3}{6} = \frac{1}{2}$$

- **Método de la frecuencia relativa**, para determinar la probabilidad por este método es en base a la proporción del número de veces que se presenta un resultado favorable en un número de observaciones, ensayos o experimentos; en este caso no interviene ninguna suposición de igualdad de posibilidades. También conocido como *método empírico*, debido a que la determinación de los valores de probabilidad se basa en la observación y recolección de datos. Fórmula:

$$P(A) = \frac{\text{número de observaciones de } A}{\text{tamaño de la muestra}} = \frac{n(A)}{n}; \text{ ejemplo:}$$

Determinar la probabilidad de que un trabajador en cierta industria, se viera afectado por el COVID 19, en este año del 2024, si en una muestra aleatoria de diez industrias del corredor comercial que emplean en total 3500 trabajadores, arrojó que 250 trabajadores se vieron afectados por COVID 19, durante el año

$$2023, P(A) = \frac{\text{número de observaciones de } A}{\text{tamaño de la muestra}} = \frac{n(A)}{n} = \frac{250}{3500} = 0.071$$

Nota: De acuerdo con L. Kazmier (4 ed. 2006). Los métodos anteriores arrojan valores de probabilidad *objetivos*, dado que los valores de probabilidad indican la tasa relativa de ocurrencia del evento a largo plazo.

- **Método subjetivo**, este método es apropiado cuando se presenta una oportunidad de que ocurra el evento, y éste sucederá o no solamente en esa ocasión; basado en las evidencias que se tienen, se obtiene cierto grado de seguridad de ocurrencia del evento; ejemplo:

Determinar la probabilidad de que un nuevo tipo de teléfono celular logre tener éxito financiero. Un futuro inversionista obtiene datos de otros modelos similares de teléfonos celulares, analiza el impacto financiero en ciertas regiones, así como el volumen de ventas requerido para lograr el éxito en las finanzas, considerando la inversión y los costos de operación. En general, el criterio del inversionista es que existe un 85% de probabilidad de que dicho teléfono celular tenga éxito financiero y un 15% de probabilidad de que no lo tenga. El inversionista al basarse en el conocimiento de toda la información dentro del lapso disponible para obtener dicha información considera aceptable la probabilidad de éxito del 85%.

4.2. Leyes de la probabilidad

Las leyes de la probabilidad son reglas matemáticas que se aplican a eventos aleatorios y que permiten calcular las posibilidades de que ocurran, siendo las más importantes:

1. **Ley de la suma:** La probabilidad de que ocurra uno de dos eventos mutuamente excluyentes es igual a la suma de sus probabilidades individuales, expresado como: $P(A \text{ o } B) = P(A) + P(B)$
2. **Ley del complemento:** La probabilidad de que un evento no ocurra es igual a uno menos la probabilidad de que sí ocurra, expresado como: $P(\bar{A}) = 1 - P(A)$

3. Ley del producto: La probabilidad de que ocurran dos eventos independientes es igual al producto de sus probabilidades individuales. Esto se expresa como:

$$P(A \text{ y } B) = P(A) \times P(B)$$

4. **Ley de la probabilidad condicional:** La probabilidad de que ocurra un evento dado que otro evento ya ha ocurrido se calcula utilizando la probabilidad condicional y para ello utilizamos la siguiente fórmula: $P(A|B) = P(A \cap B)/P(B)$ y se lee “la probabilidad de A dado que B ha ocurrido”

Es importante señalar los siguientes axiomas dentro de la probabilidad:

1. Sea **S** un espacio muestral. Entonces **$P(S) = 1$**
2. Para cualquier evento **A**, **$0 \leq P(A) \leq 1$**
3. Si **A** y **B** son eventos mutuamente excluyentes, entonces **$P(A \cup B) = P(A) + P(B)$** de forma más general, si $A_1, A_2, \dots$ son eventos mutuamente excluyentes entonces **$P(A_1 \cup A_2 \cup \dots) = P(A_1) + P(A_2) \dots$**

Abordemos el concepto de *eventos mutuamente excluyentes y no excluyentes*. Se dice que dos o más eventos son *mutuamente excluyentes o disyuntos* si no pueden ocurrir de manera conjunta. Es decir, la ocurrencia de un evento excluye en forma automática la ocurrencia del otro evento (o eventos), Kazmier (2006).

Por ejemplo, en la extracción de una carta de un juego de naipes, no puede resultar que dicha carta pueda ser as o rey al mismo tiempo; por lo tanto, se dice que son eventos mutuamente excluyentes. Sin embargo, en la extracción de una carta en un juego de naipes, ésta puede resultar as y espada. Estos eventos no son mutuamente excluyentes; sin embargo, no se concluye que todo as sea una espada o que toda espada sea un as. Al determinar la probabilidad de sacar ya sea un as o un rey en una sola extracción de un juego de naipes, siendo los eventos “as” (A) y “rey” (B), tendremos:

$$P(A \text{ o } R) = P(A) + P(B) = \frac{4}{52} + \frac{4}{52} = \frac{8}{52} = \frac{2}{13}$$

Para eventos que no son mutuamente excluyentes, la probabilidad de la ocurrencia conjunta de los dos eventos se resta de la suma de las probabilidades simples de estos dos eventos. $P(A \text{ o } B) = P(A) + P(B) - P(A \cap B)$

Al determinar la probabilidad de obtener un as (A) o una espada (E) (o ambos) en una sola extracción de una baraja de naipes, estos eventos no son mutuamente excluyentes, tendremos:

$$P(A \text{ o } E) = P(A) + P(E) - P(A \cap E) = \frac{4}{52} + \frac{13}{52} - \frac{1}{52} = \frac{16}{52} = \frac{4}{13}$$

Otro aspecto que debemos considerar en probabilidad, son los eventos *independientes y dependientes*; ya que dos eventos son *independientes* si la ocurrencia o no ocurrencia de uno no tiene efecto sobre la probabilidad de ocurrencia del otro. Dos eventos son

dependientes cuando la ocurrencia o no ocurrencia de uno *sí* afecta la probabilidad de ocurrencia del otro.

Para abordar lo anterior consideremos los resultados de lanzar dos veces seguidas una moneda legal, en este caso se considera que son eventos *independientes*, ya que el resultado del primer lanzamiento no tiene efecto sobre las respectivas probabilidades de ocurrencia de cara o sello en el segundo lanzamiento.

En el caso de extraer dos cartas *sin reemplazo* de una baraja de naipes son eventos *dependientes*, debido a que las probabilidades correspondientes con la segunda extracción dependen del resultado de la primera extracción, supongamos que deseamos saber la probabilidad de que aparezca un as en ambas cartas, para la primera el resultado es $4/52$, y para la segunda carta la probabilidad de que aparezca un as sería de $3/51$.

4.3. Aplicaciones de la probabilidad en la administración

La probabilidad tiene múltiples aplicaciones en la administración, la cual nos puede ayudar en la toma de decisiones, gestión de riesgos, análisis financiero, mercadotecnia, planificación estratégica; entre otras. Algunas de las aplicaciones que más se emplean en la administración con más detalle son:

1. **Análisis de riesgos:** La probabilidad se utiliza para cuantificar los riesgos y las incertidumbres en la toma de decisiones. Los administradores pueden utilizar la probabilidad para evaluar los riesgos asociados con diferentes estrategias y tomar decisiones informadas para minimizarlos o eliminarlos.
2. **Control de calidad:** La probabilidad se utiliza para controlar la calidad de los productos y servicios. Los administradores pueden utilizar la probabilidad para analizar los datos de calidad y determinar si los procesos de producción están bajo control.
3. **Planificación de proyectos:** La probabilidad se utiliza para planificar y programar proyectos. Los administradores se apoyan en la probabilidad para estimar el tiempo y los recursos necesarios para concretar un proyecto y para determinar las probabilidades de cumplimiento de las fechas de entrega.
4. **Investigación de mercados:** En este aspecto la probabilidad se utiliza para realizar investigaciones de mercado y analizar los datos de los consumidores. Los administradores pueden utilizar la probabilidad para estimar las preferencias de los consumidores, predecir la demanda de productos y servicios y realizar análisis de sensibilidad.
5. **Finanzas y economía:** En estas áreas la probabilidad se utiliza para modelar y predecir los precios de los activos financieros, las tasas de interés entre otros indicadores económicos. De igual forma para tomar decisiones de inversión informadas y poder evaluar los riesgos asociados con diferentes inversiones.
6. **Pronóstico de demanda:** En este caso la probabilidad se utiliza en la planificación de la producción y la gestión de inventarios para pronosticar la demanda futura de

productos y servicios, ayudando a las empresas a planificar la producción y la gestión de inventarios de manera eficiente.

7. **Diseño de experimentos:** La probabilidad se utiliza en el diseño de experimentos para determinar la probabilidad de que una variable independiente tenga un efecto significativo en una variable dependiente.
8. **Planificación financiera:** Aquí se aplica la probabilidad para evaluar el riesgo y la incertidumbre asociados con diferentes estrategias financieras; es decir, una empresa puede utilizar la probabilidad para evaluar la posibilidad de que un plan de inversión tenga éxito o fracaso en función de diferentes factores económicos y de mercado.

4.4. Árboles de probabilidad

Los árboles de probabilidad son una herramienta de análisis de decisiones utilizada en la administración para poder visualizar y evaluar las diferentes opciones y resultados posibles en un contexto; se construyen utilizando ramificaciones que representan diferentes posibilidades y sus correspondientes resultados.

Estructura básica de un árbol de probabilidad:

1. **Nodos:**

- Los representan eventos o decisiones.
- Los nodos pueden ser de dos tipos: nodos de probabilidad y nodos de decisión.

2. **Ramas:**

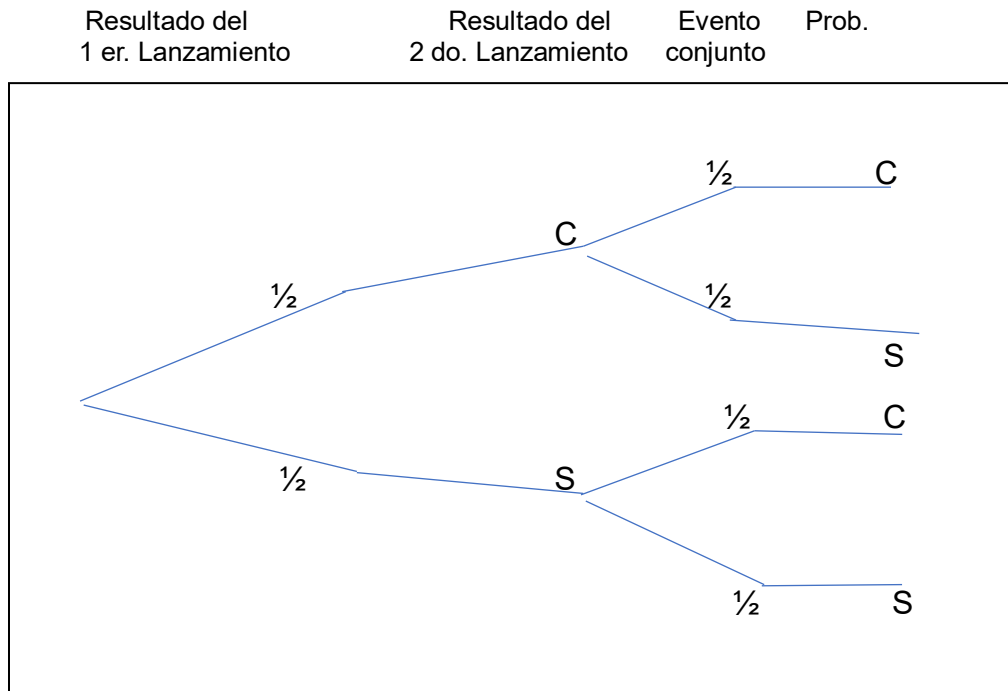
- Las ramas conectan los nodos y representan los posibles resultados de un evento o decisión.

3. **Probabilidades:**

- Se asigna una probabilidad a cada rama que indica la probabilidad de que ocurra el evento correspondiente.

4. **Calcular probabilidades condicionales:**

- Se sigue las ramas correspondientes a los eventos condicionales y se multiplican las probabilidades a lo largo de esas ramas.



4.5 Teorema de Bayes

Este teorema es aplicable en una gran gama de contextos dentro de la administración, como la gestión de proyectos, la aplicación financiera y la toma de decisiones estratégicas; por ejemplo, que una empresa desea actualizar la probabilidad de éxito de un nuevo producto en el mercado en función de la información adicional que se recopila de la competencia, las condiciones económicas y las preferencias del consumidor.

La fórmula general del teorema de Bayes es: $P(A|B) = \frac{P(A \text{ y } B)}{P(B)}$

Cabe señalar la importancia especial del teorema de Bayes en que puede aplicarse en el contexto de eventos sucesivos, para ello utilizaremos la siguiente fórmula:

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \dots + P(B|A_k)P(A_k)}$$

Consideremos el siguiente ejemplo para la aplicación de la fórmula anterior.

Ejemplo: La empresa cellphone esta considerando introducir al mercado un nuevo modelo de celular. En el pasado, el 40% de los teléfonos celulares que la empresa lanzó al mercado tuvieron éxito y el 60% no fueron exitosos. Antes de lanzar al mercado el nuevo modelo de celular, el equipo de investigación de mercados realiza un estudio y entrega un reporte, ya sea favorable o desfavorable. En el pasado, el 80% de los celulares exitosos habían recibido un reporte de investigación favorable y

el 30% de los celulares no exitosos habían recibido un reporte de investigación favorable. Para los nuevos modelos de celular bajo consideración, el equipo de investigación de mercado ha entregado un reporte favorable. ¿Cuál es la probabilidad de que el nuevo modelo de celular tenga éxito en el mercado?

Sea: evento E=equipo de celular exitoso evento F=reporte favorable

evento E´=equipo de celular no exitoso evento F´=reporte desfavorable

por lo tanto: $P(E) = 0.40$ $P(F|E) = 0.80$

$P(E') = 0.60$ $P(F|E') = 0.30$

Al aplicar la fórmula para eventos sucesivos, tendremos:

$$P(E|F) = \frac{P(F|E)P(E)}{P(F|E)P(E) + P(F|E')P(E')}$$
$$P(E|F) = \frac{(0.80)(0.40)}{(0.80)(0.40) + (0.30)(0.60)} = \frac{0.32}{0.32 + 0.18} = \frac{0.32}{0.50} = 0.64$$

La probabilidad de un equipo de celular exitoso, dado que se recibió un reporte favorable, es de 0.64

4.6 Teoría de conjuntos aplicados en la administración

La teoría de conjuntos es una herramienta matemática que se utiliza en la administración para analizar y clasificar elementos o datos en diferentes conjuntos. En la administración, la teoría de conjuntos se aplica en una variedad de contextos, como la toma de decisiones, la planificación de proyectos y la gestión de recursos.

