



Universidad Central del Ecuador
Facultad de Ciencias Administrativas
Modalidad a Distancia

UNIDAD DIDÁCTICA

Asignatura: Estadística Descriptiva

Nivel: Segundo y Tercero

Autor: Ing. Flavio Parra.

Docentes: Lic. Sofía Rodríguez

Dra. Norma Zurita

Econ. María Augusta Lara

Periodo: 2026 - 2027

Contenido

Introducción	6
Objetivos de la Asignatura.....	6
Objetivo General	6
Objetivos Específicos.....	6
Contribución de la Asignatura.....	7
Relación de la Asignatura con otras Materias	7
Bibliografía y Netgrafía.....	8
Desarrollo de las Unidades de Estudio	9
1. UNIDAD I.....	9
1. Definición y organización de los datos.....	9
1.1 Introducción.....	9
1.2 Definición.....	11
1.3 Clases de estadística	13
1.4 Tipos de variables	15
1.5 Fuentes de datos estadísticos.....	16
1.6 Niveles de medición.....	16
1.7 Descripción de los datos	22
1.7.1 Distribución de frecuencias.	22
1.7.2 Distribuciones de frecuencias de datos cualitativos	23
1.7.2.1 Representación gráfica de datos cualitativos	24
1.7.3 Distribuciones de frecuencias de datos cuantitativos.....	25
1.7.4 Representaciones de tallo y hoja.....	31
1.7.5 Representación gráfica de una distribución de frecuencias	36
1.7.6 Otras representaciones gráficas.	41
2. UNIDAD II.....	45
Medidas de tendencia central y números índices	45
2. Medidas de tendencia central	45
2.1 Media de la población	45
2.2 Media de la muestra.....	46
2.2.1 Media ponderada.	47
2.3 Mediana.....	48
2.4 Moda.....	49
2.5 Media geométrica.....	50
2.5.1 Media geométrica en un promedio en un periodo dado	50

	3
2.6 Media mediana y moda para datos agrupados	51
2.6.1 Media.....	51
2.6.2 Mediana.....	52
2.6.3 Moda.....	54
2.6.4 Selección de un promedio para datos de una distribución de frecuencias.	55
2.7 Medidas de dispersión.....	56
2.7.1 Amplitud de variación o rango	57
2.7.2 Desviación media	57
2.7.3 Varianza y desviación estándar	58
2.8 Medidas de dispersión para datos agrupados en una distribución de frecuencias	61
2.8.1 Amplitud de variación.....	61
2.8.2 Varianza y desviación estándar de una muestra	61
2.8.2 Teorema de Chevyshev.....	63
2.8.3 Regla Empírica	64
2.8.4 Dispersión relativa	65
2.8.5 Posiciones relativas de la media, la mediana y la moda.....	66
2.8.6 Asimetría o Curtosis.....	66
2.8.7 Cuartiles deciles y centiles (o percentiles)	67
2.9 Diagrama de caja.	70
2.10 Números índices	72
2.10.1 Clasificación de los números índices	72
2.10.2 Números índices simples	72
2.10.3 Números índices no ponderados	74
2.10.3.1 Promedio simple de los índices de precios.....	74
2.10.3. Índice agregado simple.....	75
2.11 Números índices ponderados	75
2.11.1 Índice de Laspeyres.....	76
2.11.2 Índice de Paasche	76
2.11.3 Índice de Fisher.....	78
2.11.3 Índices de valor.....	78
3. UNIDAD III.....	79
Introducción a la teoría de probabilidades	79
3.1 Probabilidad.....	80
3.1.1 Experimento	81
3.1.2 Resultado	81
3.1.3 Evento.....	82