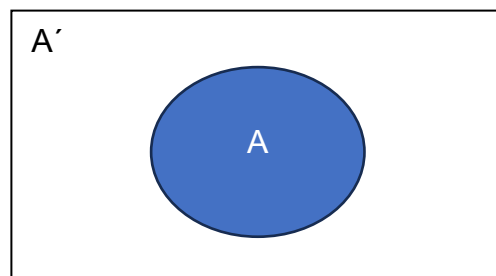
Ejemplos de como se aplica la teoría de conjuntos en la administración:

1. Segmentación de mercado: Para identificar diferentes grupos de clientes y segmentos de mercado en función de ciertos criterios, como la edad, el género, la ubicación geográfica o los hábitos de compra. Cada grupo de clientes se puede considerar como un conjunto y la empresa puede utilizar la teoría de conjuntos para analizar la intersección y la unión de estos conjuntos.
2. Asignación de recursos: Al asignar recursos a diferentes proyectos o departamentos, se puede utilizar la teoría de conjuntos para analizar la intersección y la unión de estos conjuntos para determinar como asignar los recursos de manera eficiente.
3. Evaluación de riesgos: Para evaluar la probabilidad y el impacto de diferentes riesgos asociados con diferentes eventos o escenarios al analizar la intersección y la unión de estos conjuntos.
4. Análisis de competencia: Los competidores pueden agruparse en diferentes conjuntos según su tamaño, posición en el mercado, estrategias y otros factores

relevantes, apoyando a las empresas a entender mejor su entorno competitivo y a adaptar sus estrategias de negocio.

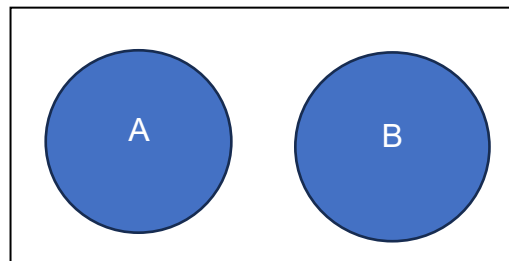
5. Gestión de inventarios: Para clasificar los productos en diferentes conjuntos según su demanda, rotación y rentabilidad; lo cual ayuda a las empresas a optimizar sus niveles de inventario y a minimizar los costos inherentes.

Para relacionar algunos aspectos de la teoría de conjuntos dentro de la administración, es necesario conocer sobre el *diagrama de ven*, el cual nos permite representar los eventos que pueden en una observación o experimento específico. Para ello se utiliza una figura cerrada, la cual representa al espacio muestral (todos los posibles eventos o resultados), y las porciones del área dentro de este espacio representa los eventos básicos.

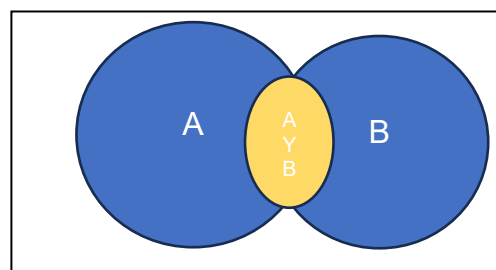


En esta figura A' representa la probabilidad del evento “no A ” y “ A ” representa la probabilidad de “ A ”.

La unión de dos eventos *mutuamente excluyente* en un diagrama de Venn, se representa



Para la representación en un diagrama de Venn, en el caso de dos eventos que *no* son *mutuamente excluyentes*, es:



Consideremos el ejemplo del libro de Levine, Krehbiel y Berenson (2006), para el caso de teoría de conjuntos; En el cual el gerente de marketing, analiza los resultados de una encuesta realizada en 1000 hogares concerniente a sus intenciones de comprar un equipo de televisión de pantalla grande en los siguientes 12 meses. Investigaciones de este tipo reciben el nombre de estudios de intención de compra. Como seguimiento encuestará los mismos hogares 12 meses después para ver si realmente compraron el equipo. Además, al gerente de marketing le interesa saber si quienes compraron la televisión de pantalla grande también compraron una televisión de alta definición, se adquirieron un equipo reproductor de DVD en los últimos 12 meses, y si estuvieran satisfechos con la compra del equipo de televisión de pantalla grande. Algunas de las preguntas que se pueden plantear son:

¿Cuál es la probabilidad de que en un hogar se planeé comprar un televisor de pantalla grande el año próximo?

¿Cuál es la probabilidad de que en ese hogar se compre realmente un televisor de pantalla grande?

¿Cuál es la probabilidad de que un hogar donde se compra un televisor de pantalla grande estará satisfecho de su compra?

Entre otras preguntas la teoría de conjunto es muy útil para dar respuesta a varias interrogantes que intervienen en la administración.

Y se pueden desarrollar futuras estrategias de ventas y de marketing.

Planea comprarlo	Realmente lo compró		Total
	Sí	No	
Sí	200	50	250
No	100	650	750
Total	300	700	1000

¿Qué es un espacio muestral? Dé ejemplos de eventos simples y eventos conjuntos.

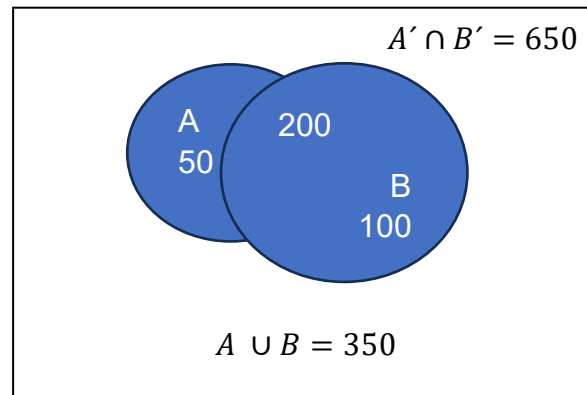
SOLUCIÓN: El espacio muestral consiste en las 1000 personas encuestadas. Los eventos simples son “planea comprarlo”, “no planea comprarlo”, “compra” y “no compra”. El complemento del evento “planea comprarlo” es “no planea comprarlo”. El evento “planea comprarlo y realmente lo compra” es un evento conjunto porque quien responde debe planear comprar la televisión y realmente comprarla.

Definir los eventos: A = planea comprar B = realmente compró

A' = no planea comprar B' = no lo compró

Al diseñar un diagrama de Venn se debe determinar el valor de la intersección de $A \cap B$ para dividir el espacio muestral en sus partes. $A \cap B$ consiste en los 200 hogares en los que se planeó comprar y realmente se compró el equipo de televisión de pantalla grande. El evento remanente A (planean comprar), consiste en 50 hogares en los que se planeó

comprar el televisor de pantalla grande pero que finalmente no lo compraron. El remanente del evento B (realmente compraron) consiste en 100 hogares en los que no se planeó comprar un equipo de televisión de pantalla grande, pero en los que finalmente se compró uno. El resto de los 650 hogares representa aquellos que ni planearon ni compraron un equipo de televisión de pantalla grande.



Probabilidad simple (marginal)

Dado que los resultados se basan en los datos recolectados en una encuesta, es posible utilizar el punto de vista de la probabilidad clásica empírica.

Debemos recordar que un evento imposible tiene una probabilidad de cero y un evento seguro de ocurrencia tiene una probabilidad de uno.

A la probabilidad simple también se le conoce como *probabilidad marginal*, por qué es posible calcular el número total de los éxitos (para este ejemplo el número total de quienes planearon comprar a partir del margen apropiado que podemos observar en la tabla de contingencia).

Al determinar la probabilidad de seleccionar un hogar en el que se plantee comprar un equipo de televisión de pantalla grande, podemos emplear la fórmula de probabilidad de ocurrencia.

$$P(A) = \frac{N(A)}{N(S)}, P(\text{planear comprar}) = \frac{250}{1000} = 0.25$$

Esto significa, que existe un 25% de probabilidad de que en un hogar se planea comprar un equipo de televisión de pantalla grande.

La probabilidad conjunta se refiere a la probabilidad de ocurrencia que implica a 2 o más eventos.

Volviendo a nuestro ejemplo, aquellas familias que planearon comprar y realmente compraron el televisor de pantalla grande se identifican con los resultados de la celda “sí- planearon comprar y sí - realmente lo compraron”. Por lo tanto, la probabilidad de elegir un hogar que planean comprar y realmente lo compra es:

$$P(\text{planea comprar y realmente lo compra}) = \frac{200}{1000} = \mathbf{0.20 \text{ o } 20\%}$$

Dado que la probabilidad de ocurrencia utiliza el número total de resultados posibles, en muchas ocasiones existe un gran número de posibles resultados, siendo por lo tanto complicado determinar el número exacto; para ello observaremos algunas reglas para contar el número posible de resultados.

Regla de conteo 1. Si cualquiera de los eventos K mutuamente excluyentes y colectivamente exhaustivos pueden ocurrir en cada uno de los ensayos n , el número de posibles resultados es igual a K^n .

Ejemplo: Suponga que se lanza una moneda al aire 5 veces. ¿Cuál es el número de diferentes resultados posibles (la secuencia de caras y cruces)? $R = 2^5 = 32$

Regla de conteo 2. Si hay k_1 Eventos en el primer ensayo, k_2 eventos en el segundo ensayo, ... y k_n eventos en el n -ésimo ensayo, entonces el número posible de resultados es $(k_1)(k_2) \dots (k_n)$.

Ejemplo: El menú de un restaurante tiene un precio fijo para las cenas completas que consiste en un aperitivo, un platillo principal, una bebida y un postre. Se tiene la posibilidad de elegir entre 5 aperitivos, 10 platillos, 3 bebidas y 6 postres. Determine el número total de cenas posibles. $R = (5)(10)(3)(6) = 900$

Regla de conteo 3. El número de maneras en el que n objetos pueden arreglarse en orden es, $n! = (n)(n-1)(n-2) \dots (n-n+1)$

Ejemplo: Si un conjunto de 5 libros de texto se colocan en un librero, ¿de cuántas formas es posible ordenar estos 5 libros de texto? $R = 5! = (5)(4)(3)(2)(1) = 120$

Regla de conteo 4. En muchos casos se necesita saber el número de formas en las que un subconjunto de un grupo completo de elementos puede arreglarse en *orden*. Cada posible arreglo es llamado **permutación**, Para ello se aplica $p_x^n = \frac{n!}{(n-x)!}$

Ejemplo: Suponga el ejemplo anterior de los 6 libros de texto, pero sólo hay espacio para cuatro libros en un librero, ¿de cuántas maneras es posible acomodar estos libros en el librero? $p_4^6 = \frac{n!}{(n-x)!} = \frac{6!}{(6-4)!} = \frac{(6)(5)(4)(3)(2)(1)}{(2)(1)} = 360$

Regla de conteo 5. En muchas situaciones *no* interesa el *orden* de los resultados, únicamente el número de maneras en las que x objetos pueden seleccionarse a partir de

n elementos, *sin consideración de orden*. A esta regla se le llama la regla de las **combinaciones**, Para ello se aplica $C_x^n = \frac{n!}{x!(n-x)!}$

Ejemplo: Suponga que en el ejemplo de los 6 libros y queremos acomodar a sólo cuatro de ellos sin importar el orden, ¿de cuántas maneras es posible realizar dicha operación?

$$C_4^6 = \frac{6!}{4!(6-4)!} = \frac{(6)(5)(4)(3)(2)(1)}{(4)(3)(2)(1)(2)(1)} = 15$$

Ejercicios:

1. En cada una de las siguientes situaciones, indique cuál de los métodos (clásico, de la frecuencia relativa o subjetivo) sería el más útil para la determinación del valor de probabilidad requerido.
 - a) La probabilidad de una recesión el próximo año.
 - b) La probabilidad de que un dado con 6 caras muestre ya sea un 6 o un uno en un solo lanzamiento.
 - c)Cuál es la probabilidad de qué de un lote de 20 piezas, hoy de las cuales se sabe cuenta que una de ellas está defectuosa, una pieza elegida de manera aleatoria resulte ser la defectuosa.
 - d) La probabilidad de que una pieza tomada en forma aleatoria de un lote grande de piezas resulte defectuosa.
 - e) La probabilidad de que una persona, elegida de manera aleatoria, que entra en una tienda grande de departamentos efectúe una compra.
 - f) La probabilidad de que el promedio industrial aumente por lo menos 50 puntos durante los siguientes 6 meses.

Solución:

- a) Subjetivo.
 - b) Clásico.
 - c) Clásico.
 - d) Frecuencia relativa.
 - e) Frecuencia relativa
 - f) Subjetivo.
2. Determine el valor de probabilidad aplicable a cada una de las siguientes situaciones.
 - a) La probabilidad de sufrir un accidente de trabajo en una industria específica a lo largo de 1 año. Una muestra aleatoria de 10 empresas que emplean en total 8000 personas reveló que ocurrieron 400 accidentes de trabajo en un periodo reciente de 12 meses.
 - b) La probabilidad de apostar el número ganador en la ruleta. Los números de la rueda son 0, 00 y del uno hasta el 36.

Solución:

$$\begin{aligned} \text{a) } P &= \frac{400}{8000} = 0.05 \\ \text{b) } P &= \frac{1}{38} \end{aligned}$$

3. Para cada una de las siguientes oportunidades relativas determine el valor de la probabilidad correspondiente, y para cada uno de los valores de probabilidad determine las oportunidades relativas equivalentes.
- a) Un agente de compras estima que las oportunidades relativas de que un flete llegue a tiempo son 2:1
 - b) La probabilidad de que un nuevo componente no funcione de manera apropiada cuando se ensamble se calcula que es $P = 1/5$.
 - c) Las oportunidades relativas de que un nuevo producto tenga éxito se estima que son 3:1
 - d) La probabilidad de que el equipo local gane el juego inicial de la temporada se calcula que es $1/3$.

Solución:

- a) *La probabilidad de que el flete llegue a tiempo es $P = \frac{2}{2+1} = \frac{2}{3} \cong 0.67$*
- b) *Las oportunidades de que no funcione de manera apropiada son 1:4*
- c) *La probabilidad de que el producto tenga éxito es $P = \frac{3}{3+1} = \frac{3}{4} = 0.75$*
- d) *Las oportunidades relativas de que el equipo gane son 1:2*

4. De 300 estudiantes de administración, 100 se encuentran inscritos en contabilidad y 80 están inscritos en estadística para la administración. Estas cifras incluyen a 30 estudiantes que están inscritos en ambos cursos. ¿Cuál es la probabilidad de que un estudiante elegido de manera aleatoria este inscrito en contabilidad (A) o en estadística para la administración (B)?

Solución:

$$P(A \text{ o } B) = P(A) + P(B) - P(A \text{ y } B) = \frac{100}{300} + \frac{80}{300} - \frac{30}{300} = \frac{150}{300} = 0.50$$

5. Cuando de una población finita se toma una muestra *sin* reemplazo, los valores de probabilidad correspondientes a los diferentes eventos depende de qué eventos (elementos muestreados) ya hayan ocurrido. Por otro lado, cuando el muestreo se hace *con* reemplazo, los eventos siempre son independientes.
- a) Suponga que se escogen al azar y sin reemplazo 3 cartas de una baraja de 52 naipes. ¿Cuál es la probabilidad de que las 3 cartas sean ases?
 - b) Suponga que se escogen en forma aleatoria 3 cartas de una baraja de 52 naipes, pero después de elegida, cada carta se devuelve al paquete que se baraja antes de la siguiente extracción de una carta. ¿Cuál es la probabilidad de que las 3 cartas sean ases?

Solución:

a) Se aplica la regla de multiplicación para eventos dependientes:

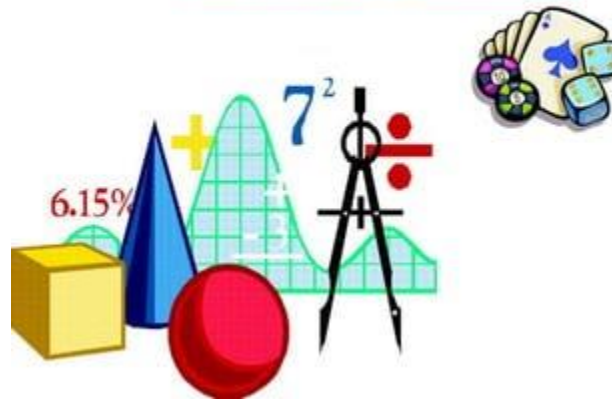
$$\begin{aligned} P(A_1 \text{ y } A_2 \text{ y } A_3) &= P(A_1)P(A_2|A_1)P(A_3|A_1 \text{ y } A_2) \\ &= \frac{4}{52} \times \frac{3}{51} \times \frac{2}{50} = \frac{24}{132600} = \frac{1}{5525} \cong 0.0002 \end{aligned}$$

b) Se aplica la regla de multiplicación para eventos independientes:

$$P(A_1 \text{ y } A_2 \text{ y } A_3) = P(A)P(A)P(A) = \frac{4}{52} \times \frac{4}{52} \times \frac{4}{52} = \frac{4}{140608} \cong 0.0005$$

Unidad 5

Distribución de la probabilidad



Distribuciones de probabilidad aplicadas en la administración

5.1. Distribuciones para variables discretas

Recordemos de la unidad número uno, que una variable aleatoria se define como un evento numérico cuyo valor se determina mediante un proceso aleatorio. Si a cada uno de los posibles valores numéricos de una variable aleatoria X se le asigna un valor de probabilidad, ya sea mediante una lista o una función matemática, el resultado es una *distribución de probabilidad*. La suma de las probabilidades de todos los posibles resultados numéricos debe ser 1. Cada uno de los valores de probabilidad se puede denotar mediante el símbolo $f(x)$, que indica que interviene una función matemática, mediante $P(x)$.

Sí se trata de una *variable aleatoria discreta*, los valores observados sólo pueden presentarse en puntos aislados a lo largo de una escala de valores.

Ejemplo: Demanda diaria de alquiler de vehículos de la empresa *Hertz* durante un periodo de 50 días.

	A	B	C
1	Demanda posible X	No. De días	P(X)
2	3	3	0.06
3	4	7	0.14
4	5	12	0.24
5	6	14	0.28
6	7	10	0.20
7	8	4	0.08
8		50	1.00
9			

Con frecuencia es útil describir una variable aleatoria en términos de su *media* y de su *varianza* o *desviación estándar*. La media (a largo plazo) de una variable aleatoria X se conoce como *valor esperado* y se denota por $E(X)$. En el caso de una variable aleatoria discreta, la media es el promedio ponderado de todos sus posibles valores numéricos empleando sus respectivas probabilidades como factores de ponderación, quedando de la siguiente manera: $E(X) = \sum XP(X)$

	A	B	C	D
	Demanda posible X	No. De días	P(X)	Valor ponderado
				[XP(X)]
	3	3	0.06	0.18
	4	7	0.14	0.56
	5	12	0.24	1.2
	6	14	0.28	1.68
	7	10	0.20	1.4
	8	4	0.08	0.64
		50	1.00	5.66

Para determinar el valor de la varianza de una variable aleatoria discreta, en donde no se requiere determinar las desviaciones respecto de la media, se emplea la siguiente fórmula: $V(X) = \sum x^2P(X) - [\sum XP(X)]^2$

$$V(X) = E(X^2) - [E(X)]^2$$

	A	B	C	D	E	F
1	Demanda posible X	No. De días	P(X)	Valor ponderado	Cuadrado de la	Cuadrado del valor
2				[XP(X)]	demanda (X²)	ponderado [X²P(X)]
3	3	3	0.06	0.18	9	0.54
4	4	7	0.14	0.56	16	2.24
5	5	12	0.24	1.2	25	6
6	6	14	0.28	1.68	36	10.08
7	7	10	0.20	1.4	49	9.8
8	8	4	0.08	0.64	64	5.12
9		50	1.00	5.66		33.78

$$V(X) = E(X^2) - [E(X)]^2 = 33.78 - (5.66)^2 = 33.78 - 32.04 = 1.74$$

Una vez conocida el valor de la varianza, podemos determinar el valor de la desviación estándar únicamente obteniendo la raíz cuadrada de la varianza:

$$\sigma = \sqrt{V(X)} = \sqrt{1.74} = 1.32 \text{ veh\u00edculos}$$

En la administraci\u00f3n, se presenta m\u00e1s com\u00fanmente dos tipos de distribuciones de variables aleatorias discretas, siendo estas:

1.- Distribuci\u00f3n binomial: se utiliza para modelar eventos con 2 posibles resultados, como el \u00e9xito o el fracaso, y se puede aplicar en situaciones donde se puede repetir un ensayo o experimento un n\u00famero fijo de veces con una probabilidad constante de \u00e9xito en cada intento. Un ejemplo en la administraci\u00f3n, la distribuci\u00f3n binomial se puede utilizar para modelar la probabilidad de que un nuevo producto tenga \u00e9xito o fracaso en el mercado, as\u00ed como, si se desea conocer la probabilidad de que alg\u00fan producto pueda permanecer en el gusto de los consumidores. En este tipo de distribuci\u00f3n se hace uso de la siguiente f\u00f3rmula:

$$P(X|n, p) = C_x^n p^x q^{n-x}$$

$$P(X|n, p) = \frac{n!}{X! (n - X)!} p^x q^{n-x}$$

Donde:

X = n\u00famero determinado de \u00e9xitos, n = n\u00famero de ensayos, p = probabilidad de \u00e9xito para cada prueba y $q = (1 - p)$

Ejemplos de distribuci\u00f3n binomial

1. La probabilidad de que un prospecto elegido al azar para un agente de ventas realice una compra es 0.20. Si un representante de ventas llama a 6 prospectos determine la probabilidad de que hagan exactamente cuatro ventas.

Para este ejemplo: $n = 6$, $p = 0.20$, $q = 1 - 0.20 = 0.80$ y $x = 4$

$$P(4|6, 0.20) = C_4^6 (0.20)^4 (0.80)^2 \cong 0.015$$

Con el uso de tablas para el c\u00e1lculo de distribuci\u00f3n binomial, se facilita en gran medida su determinaci\u00f3n.

Tabla D.1: DISTRIBUCION BINOMIAL: f(x)												
x	p	0,01	0,05	0,10	0,20	0,25	0,30	0,40	0,50	0,60	0,70	0,75
1	0	0,9900	0,9500	0,9000	0,8000	0,7500	0,7000	0,6000	0,5000	0,4000	0,3000	0,2500
	1	0,0100	0,0500	0,1000	0,2000	0,2500	0,3000	0,4000	0,5000	0,6000	0,7000	0,7500
2	0	0,9801	0,9025	0,8100	0,6400	0,5625	0,4900	0,3600	0,2500	0,1600	0,0900	0,0625
	1	0,0198	0,0950	0,1800	0,3200	0,3750	0,4200	0,4800	0,5000	0,4800	0,4200	0,3750
	2	0,0001	0,0025	0,0100	0,0400	0,0625	0,0900	0,1600	0,2500	0,3600	0,4900	0,5625
3	0	0,9703	0,8574	0,7290	0,5120	0,4219	0,3430	0,2160	0,1250	0,0640	0,0270	0,0156
	1	0,0294	0,1354	0,2430	0,3840	0,4219	0,4410	0,4320	0,3750	0,2880	0,1890	0,1406
	2	0,0003	0,0071	0,0270	0,0960	0,1406	0,1890	0,2880	0,3750	0,4320	0,4410	0,4219
	3	0,0000	0,0001	0,0010	0,0080	0,0156	0,0270	0,0640	0,1250	0,2160	0,3430	0,4219
4	0	0,9606	0,8145	0,6561	0,4096	0,3164	0,2401	0,1296	0,0625	0,0256	0,0081	0,0039
	1	0,0388	0,1715	0,2916	0,4096	0,4219	0,4116	0,3456	0,2500	0,1536	0,0756	0,0469
	2	0,0006	0,0135	0,0486	0,1536	0,2109	0,2646	0,3456	0,3750	0,3456	0,2646	0,2109
	3	0,0000	0,0005	0,0036	0,0256	0,0469	0,0756	0,1536	0,2500	0,3456	0,4116	0,4219
	4		0,0000	0,0001	0,0016	0,0039	0,0081	0,0256	0,0625	0,1296	0,2401	0,3164
5	0	0,9510	0,7738	0,5905	0,3277	0,2373	0,1681	0,0778	0,0313	0,0102	0,0024	0,0010
	1	0,0480	0,2036	0,3281	0,4096	0,3955	0,3602	0,2592	0,1563	0,0768	0,0284	0,0146
	2	0,0010	0,0214	0,0729	0,2048	0,2637	0,3087	0,3456	0,3125	0,2304	0,1323	0,0879
	3	0,0000	0,0011	0,0081	0,0512	0,0879	0,1323	0,2304	0,3125	0,3456	0,3087	0,2637
	4		0,0000	0,0005	0,0064	0,0146	0,0284	0,0768	0,1563	0,2592	0,3602	0,3955
	5			0,0000	0,0003	0,0010	0,0024	0,0102	0,0313	0,0778	0,1681	0,2373
6	0	0,9415	0,7351	0,5314	0,2621	0,1780	0,1176	0,0467	0,0156	0,0041	0,0007	0,0002
	1	0,0571	0,2321	0,3543	0,3932	0,3560	0,3025	0,1866	0,0938	0,0369	0,0102	0,0044
	2	0,0014	0,0305	0,0984	0,2458	0,2966	0,3241	0,3110	0,2344	0,1382	0,0595	0,0330
	3	0,0000	0,0021	0,0146	0,0819	0,1318	0,1852	0,2765	0,3125	0,2765	0,1852	0,1318
	4		0,0001	0,0012	0,0154	0,0330	0,0595	0,1382	0,2344	0,3110	0,3241	0,2966
	5		0,0000	0,0001	0,0015	0,0044	0,0102	0,0369	0,0938	0,1866	0,3025	0,3560
	6			0,0000	0,0001	0,0002	0,0007	0,0041	0,0156	0,0467	0,1176	0,1780

Localizando en la tabla el valor de $n=6$, $x=4$ y $p=0.20$, se desplaza hasta la intersección de los valores de x y p , resultando **0.0154**

Con el uso de Excel, quedaría:

F11 : f_x =D11-E11+DISTR.BINOM.SERIE(B3,C3,A3)

Variable de interés	Núm. De ensayos	Prob. De éxito	Prob. Acumulada
x	n	p	q
4	6	0.2	0.8

Argumentos de función

DISTR.BINOM.SERIE

Ensayos B3 = 6

Probabilidad_s C3 = 0.2

Número_s A3 = 4

Número_s2 = número

= 0.01536

Devuelve la probabilidad de un resultado de prueba que usa una distribución binomial.

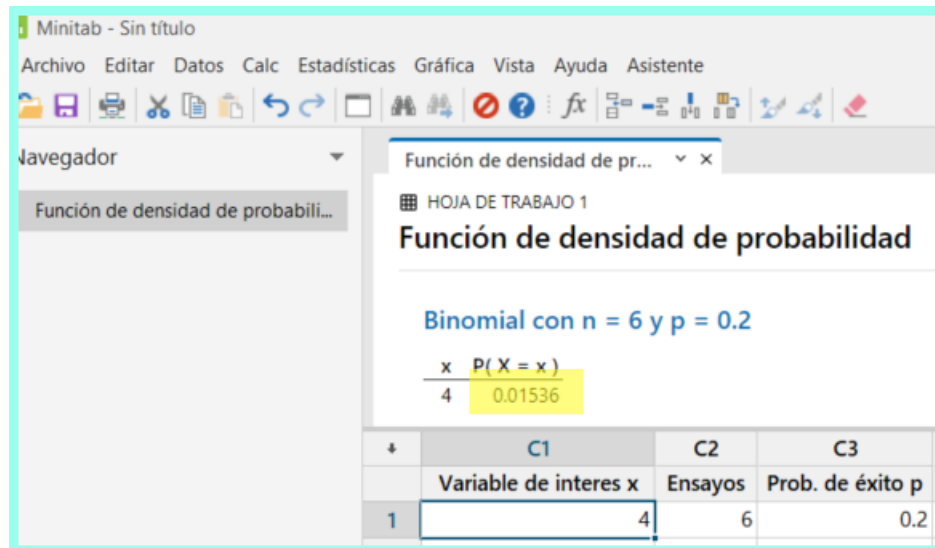
Número_s Es la cantidad de resultados satisfactorios en las pruebas.

Resultado de la fórmula = 0.01696

[Ayuda sobre esta función](#)

Aceptar Cancelar

Con el uso de Minitab, resulta:



Para resolución de los ejercicios podemos apoyarnos con algunas de las herramientas anteriores; fórmula, tablas, Excel o Minitab.

2. Del ejemplo anterior, considerando que se quiere determinar la probabilidad en la cual representante desea saber que se realicen:
 - g) Cuatro o más ventas
 - h) A lo más tres ventas
 - i) Entre dos y tres ventas
 - j) Ninguna venta

Solución:

$$\begin{aligned} \text{a) } P(X \geq 4 | n = 6, p = 0.20) &= P(X = 4) + P(X = 5) + P(X = 6) \\ &= 0.0154 + 0.0015 + 0.0001 \cong \mathbf{0.017} \end{aligned}$$

$$\begin{aligned} \text{b) } P(X \leq 3 | n = 6, p = 0.20) &= p(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) \\ &= 0.2621 + 0.3931 + 0.2458 + 0.0819 \cong \mathbf{0.983} \end{aligned}$$

$$\begin{aligned} \text{c) } P(X = 2, X = 3 | n = 6, p = 0.20) &= P(X = 2) + P(X = 3) \\ &= 0.2458 + 0.0819 \cong \mathbf{0.328} \end{aligned}$$

$$\text{d) } P(X = 0 | n = 6, p = 0.20) = P(X = 0) = \mathbf{0.2621}$$

II.- Distribución de Poisson

Este tipo de distribución se usa para determinar la probabilidad de la ocurrencia de un número determinado de eventos cuando estos ocurren en un continuo espacio o tiempo. Suponiendo que los eventos son independientes y que el proceso es estacionario; para determinar la probabilidad bajo este proceso sólo se requiere

conocer el número medio de eventos a largo plazo en un lapso específico o en una dimensión de espacio que interese; lo cual por lo general se representa por la letra griega lambda (λ) o también por μ . La fórmula para determinar la probabilidad de un número determinado X de éxitos en una distribución de Poisson es:

$$P(X|\lambda) = \frac{\lambda^X e^{-\lambda}}{X!}$$

El valor de λ que se use debe ser aplicable al período de interés.

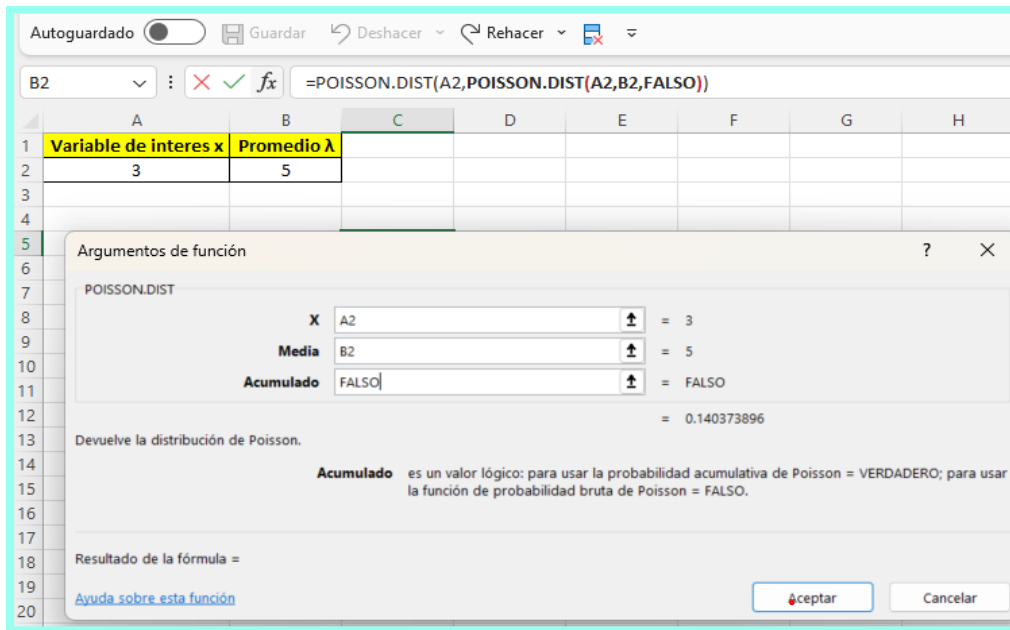
Ejemplo: En el departamento de recursos financieros se recibe en promedio 5 solicitudes de pago por hora. Determinar la probabilidad de que se reciban exactamente 3 solicitudes de pago en 1 hora elegida al azar, se obtiene de la siguiente manera:

$$P(X = 3|\lambda = 5) = \frac{(5)^3 e^{-5}}{3!} \cong \mathbf{0.1404}$$

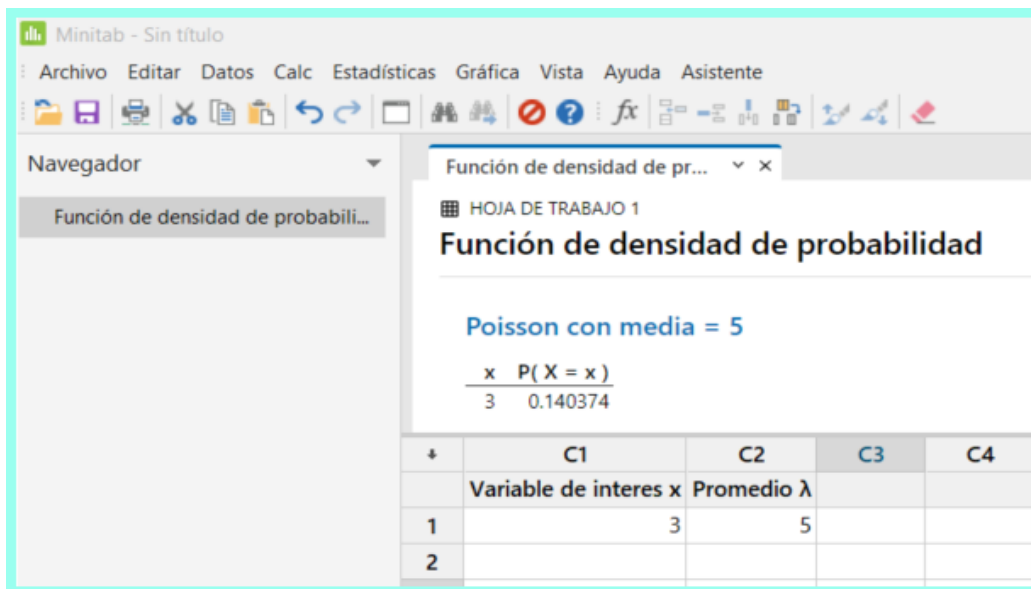
Haciendo uso de las tablas de Poisson, resulta:

x	$\lambda = 4,1$	$\lambda = 4,2$	$\lambda = 4,3$	$\lambda = 4,4$	$\lambda = 4,5$	$\lambda = 4,6$	$\lambda = 4,7$	$\lambda = 4,8$	$\lambda = 4,9$	$\lambda = 5,0$
0	0,0166	0,0150	0,0136	0,0123	0,0111	0,0101	0,0091	0,0082	0,0074	0,0067
1	0,0679	0,0630	0,0583	0,0540	0,0500	0,0462	0,0427	0,0395	0,0365	0,0337
2	0,1393	0,1323	0,1254	0,1188	0,1125	0,1063	0,1005	0,0948	0,0894	0,0842
3	0,1904	0,1852	0,1798	0,1743	0,1687	0,1631	0,1574	0,1517	0,1460	0,1404
4	0,1951	0,1944	0,1933	0,1917	0,1898	0,1875	0,1849	0,1820	0,1789	0,1755
5	0,1600	0,1633	0,1662	0,1687	0,1708	0,1725	0,1738	0,1747	0,1753	0,1755
6	0,1093	0,1143	0,1191	0,1237	0,1281	0,1323	0,1362	0,1398	0,1432	0,1462
7	0,0640	0,0686	0,0732	0,0778	0,0824	0,0869	0,0914	0,0959	0,1002	0,1044
8	0,0328	0,0360	0,0393	0,0428	0,0463	0,0500	0,0537	0,0575	0,0614	0,0653
9	0,0150	0,0168	0,0188	0,0209	0,0232	0,0255	0,0281	0,0307	0,0334	0,0363

Con la herramienta de Excel:



Con el apoyo de Minitab resulta:



Al igual que en la distribución binomial, para determinar la probabilidad en un caso de Poisson se cuenta con cuatro herramientas; fórmula, tablas, Excel y Minitab.

Ejemplos: Al considerar el caso anterior del departamento de Recursos Financieros en la recepción de cinco solicitudes de pago por hora, consideremos los siguientes casos:

Determinar la probabilidad de que se reciban en un intervalo de una hora elegida al azar:

- a) A lo más dos solicitudes de pago
- b) Cuando menos tres solicitudes de pago
- c) Ninguna solicitud de pago

Solución: a) $P(X \leq 2 | \lambda = 5) = P(X = 0) + P(X = 1) + P(X = 2)$
 $0.0067 + 0.0337 + 0.0842 \cong \mathbf{0.125}$

b) $P(X \geq 3 | \lambda = 5) = P(X = 3) + P(X = 4) + \dots + P(X = \infty)$

Podemos observar que esta distribución de Poisson sigue un proceso continuo, por lo tanto, se debe determinar las probabilidades de 0 a 2, sumarlas y posteriormente restárselas a uno.

$$P(X \leq 2 | \lambda = 5) = P(X = 0) + P(X = 1) + P(X = 2)$$
$$0.0067 + 0.0337 + 0.0842 \cong \mathbf{0.125}$$

$$\mathbf{1 - 0.125 = 0.875}$$

c) $P(X = 0 | \lambda = 5) = P(X = 0) \cong \mathbf{0.0067}$

5.2. Distribuciones para variables continuas

En la administración, hay varias distribuciones de variables continuas que se utilizan para modelar diferentes situaciones; algunas de las distribuciones de variables continuas más comunes son las siguientes:

1. Distribución normal: la distribución normal es una de las distribuciones más empleadas en estadística y administración. Aplicándose para modelar datos continuos que presenta una distribución simétrica alrededor de la media o promedio. En administración, se puede aplicar para modelar la distribución de ingresos de una empresa, el tiempo de producción de un producto, o el rendimiento de los empleados en una prueba.
2. Distribución exponencial: esta se utiliza para modelar la cantidad de tiempo entre eventos que ocurren de manera aleatoria independiente en un intervalo de tiempo o espacio determinado. En la administración, se puede utilizar para modelar el tiempo de espera en una fila, el tiempo entre fallas de un equipo, o el tiempo de respuesta de un centro de atención al cliente.
3. Distribución uniforme: Tipo de distribución que se utiliza para modelar datos que presentan una distribución uniforme en un intervalo determinado. Dentro de la

administración, se puede utilizar para modelar la distribución de precios de productos, la distribución de salarios en una empresa, o la distribución de tiempo de respuesta de un sistema.

Estas son solo algunas de las distribuciones de variables continuas que se utilizan en la administración. La elección del tipo de distribución adecuada dependerá de la situación específica y los datos disponibles.

Enfoquémonos a la distribución normal siendo la que más se utiliza en estadística; este tipo de distribución es de vital importancia dentro de la estadística por tres razones principales:

- Muchas variables continuas comunes en el mundo de los negocios presentan distribuciones que se asemejan estrechamente a la distribución normal.
- La distribución normal sirve para acercarse a diversas distribuciones de probabilidad discreta, como la distribución binomial y la distribución de Poisson.
- La distribución normal proporciona la base para la *estadística inferencial clásica* por su relación con el *teorema de límite central*.

La distribución normal tiene importantes propiedades teóricas:

- Tiene una apariencia de forma de campana y por lo tanto es simétrica.
- Sus medidas de tendencia central (media, mediana y moda) son idénticas.
- Su “50% central” es igual a 1.33 desviaciones estándar. Esto significa que el rango intercuartil está contenido dentro de un intervalo de dos tercios de una desviación estándar por debajo de la media y de dos tercios de una desviación estándar por encima de la media.
- Su variable aleatoria asociada tiene un rango infinito ($-\infty < X < \infty$)

La función de densidad de la probabilidad normal se representa por:

$$f(X) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left[\frac{(x-\mu)}{\sigma}\right]^2}$$

Donde: e es la constante matemática $\cong 2.71828$

π es la constante matemática $\cong 3.14159$

μ es la media

σ es la desviación estándar

x es cualquier valor de la variable continua, donde ($-\infty < x < \infty$)

Para no aplicar esta función matemática en el cálculo de la probabilidad para una variable continua dentro de la distribución normal, utilizamos la siguiente fórmula de

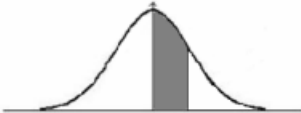
transformación, para convertir cualquier variable aleatoria normal X en una **variable aleatoria normal estandarizada Z** .

Fórmula de transformación: $Z = \frac{X - \mu}{\sigma}$

Convertida la variable aleatoria normal X en una variable aleatoria normal estandarizada Z , se determina las probabilidades deseadas usando la tabla para la **distribución normal estandarizada**.

Distribución Normal

En los ejes están los valores de z y la tabla muestra el área del eje central a la derecha.

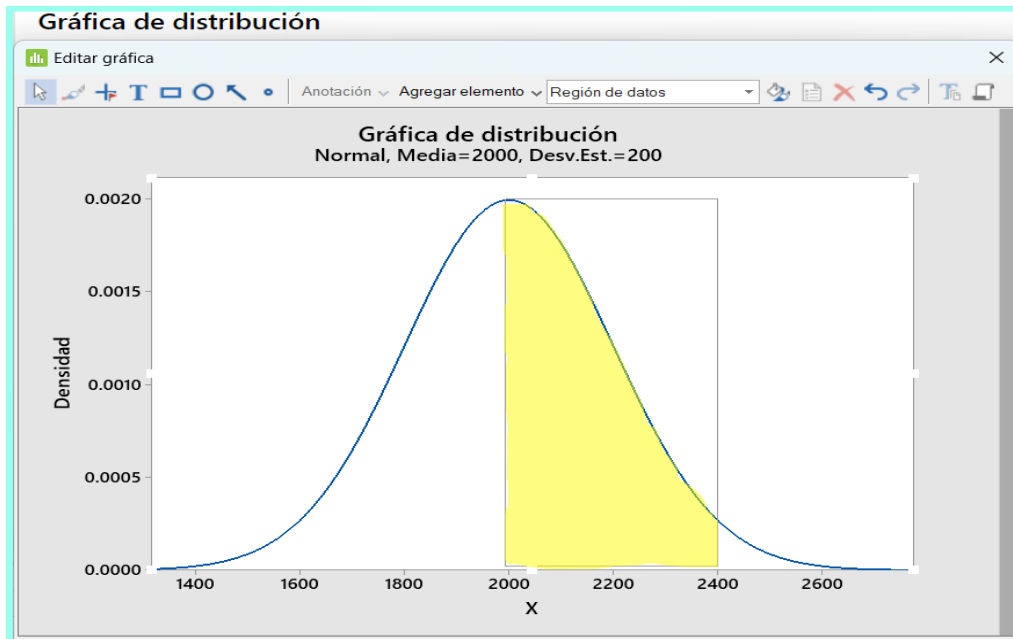


z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545

Un valor normal estándar Z Reformula el valor original X en términos del número de unidades de desviaciones estándar en las que el valor original difiere de la media de la distribución. Un valor negativo de z indicará que el valor X original es menor que el valor de la media ubicándose por lo tanto a la izquierda de la gráfica de distribución normal, debemos de tomar en cuenta a la hora de estandarizar nuestra variable de interés; Para cuando utilizemos la tabla de distribución normal se va a trabajar con valores absolutos.

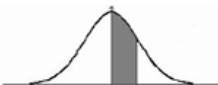
Ejemplo: La vida útil de un componente electrónico para un teléfono celular se ajusta a una distribución normal con media $\mu = 2000$ horas y desviación estándar $\sigma = 200$ horas. Para determinar la probabilidad de que un componente tomado al azar dure entre 2000 y 2400 horas se determina de la siguiente manera.

$$Z = \frac{X - \mu}{\sigma} = \frac{2400 - 2000}{200} = \frac{400}{200} = +2.0$$



El valor $Z = +2.0$ indica que 2400 horas se encuentra a 2 desviaciones estándar sobre la media que es 2000 horas aplicando la tabla de distribución normal ubicamos este valor de X estandarizada.

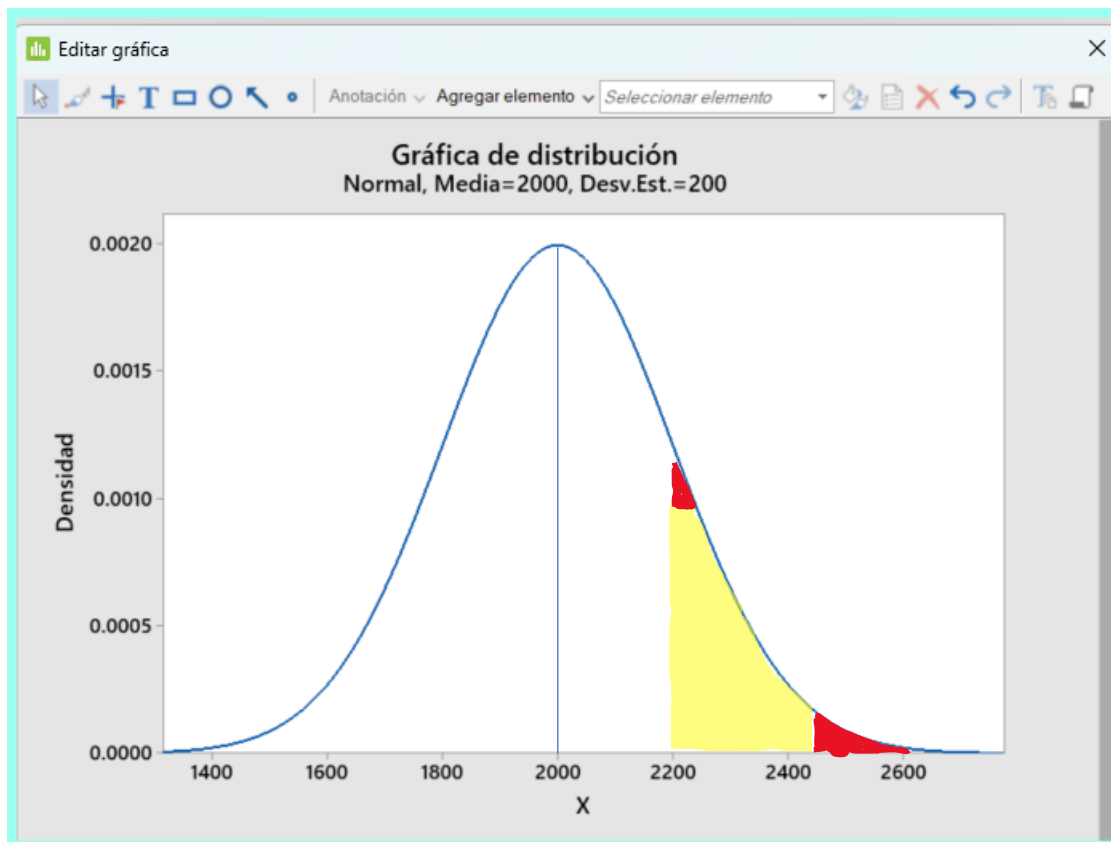
Distribución Normal
En los ejes están los valores de z y la tabla muestra el área del eje central a la derecha.



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817

Por lo tanto, la probabilidad de que un componente tomado al azar dure entre 2000 y 2400 horas es del **47.72%** $P(0 \leq Z \leq +2.0) = 0.4772$ o $P(2000 \leq X \leq 2400) = 0.4772$

Considerando el ejemplo anterior del componente electrónico, suponga que se está interesado en la probabilidad de que un componente tomado al azar dure más de 2200 horas, se procede de la siguiente manera:



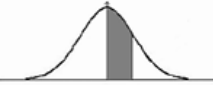
Estamos interesados en el área sombreada bajo la curva de distribución normal, lo que procede a continuación es estandarizar el valor de $X=2200$ Aplicando la fórmula de Z .

$$Z = \frac{2200 - 2000}{200} = +1.0$$

El valor resultante de z lo ubicamos en la tabla de distribución normal,

Distribución Normal

En los ejes están los valores de z y la tabla muestra el área del eje central a la derecha.



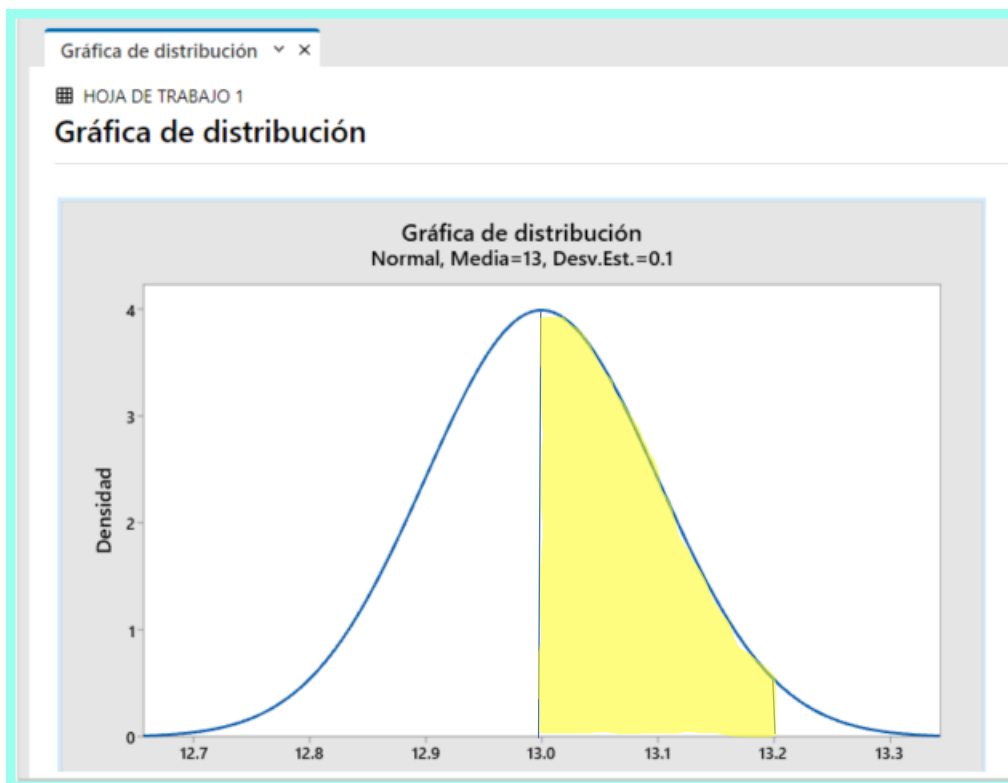
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015

$$P(0 \leq z \leq +1.0) = 0.3413 \text{ de acuerdo a la tabla}$$

$$P(z > +1.0) = 0.5000 - 0.3413 = \mathbf{0.1587 \text{ o } 15.87\%}$$

Ejercicios:

1.- una empresa de dulces para piñatas se ha ajustado el proceso de empaque de modo que en cada paquete se coloquen en promedio $\mu=13.0$ onzas de dulce. Debido a las fuentes aleatorias de variabilidad no todos los paquetes tienen en promedio 13.0 oz; por lo tanto, presentan una desviación estándar del peso neto verdadero de $\sigma=0.1$ oz. Se conoce que la distribución de pesos san justa a una distribución de probabilidad normal. Determine la probabilidad de que un paquete elegido al azar contenga entre 13.0 y 13.2 oz de dulces.



$$Z = \frac{X - \mu}{\sigma} = \frac{13.2 - 13.0}{0.1} = +2.0$$

$$P(13.0 \leq X \leq 13.2) = P(0 \leq Z \leq +2.0) = \mathbf{0.4772}$$

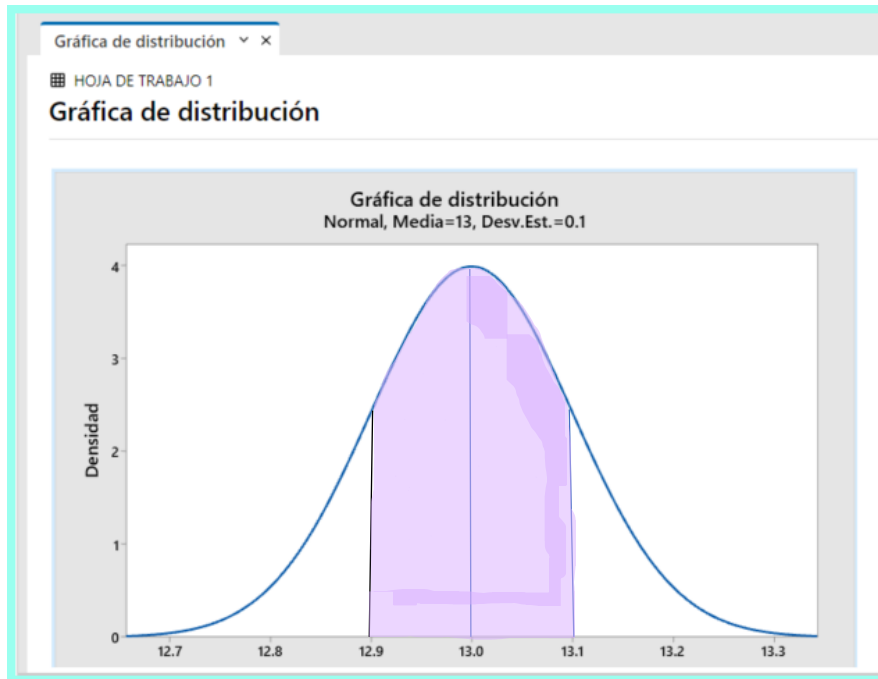
Utilizando la tabla de distribución normal.

2.- Del ejercicio anterior cuál es la probabilidad de que el peso de la bolsa de dulces se localice entre 12.9 oz y 13.1 oz

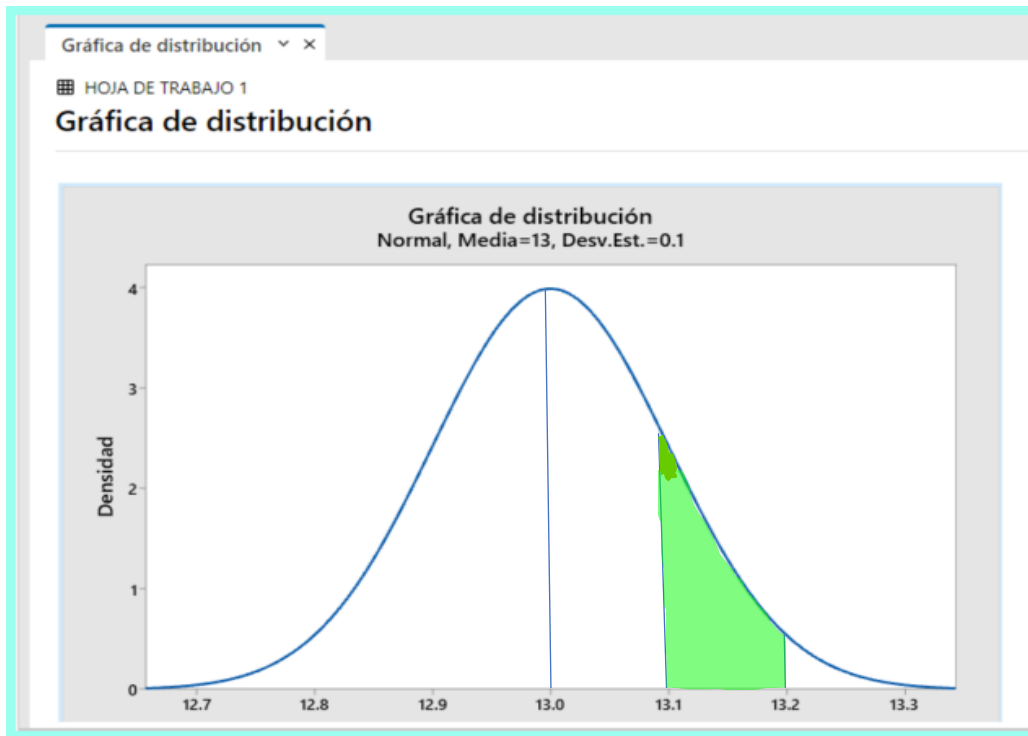
$$Z_1 = \frac{X_1 - \mu}{\sigma} = \frac{12.9 - 13.0}{0.1} = -1.0$$

$$Z_2 = \frac{X_2 - \mu}{\sigma} = \frac{13.1 - 13.0}{0.1} = +1.0$$

$$P(12.9 \leq X \leq 13.1) = P(-1.0 \leq Z \leq +1.0) = 0.3413 + 0.3413 = \mathbf{0.6826}$$



3.- regresando al ejercicio número uno considere que se desea conocer la probabilidad de que un paquete de dulces contenga entre 13.1 oz y 13.2 oz. La gráfica de distribución normal quedaría de la siguiente manera,



$$Z_1 = \frac{X_1 - \mu}{\sigma} = \frac{13.1 - 13.0}{0.1} = +1.0$$

$$Z_2 = \frac{X_2 - \mu}{\sigma} = \frac{13.2 - 13.0}{0.1} = +2.0$$

$$P(13.1 \leq X \leq 13.2) = P(+1.0 \leq Z \leq 2.0) = 0.4772 - 0.3413 = \mathbf{0.1359}$$

Ejercicios:

1. Un vendedor de seguros vende pólizas a 5 hombres, todos de la misma edad y con buena salud. De acuerdo con las tablas actuariales, la probabilidad de que un hombre de esta edad esté vivo en 30 años es $\frac{2}{3}$. Encontrar la probabilidad de que en 30 años estén vivos;
 - a) los 5 hombres,
 - b) por lo menos 3 de estos hombres,
 - c) sólo 2 de estos hombres y
 - d) por lo menos uno de ellos.

Resultados:

a) $\frac{32}{243} = 0.132$

b) $\frac{192}{243} = 0.790$

c) $\frac{40}{243} = 0.165$

d) $\frac{242}{243} = 0.996$

2. Suponga que 40% de los empleados de una empresa están a favor de sindicalizarse, y que se toma una muestra aleatoria de 10 empleados y se les pide una respuesta anónima. ¿Cuál es la probabilidad de que:

a) la mayoría de los interrogados y

b) menos de la mitad de los interrogados estén a favor de la sindicalización?

Resultados:

a) 0.1663

b) 0.6330

3. En una tienda de departamentos, seis personas por hora en promedio usan un cajero automático durante las horas pico. ¿Cuál es la probabilidad de que:

a) exactamente seis personas usen el cajero durante una hora tomada al azar?

b) menos de cinco personas usen el cajero durante una hora tomada al azar?

c) nadie use el cajero durante un intervalo de 10 minutos? Y

d) nadie use el cajero durante un intervalo de 5 minutos?

Resultados:

a) 0.1606

b) 0.2851

c) 0.3679

d) 0.6065

4. Suponga que el manuscrito de un libro de texto tiene en total 50 errores tipográficos en sus 500 páginas, y que estos errores están distribuidos en forma aleatoria a lo largo del libro. ¿Cuál es la probabilidad de que

a) un capítulo de 30 páginas tenga dos o más errores?

b) un capítulo de 50 páginas tenga dos o más errores?

c) una página seleccionada al azar no tenga ningún error?

Resultados:

- a) 0.8008
- b) 0.9596
- c) 0.9048

5. Se ha encontrado que la vida útil de cierta marca de llantas de alto rendimiento se ajusta a una distribución normal con $\mu = 38000$ millas y de $\sigma = 3000$ millas. a) ¿cuál es la probabilidad de que una llanta elegida en forma aleatoria tenga una vida útil de por lo menos 35000 millas? b) ¿cuál es la probabilidad de que dure más de 45000 millas?

Respuestas:

- a) 0.8413
 - b) 0.0099
5. Se ha encontrado que el tiempo que requiere una persona en la ventanilla de un cajero de banco tiene una distribución aproximadamente normal con media igual $\mu=130$ segundos y una $\sigma = 45$ segundos. ¿Cuál es la probabilidad de que una persona elegida en forma aleatoria a) requiera menos de 100 segundos para realizar una transacción? b) demore entre 2 y 3 minutos en la ventanilla del cajero?

Respuestas:

- a) 0.2514
- b) 0.4536

Unidad 6



Muestreo y estimación aplicado al control estadístico de procesos

6.1 Fundamentos teóricos del muestreo y estimación

Con este tema estaremos entrando al campo de la estadística inferencial, debido a factores tales como el tiempo, y el costo; los parámetros de una población o de un proceso suelen estimarse basándose en estadísticos muestrales. Un *parámetro* es un valor representativo de una población, mientras que un *estadístico muestral* es un valor representativo de una muestra. Para poder usar un estadístico muestral como estimador de un parámetro la muestra debe ser una *muestra aleatoria*, la cual es obtenida del total de conjuntos o elementos a analizar o a estudiar.

Una *distribución de muestreo* se refiere la distribución de los distintos valores que un estadístico muestral, o estimador tomaría en muchas muestras distintas, todas del mismo tamaño. Aun cuando por lo general o sea tendrá sólo una muestra aleatoria o un subgrupo racional, se debe aceptar que el estadístico muestral particular que se determina, como puede ser la media o la mediana muestral, no es exactamente igual al parámetro poblacional respectivo; además, *el valor de un estadístico muestral varía de una muestra a otra* debido a la variabilidad en el muestreo aleatorio o *error de muestreo*.

Estimadores puntuales que se usan con frecuencia

Parámetro poblacional	Estimador
Media, μ	$\bar{X}$
Diferencia entre las medias de 2 poblaciones, $\mu_1 - \mu_2$	$\bar{X}_1 - \bar{X}_2$
Proporción, π	$\hat{P}$
Diferencia entre las proporciones de 2 poblaciones, $\pi_1 - \pi_2$	$\hat{P}_1 - \hat{P}_2$
Varianza, σ^2	s^2
Desviación estándar, σ	s

6.2 Distribución del muestreo, características y aplicación en el área administrativa

La distribución de muestreo se refiere a la distribución de las estadísticas calculadas a partir de múltiples muestras las cuales fueron tomadas de la misma población. Esta distribución proporciona información sobre la variabilidad de las estadísticas muestrales las cuales nos ayuda a determinar inferencias sobre los parámetros poblacionales.

Características de la distribución del muestreo:

- 1.- Media: la media de la distribución del muestreo suele ser igual al valor del parámetro poblacional; es decir, si se toman múltiples muestras de una población y se determina la media de cada muestra, en promedio, las medias de las muestras deberían ser iguales a la media poblacional.
- 2.- Forma: la forma de la distribución del muestreo depende de la forma de la población y del tamaño de la muestra. Si la población sigue una distribución normal y el tamaño de la muestra es lo suficientemente grande, (según el Teorema del Límite Central), la distribución del muestreo de la media será aproximadamente normal, independientemente de la forma de la población.
- 3.- Varianza: la varianza de la distribución del muestreo es inversamente proporcional al tamaño de la muestra. A medida que el tamaño de la muestra se incrementa, la distribución del muestreo se vuelve más concentrada alrededor del valor del parámetro poblacional.

Aplicación en el área administrativa:

Dentro del área administrativa, la distribución del muestreo se utiliza para realizar inferencias sobre los parámetros poblacionales y tomar decisiones basadas en datos. Algunas aplicaciones comunes dentro de esta área resultan:

- 1.- Estimación de parámetros: la distribución del muestreo se utiliza para construir intervalos de confianza para estimar parámetros poblacionales, como la media, la proporción o la desviación estándar. Estos intervalos proporcionan un rango dentro del cual se espera que se encuentre el valor real o verdadero del parámetro con cierto nivel de confianza.
- 2.- Pruebas de hipótesis: la distribución de muestreo también se utiliza para realizar comparaciones estadísticas muestrales con los valores esperados bajo una hipótesis nula para evaluar si hay evidencia suficiente para rechazar o no la hipótesis.
- 3.- Toma de decisiones: la distribución del muestreo ayuda a los administradores a tomar decisiones basadas en datos. Por ejemplo, al evaluar el rendimiento de un proyecto o de una inversión se puede calcular una estimación puntual y un intervalo de confianza para la medida de rendimiento relevante y tomar decisiones fundamentadas en esta información y con los resultados esperados.

Fórmulas que emplear para estos temas:

valor esperado $E(\bar{X}) = \mu$

desviación estándar de la distribución de las medias muestrales

o error estándar de la media $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

factor de corrección finita, omitir cuando $n < 0.05N$, $\sigma_{\bar{x}} = \sqrt{\frac{N-n}{N-1}}$

si la dist'n. de muestreo para la media es una dist'n. normal $Z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}}$

Intervalos de confianza para la media empleando la distribución normal:

Cuando los parámetros poblacionales como la media y la desviación estándar poblacional hoy son conocidos hoy se emplea un *razonamiento deductivo* para el resultado muestral, cuando hacemos uso de los datos muestrales para hacer afirmaciones acerca del valor de la media poblacional estaremos empleando un *razonamiento inductivo*.

cuando se conoce la σ poblacional $\bar{X} \pm z\sigma_{\bar{x}}$

al no conocer la σ poblacional $\bar{X} \pm z s_{\bar{x}}$

Los valores $\pm z\sigma_{\bar{x}}$ o $\pm z s_{\bar{x}}$ son el margen de error del intervalo de confianza

Los intervalos de confianza que más se usan son los intervalos de 90%, 95% y 99%.

Proporciones de área bajo la curva normal

Z (número de unidades de desviación estándar a partir de la media)	Proporción del área en el intervalo $\mu \pm z\sigma$
1.645	0.90
1.96	0.95
2.58	0.99

Uso de la distribución t y los intervalos de confianza para la media:

$$\bar{X} \pm t_{gl} s_{\bar{x}}$$

$$gl = n - 1$$

Tabla resumen de la estimación de un intervalo para la media poblacional

Población	Tamaño de la muestra	σ conocida	σ no conocida
Con distribución normal	Grande ($n \geq 30$)	$\bar{x} \pm z\sigma_{\bar{x}}$	$\bar{x} \pm t_{s_{\bar{x}}} o \bar{x} \pm z s_{\bar{x}} **$
	pequeña ($n < 30$)	$\bar{x} \pm z\sigma_{\bar{x}}$	$\bar{x} \pm t_{s_{\bar{x}}}$
Con una distribución diferente de la normal	Grande ($n \geq 30$)	$\bar{x} \pm z\sigma_{\bar{x}} *$	$\bar{x} \pm t_{s_{\bar{x}}} o \bar{x} \pm z s_{\bar{x}} ***$
	pequeña ($n < 30$)	Generalmente se emplea un procedimiento no paramétrico para determinar la mediana.	

*Se invoca el teorema del límite central

**Se usa z como aproximación a t

***Se invoca el teorema del límite central y se usa z como aproximación a t

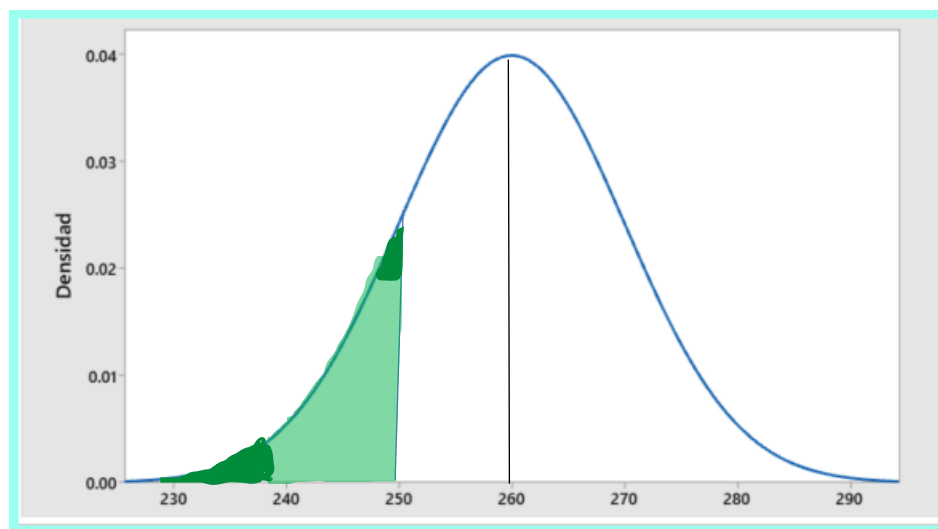
Ejemplos:

1.- Determinar el error estándar de la distribución de muestreo para la media, en el caso donde un auditor toma una muestra aleatoria de tamaño $n = 16$ de un conjunto de $N = 1500$ cuentas por cobrar y se desconoce la desviación estándar de los montos en todas las cuentas por cobrar.

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{57.00}{\sqrt{16}} = \$14.25$$

2.- Un auditor extrae una muestra aleatoria de tamaño $n=36$ de una población de 1000 cuentas por cobrar. Se conoce que la media poblacional es de $\mu = \$260.00$ y $\sigma = \$45.00$.

¿Cuál es la probabilidad de que la media muestral sea menor que \$250?



$$E(\bar{X}) = \mu = \$260$$

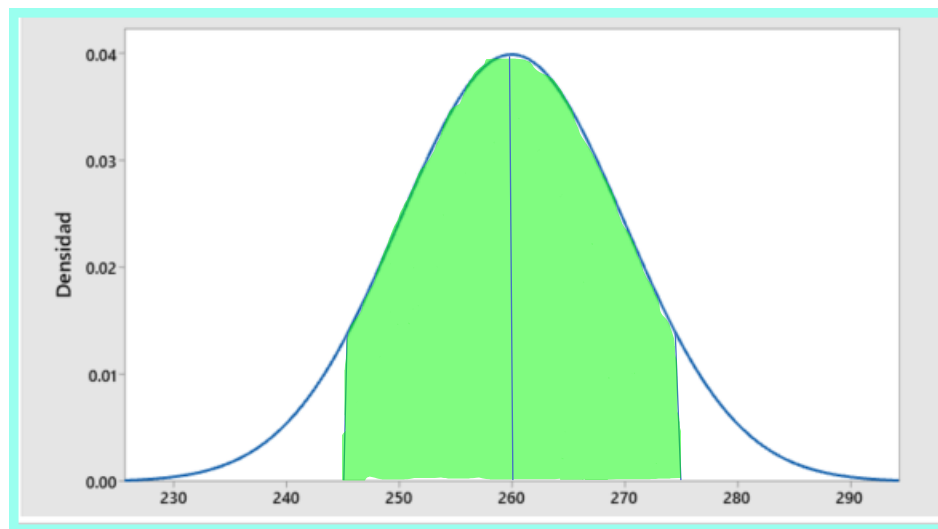
$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{\$45.00}{\sqrt{36}} = \$7.50$$

$$Z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{\$250 - \$260}{\$7.50} = -1.33$$

$$P(\bar{X} < 250.00 | \mu = 260.00, \sigma_{\bar{x}} = 7.50) = P(z < -1.33)$$

$$P(z < 1.33) = 0.5000 - P(-1.33 \leq z \leq 0) \\ = 0.5000 - 0.4082 = \mathbf{0.0918}$$

3.- Del ejemplo anterior, ¿cuál es la probabilidad de que la media muestral fluctúe a \$15.00 de la media poblacional?



$$P(245 \leq \bar{X} \leq 275 | \mu = 260, \sigma_{\bar{x}} = 7.50)$$

$$z_1 = \frac{245 - 260}{7.50} = -2.00$$

$$z_2 = \frac{275 - 260}{7.50} = +2.00$$

$$P(-2.00 \leq z \leq +2.00) = 0.4772 + 0.4772 \cong \mathbf{95\%}$$

4.- Estime el salario medio del total de trabajadores de una empresa con un 95% de confianza de que el intervalo comprenda el valor medio poblacional, para una muestra aleatoria de 30 empleados por hora, donde se conoce que la media muestral es de $\bar{X} = \$180.00$ y $s = \$14.00$.

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{14.00}{\sqrt{30}} = 2.56$$

$$\bar{X} \pm 1.96s_{\bar{x}} = 180.00 \pm 1.96(2.56) = \text{\$174.98 a \$185.02}$$

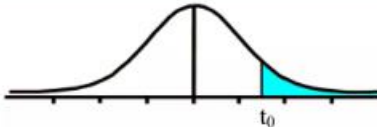
con un nivel de confianza en este estimador de 95% podemos afirmar que el salario medio de todos los trabajadores está entre \$174.98 y \$185.02

5.- Se conoce la vida promedio de operación en una muestra aleatoria de $n= 10$ tablillas electrónicas es $\bar{X} = 4000 h$ y $s = 200 h$. Se supone que la vida de operación de las tablillas electrónicas presenta una distribución aproximadamente normal. Mediante un intervalo de 95% de confianza estime la vida media de operación de la población de las tablillas electrónicas de la que se extrajo la muestra.

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{200}{10} = 63.3$$

$$t_{gl} = t_{n-1} = t_9 = 2.262$$

Tabla t-Student



Grados de libertad

	0.25	0.1	0.05	0.025	0.01	0.005
1	1.0000	3.0777	6.3137	12.7062	31.8210	63.6559
2	0.8165	1.8856	2.9200	4.3027	6.9645	9.9250
3	0.7649	1.6377	2.3534	3.1824	4.5407	5.8408
4	0.7407	1.5332	2.1318	2.7765	3.7469	4.6041
5	0.7267	1.4759	2.0150	2.5706	3.3649	4.0321
6	0.7176	1.4398	1.9432	2.4469	3.1427	3.7074
7	0.7111	1.4149	1.8946	2.3646	2.9979	3.4995
8	0.7064	1.3968	1.8595	2.3060	2.8965	3.3554
9	0.7027	1.3830	1.8331	2.2622	2.8214	3.2498
10	0.6998	1.3722	1.8125	2.2281	2.7638	3.1693
11	0.6974	1.3634	1.7959	2.2010	2.7181	3.1058
12	0.6955	1.3562	1.7823	2.1788	2.6810	3.0545

$$Int. de 95\% = \bar{X} \pm t_{gl}s_{\bar{x}}$$

$$= 4000 \pm (2.262)(63.3)$$

$$\cong \text{3857 a 4143 h}$$

6.3. Teorema del límite central

El Teorema del Límite Central es de vital importancia en estadística y establece que, bajo ciertas condiciones, la distribución de la media muestral de una muestra aleatoria tiende a aproximarse a una distribución normal a medida que el tamaño de la muestra se incrementa, independientemente de la forma de la distribución de la población original.

El Teorema del Límite Central se basa en los siguientes supuestos:

- 1.- Independencia: Los elementos de la muestra deben ser seleccionados de forma independiente y no estar correlacionados entre sí.
- 2.- Tamaño de la muestra: el tamaño de la muestra debe ser suficientemente grande; si bien no hay un tamaño de muestra mínimo definido, se considera que una muestra con un tamaño de al menos 30 elementos es suficiente para que el teorema se aplique de manera aproximada.
- 3.- Suma o media: El Teorema del Límite Central se aplica tanto a la suma de variables aleatorias como a la media de variables aleatorias. En el caso de la media, el teorema es válido incluso si las variables no tienen una distribución normal.

Las implicaciones del teorema del límite central son significativas. Ya que, aunque la población original no siga una distribución normal, la distribución de las medias muestrales se aproxima cada vez más a una distribución normal a medida que el tamaño de la muestra se incrementa. Esto es muy importante porque la distribución normal es de las más aplicadas en cualquier área de investigación o del conocimiento. Lo cual permite realizar inferencias y cálculos estadísticos más confiables y precisos para la hora de tomar una decisión.

El teorema del límite central tiene muchas aplicaciones prácticas en estadística e investigación. Por ejemplo, se utiliza para estimar parámetros poblacionales, construir intervalos de confianza, realizar pruebas de hipótesis y simular distribuciones muestrales. Es una herramienta fundamental en el análisis estadístico contribuyendo en gran medida al desarrollo de la teoría y las aplicaciones estadísticas.

6.4. Tipos de estimación y características

En el campo de la estadística, existen varios tipos de estimación y cada uno de ellos con sus propias características y aplicaciones dependiendo del tipo de estudio o análisis que se quieran llevar a cabo a continuación definiremos algunos de ellos:

- 1.- Estimación puntual: consiste en obtener un valor único como estimación del parámetro desconocido de interés. Por ejemplo, la media muestral como estimación de la media poblacional. La estimación puntual proporciona un valor

específico, pero no ofrece información sobre la precisión o incertidumbre asociada a la estimación.

- 2.- Intervalo de confianza: en lugar de proporcionar una única estimación, un intervalo de confianza establece un rango de valores dentro del cual se espera que se encuentre el parámetro fue un cierto nivel de confianza. Estos niveles de confianza más utilizados son de él 90%, 95% y 99% para la media poblacional. Este tipo de estimación toma en cuenta la incertidumbre asociada a la estimación y proporciona una buena medida de precisión.
- 3.- Estimación por intervalo de producción: similar al intervalo de confianza, pero se utiliza para predecir el rango de valores en el cual se espera que se encuentre una nueva observación. Resulta muy práctico en problemas de pronóstico y predicción.
- 4.- Estimación bayesiana: se basa en la teoría de la probabilidad bayesiana y utiliza información previa y datos observados para obtener una distribución posterior del parámetro de interés. Este tipo de estimador proporciona una forma de actualizar las creencias previas a medida que se obtienen nuevos datos.
- 5.- Estimación por máxima verosimilitud: consiste en encontrar el valor del parámetro que maximiza la verosimilitud de los datos observados. Es ampliamente utilizado y presenta buenas propiedades asintóticas.
- 6.- Estimación por mínimos cuadrados: utilizado en análisis de regresión, buscan minimizar la suma de los cuadrados de las diferencias entre los valores observados y los predichos por un modelo.

Características de un buen estimador:

- Precisión: la estimación debe ser lo más cercana posible al valor real del parámetro que se quiera estimar.
- Sesgo: la estimación debe ser lo más imparcial posible, es decir, no debe estar sistemáticamente por encima o por debajo del valor real del parámetro.
- Eficiencia: la estimación debe tener la menor varianza posible, es decir para sentar en gran medida lo más consistente posible.
- Robustez: la estimación debe ser resistente a datos atípicos o errores de medición.
- Consistencia: a medida que se aumente el tamaño de la muestra, la estimación debe converger al valor real del parámetro.

6.5. Determinación del tamaño de la muestra

Para determinar el tamaño de la muestra, va a depender de varios factores, incluyendo el objetivo del estudio, el nivel de precisión deseado, el tipo de variable que se está

estudiando y la metodología estadística que se utilizará. A continuación, se enumeran algunos pasos para determinar el tamaño de una muestra:

- 1.- Definir el objetivo del estudio: es importante tener Claro qué es lo que se quiere investigar y cuál es la pregunta o hipótesis que se pretende responder. Esto nos ayuda a determinar el tipo de análisis estadístico que se utilizará y la precisión requerida.
- 2.- Estimar el nivel de confianza y el margen de error: el nivel de confianza se refiere a la probabilidad de que el intervalo de confianza contenga el valor del parámetro. El margen de error indica la precisión deseada para la estimación. Estos valores suelen establecerse de antemano en función de las necesidades del estudio y la confiabilidad requerida.
- 3.- Estimar la variabilidad de la población: si se tiene información previa de la variabilidad de la población, esto se puede utilizar para estimar el tamaño de la muestra. En caso contrario, conservadoramente se puede emplear una estimación o realizar un estudio piloto para obtener una estimación inicial de la variabilidad.
- 4.- Seleccionar el método de cálculo de tamaño de la muestra: existen diferentes métodos, como son; el método de fórmulas analíticas, el método de muestreo aleatorio simple o el método de simulación. El método adecuado va a depender del tipo de estudio y las suposiciones que se realicen estadísticamente.
- 5.- Realizar los cálculos: esto implica ingresar los valores correspondientes a nivel de confianza, al margen de error, la variabilidad estimada así como otros parámetros de gran relevancia.

Estos pasos anteriores no son exclusivos para estimar el tamaño de la muestra, ya que en la práctica se pueden ajustar a medida que se obtienen los datos reales y se evalúa la precisión y la representatividad de la muestra. Además, el tamaño de la muestra también puede depender de limitaciones prácticas, como recursos y tiempo disponibles para la recolección de datos.

Al suponer que se especifica el tamaño del intervalo de confianza y el nivel de confianza deseado. Sí se conoce hoy la desviación estándar poblacional o se puede estimar, la fórmula para determinar el tamaño de la muestra en el uso de la distribución normal es:

$$n = \left(\frac{z\sigma}{E} \right)^2$$

Ejemplo: El área de recursos humanos de una empresa desea estimar con un margen de tolerancia de 3 horas y con 90% de confianza el número promedio de horas por semestre de capacitación hoy a los supervisores de una sección de la empresa. Se estima que la desviación estándar de las horas de capacitación en otras áreas es de 20 horas. Determine el tamaño de muestra mínimo que se requiere.

$$n = \left(\frac{z\sigma}{E} \right)^2 = \left(\frac{1.645 * 20}{3} \right)^2 = 120.27 \cong 121$$

Para este tipo de ejercicios se recomienda que el resultado se aproxime al entero inmediato superior, y cuando el resultado nos arroja un valor inferior a 30, debemos asumir como valor mínimo 30 elementos para la muestra, dado que las operaciones que se desarrolla están basadas en aproximaciones a una distribución normal.

6.6. Intervalos de confianza aplicados al control estadístico de procesos

En el control estadístico de procesos (CEP), los intervalos de confianza se utilizan para estimar los límites de control y monitorear la variabilidad del proceso. Los intervalos de confianza más comunes aplicados en el CEP son:

- 1.- Intervalo de confianza para la media: se utiliza para estimar el rango dentro del cual se espera que esté la media de una característica del proceso. Se calcula utilizando la media muestral, la desviación estándar muestral y el tamaño de la muestra. Los límites de control se establecen utilizando un nivel de confianza y se utilizan para identificar cuando el proceso está fuera de los límites de control.
- 2.- Intervalo de confianza para la variabilidad: se utiliza para estimar el rango dentro del cual se espera que esté la variabilidad del proceso. Se calcula utilizando el rango muestral, la desviación estándar muestral o el coeficiente de variación muestral. Se establecen los límites de control para identificar de forma adecuada las variaciones inusuales en la variabilidad del proceso.
- 3.- Intervalo de confianza para la proporción: en algunos casos, el interés se centra en estimar la proporción de unidades defectuosas en un proceso. En este caso, se calcula un intervalo de confianza para la proporción, empleando la proporción muestral y el tamaño de la muestra. Este intervalo proporciona una estimación de la variabilidad en la proporción y es utilizado para medir el número de unidades defectuosas dentro de los límites aceptables.

Es importante considerar que los intervalos de confianza en el CEP están relacionados con los límites de control utilizados para poder detectar variaciones inusuales en el proceso. Estos límites se establecen en función de la variabilidad esperada del proceso y el nivel de confianza deseado. Los intervalos de confianza proporcionan una forma de estimar dicha variabilidad lo cual nos permite determinar si el proceso se encuentra dentro de control estadístico.

6.7. Gráficas de control y tipos de variaciones en los procesos

Las gráficas de control son herramientas indispensables y de gran relevancia utilizadas en el control estadístico de los procesos, (CEP) Para poder monitorear la variabilidad del

proceso y así detectar cualquier desviación de este. Las más comunes son las siguientes:

- 1.- Gráfica de control para variables: esta gráfica se utiliza cuando se mide una característica de calidad del proceso que se pueden medir en una escala numérica continua, como la longitud, el peso, la temperatura, entre otras. En este tipo de gráfica se grafican los valores de las muestras tomadas a lo largo del tiempo y se establecen límites de control para detectar desviaciones que son significativas de la media del proceso.
- 2.- Gráfica de control para atributos: esta gráfica se utiliza cuando se mide una característica de calidad del proceso que se puede clasificar en categorías discretas, como defectuoso o no defectuoso, aceptable o no aceptable, entre otras. En esta gráfica se grafica la proporción de unidades defectuosas o no conformes en cada muestra tomada a lo largo del tiempo y se establecen límites de control para detectar desviaciones significativas de la proporción media de unidades defectuosas o no conformes en el proceso.

Los tipos de variaciones en los procesos se pueden clasificar en dos categorías: variaciones comunes y variaciones especiales. Siendo las variaciones comunes aquellas que están presentes en el proceso de manera constante y predecible, y pueden ser atribuidas a la variabilidad inherente del proceso. Por otra parte, las variaciones especiales, son aquellas que ocurren de manera impredecible y que no son inherentes al proceso. Las cuales pueden ser causadas por factores externos, como cambios en las condiciones ambientales, errores en la medición, entre otros.

Es importante que las gráficas de control se utilizan para detectar desviaciones significativas de las variaciones comunes y no de las variaciones especiales, ya que las primeras están presentes en el proceso y no necesariamente indican que el proceso esté fuera de control. Al detectar una desviación significativa de las variaciones comunes, se pueden asumir medidas para corregir y mejorar el proceso de la calidad del producto o servicio.

Cuando *media y desviación estándar del proceso son conocidas*.

$$\text{Límites de control} = \mu \pm 3 \frac{\sigma}{\sqrt{n}}$$

Cuando *media y desviación estándar del proceso son desconocidas*.

$$\bar{\bar{X}} = \frac{\sum \bar{X}}{k}$$

$$\bar{s} = \frac{\sum s}{k}$$

Ejemplo: La siguiente tabla presenta los pesos en onzas en una sucesión de 15 muestras de subgrupo racional de papas fritas, con $n=4$ en cada muestra. también se conocen las medias, las desviaciones estándar. Suponga que las especificaciones de empaque requieren un peso promedio por paquete de media igual $\mu = 15oz$ y una desviación estándar de $\sigma = 0.1oz$. Use estos valores paramétricos para determinar la línea central y los límites de control inferior y superior para la gráfica $\bar{X}$.

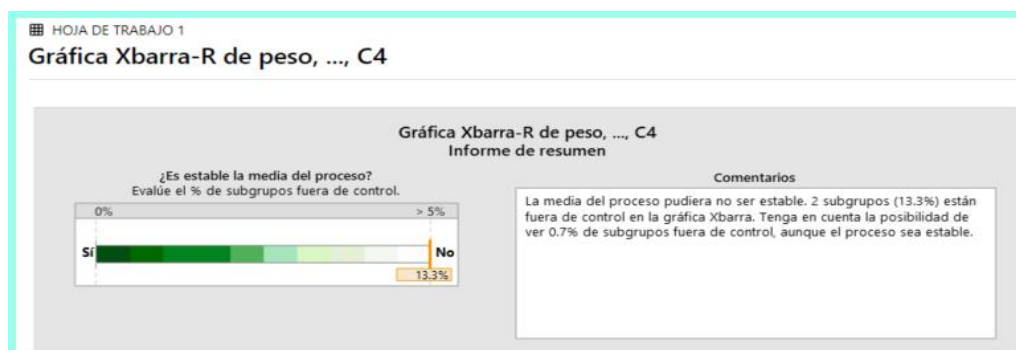
A	B	C	D	E	F	G
Muestra	Peso de los paquetes (onzas)				x	s
1	15.01	14.98	15.16	14.8	14.99	0.148
2	15.09	15.14	15.08	15.03	15.08	0.045
3	15.04	15.1	14.93	15.13	15.05	0.088
4	14.9	15.03	14.94	14.92	14.95	0.057
5	15.04	15.05	15.08	14.98	15.04	0.042
6	14.96	14.81	14.96	14.91	14.91	0.071
7	15.01	15.1	14.9	15.03	15.01	0.083
8	14.71	14.92	14.77	14.95	14.84	0.116
9	14.81	14.8	14.64	14.95	14.8	0.127
10	15.03	14.89	14.99	15.03	14.98	0.066
11	15.16	14.91	14.95	14.83	14.96	0.141
12	14.92	15.05	15.01	15.02	15	0.056
13	15.06	15.03	14.95	15.02	15.02	0.047
14	14.99	15.14	15.04	15.11	15.07	0.068
15	14.94	15.08	14.9	15.17	15.02	0.125

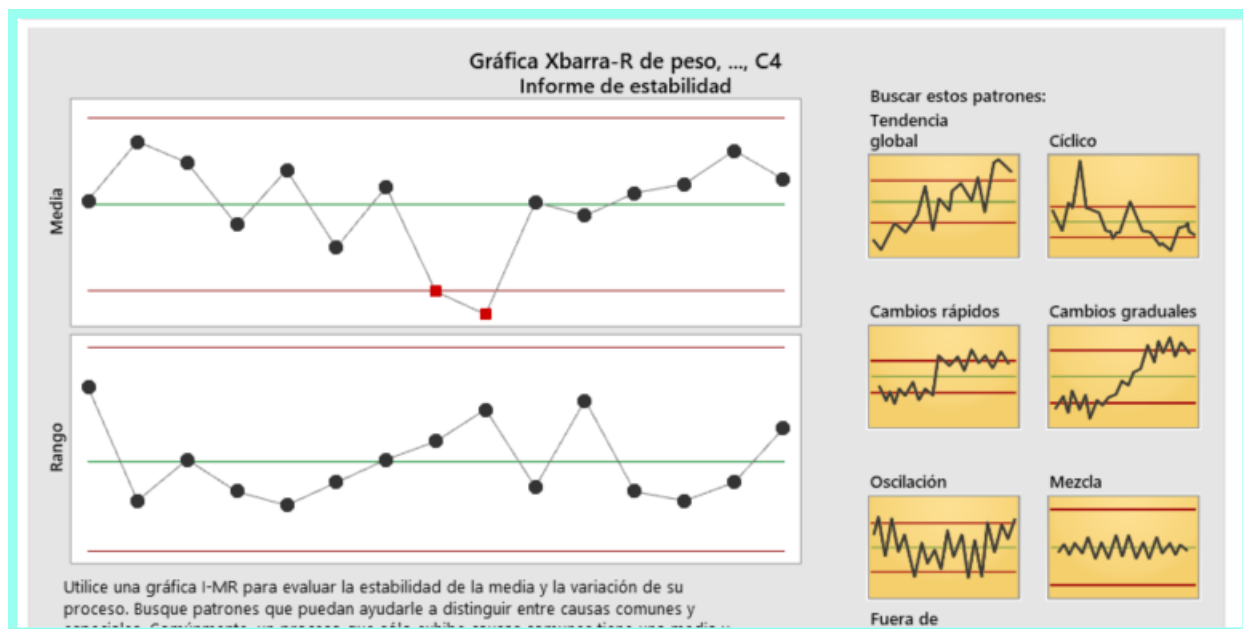
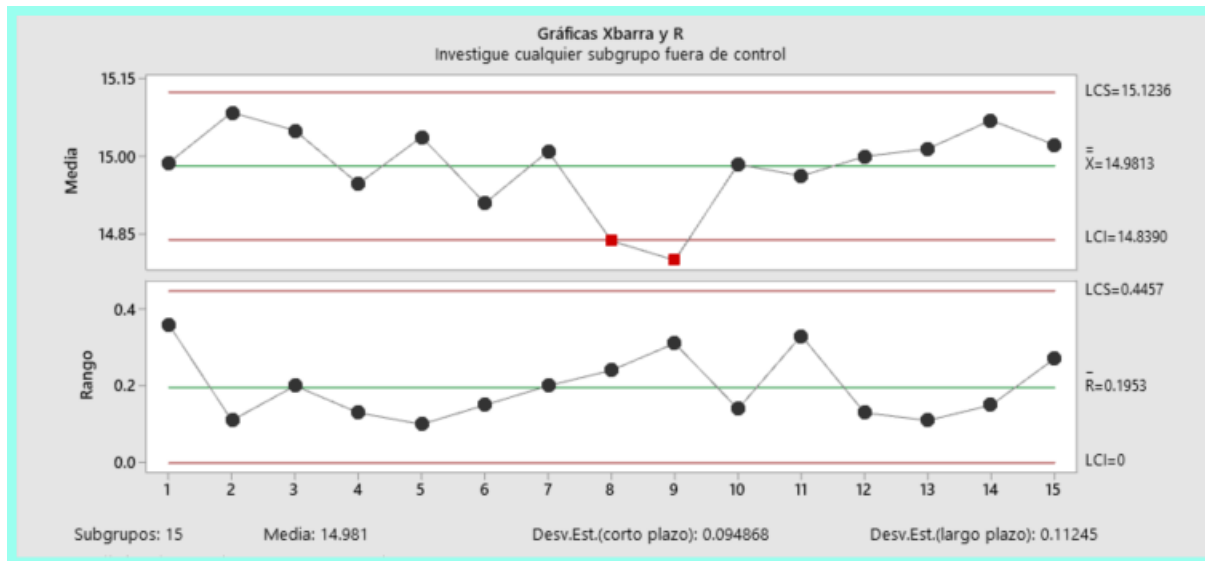
Línea central = $\mu = 15.0$ onzas

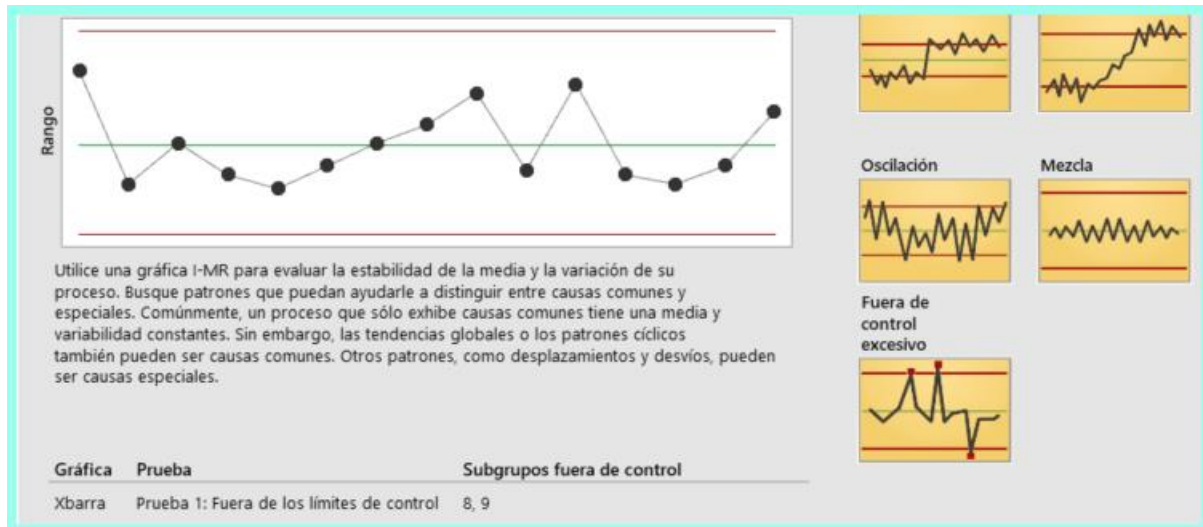
$$\text{Límites de control} = \mu \pm 3 \frac{\sigma}{\sqrt{n}}$$

$$= 15.0 \pm 3 \frac{0.1}{\sqrt{4}} = 14.85 \text{ y } 15.15 \text{ onzas}$$

Con la aplicación de Minitab tenemos:







Gráfica Xbarra-R de peso, ..., C4 Tarjeta de informe		
Verificar	Estado	Descripción
Estabilidad		La variación del proceso es estable. No hay subgrupos fuera de control en la gráfica R. Sin embargo, la media del proceso pudiera no ser estable. 2 subgrupos (13.3%) están fuera de control en la gráfica Xbarra (hay la posibilidad de que vea 0.7% de subgrupos fuera de control, aunque el proceso sea estable). Debe investigar los subgrupos fuera de control y omitir de los cálculos aquellos con causas especiales.
Cantidad de datos		Es posible que usted no tenga suficientes datos para estimar límites de control precisos. Por lo menos 70 puntos de datos se deberían incluir en los cálculos.
Datos correlacionados		Sus datos pasaron la prueba de correlación. La correlación entre puntos de datos consecutivos dentro de cada subgrupo es menor que 0.2.
Gráficas alternativas		Esta gráfica sirve para supervisar el control del proceso. Si su principal objetivo es explorar sus datos o comparar su proceso antes y después de un cambio, utilice las Gráficas de control disponibles en Análisis gráfico o las Gráficas de control de antes/después.

Ejercicios:

1.- Suponga que se desea estimar el monto medio de las ventas al menudeo por plaza de un producto durante el año pasado. El número de plazas de venta al menudeo es grande. Determine el intervalo de 95% de confianza suponiendo que los montos de venta tienen distribución normal, $\bar{X} = \$3425$, $\sigma = \$200$ y $n = 25$

Respuesta: \$3346.60 a \$3503.40

2.- Con referencia al problema número uno, determine el intervalo de 95% de confianza suponiendo que la población tiene distribución normal, $\bar{X} = \$3425$, $s = \$200$ y $n = 25$

Respuesta: \$3342.44 a \$3507.56

3.- considerando el ejercicio número uno determine el intervalo de 95% de confianza suponiendo que la población *no* tiene distribución normal, $\bar{X} = \$3425$, $s = \$200$ y $n = 50$

Respuesta: \$3369.55 a \$3480.45

4.- Un analista de un departamento de personal selecciona de manera aleatoria los registros de 16 empleados que trabajan por horas y encuentra que la tarifa salarial media por hora es \$9.50. Se supone que las tarifas salariales en la empresa tienen distribución normal. Si se sabe que la desviación estándar de las tarifas salariales es de \$1.00, estime la tarifa salarial media en esa empresa usando el intervalo de 80% de confianza.

Respuesta: \$9.18 a \$9.82

5.- un analista desea estimar con un margen de error de 25 centavos y 90% de confianza la tarifa horaria media de los trabajadores de una empresa. Se estima que la desviación estándar de las tarifas salariales no es mayor que \$1.00. ¿Cuál es el número mínimo de registros de personal que debe tomarse como muestra para satisfacer este objetivo de la investigación?

Respuesta: $43.56 \cong 44$

Referencias bibliográficas:

- Anderson, D.R., Sweeney, D. J. y Williams T. A. (2009). Estadística para Administración y Economía. 10ª edición. México, D.F. Cengage Learning Editores, S. A. de C.V.
- Levine, David M., Krehbiel, Timothy C., Berenson, Mark L. Estadística para administración. Ed. PEARSON Prentice Hall, 2006 4 edición.
- Martínez, E. Estadística. Ed. Santiago de los Caballeros: Universidad Abierta para Adultos (UAPA), 2020. 221 págs. Disponible en: <https://elibro.net/en/ereader/itcdcuauhtemoc/175596?page=16>
- Mason, D.R. (1998) Estadística para Administración y Economía. Santafé de Bogotá, D. C. Alfaomega Grupo Editor, S.A. de C.V.
- Mendenhall, W. R. (1987). Estadística para Administración y Economía. México, D. F. Grupo Editorial Iberoamérica.
- Montgomery, D. et al. (1997) Probabilidad y Estadística para Ingeniería y Administración. México, CECSA.
- Murray R. Spiegel., Larry J. Stephens (2009) 4ª Estadística, serie Schaum, Ed. Mc Graw Hill
- Jay L. Devore. Fundamentos de probabilidad y estadística. Ed. CENGAGE 2018 1 edición.
- Kazmier, Leonard J. Estadística aplicada a administración y economía. Ed. McGraw-Hill Interamericana, 2006 4 edición.