	4
3.2 Enfoques de probabilidad	83
3.2.1 Probabilidad clásica	83
Mutualmente excluyente	83
Colectivamente exhaustivo.....	83
3.2.2 Probabilidad empírica.....	84
3.2.3 Probabilidad subjetiva	85
3.3 Reglas de probabilidad	88
Reglas de adición	88
3.3.1 Regla especial de Adición	88
Diagrama de Venn	90
3.3.2 Regla general de la adición	93
3.4 Reglas de multiplicación	96
3.4.1 Regla especial de la multiplicación	96
Independiente	97
3.5 Eventos dependientes	97
3.5.1 Probabilidad condicional	98
3.6 Tablas de contingencias.....	100
3.6.1 Diagramas de árbol (arborigramas)	102
3.6.2 Teorema de Bayes	105
3.7 Principios de conteo	110
3.7.1 Formula de la multiplicación.....	110
3.7.2 Permutaciones.....	112
3.7.2 Combinaciones	113
4. UNIDAD IV	114
4. Distribución de probabilidad discreta y continua	114
4.1 Distribución probabilística	114
4.2 Variables aleatorias	116
4.2.1 Variable aleatoria discreta.....	116
4.2.2 Variable aleatoria continua	117
4.2.3 Media, varianza y desviación estándar de una distribución de probabilidad	117
4.3 Distribución probabilística binomial	121
4.3.2 Distribuciones probabilísticas acumulativas.	128
4.3.3 Distribución probabilística hipergeométrica.....	129
4.3.4 Distribución probabilística de Poisson.....	132
4.4 Distribuciones probabilísticas continuas	135
4.4.1 Distribuciones de probabilidad uniforme	135

	5
4.4.2 Distribuciones de probabilidad normal.....	137
4.4.2.1 Distribución probabilística estándar.....	138
4.4.3 Regla empírica	140
4.4.4 Aproximación de normal a binomial.....	149
4.4.4.1 Factor de corrección de continuidad.....	150
4.5 Distribución probabilística exponencial.....	154
5. Anexo: Correlación y Regresión Lineal	158
¿Qué es el análisis de correlación?.....	158
Coeficiente de correlación.....	159
Análisis de regresión.....	161
El coeficiente de determinación	163
Bibliografía.....	164

Introducción

La asignatura de Estadística Descriptiva constituye un componente formativo esencial en las tres carreras de la Facultad de Ciencias Administrativas, pues proporciona herramientas cuantitativas para organizar, describir e interpretar información relevante para la toma de decisiones contables, empresariales y de gestión pública. En Contabilidad y Auditoría permite estructurar, analizar y presentar datos financieros y contables mediante tablas y gráficos, apoyando procesos como la selección de muestras de auditoría, el control de costos y la evaluación de variaciones en los estados financieros. En Administración de Empresas aporta insumos para el análisis de mercados, la gestión de la producción, la evaluación de riesgos y el monitoreo de indicadores de desempeño organizacional, facilitando decisiones coherentes con los lineamientos de desarrollo local y nacional. En Administración Pública resulta crucial para el tratamiento de información demográfica, socioeconómica y presupuestaria, permitiendo asignar recursos de forma eficiente y diseñar políticas públicas alineadas con las necesidades reales de la población.

A través del estudio de distribuciones de frecuencia, medidas descriptivas, regresión, correlación, series de tiempo, probabilidades y números índice, el estudiantado desarrolla capacidades para analizar datos reales provenientes de organizaciones públicas y privadas, aportando evidencia objetiva para la planificación, el control y la evaluación de resultados. Estos contenidos se trabajan articulando teoría, resolución de problemas y uso de software estadístico, de modo que el estudiante aprenda a transformar datos en información comprensible, a interpretar indicadores y a comunicar hallazgos de manera rigurosa y ética en distintos contextos profesionales.

Objetivos de la Asignatura

Objetivo General

Desarrollar en el estudiante competencias para organizar, analizar e interpretar datos cuantitativos mediante técnicas de estadística descriptiva, a fin de apoyar la toma de decisiones en los ámbitos contable-financiero, empresarial y de la administración pública.

Objetivos Específicos

- Reconocer los conceptos fundamentales de la estadística, tipos de variables, poblaciones y muestras, aplicados a problemas administrativos reales.
- Organizar y representar conjuntos de datos mediante tablas de frecuencia y gráficos estadísticos, utilizando herramientas tecnológicas básicas.
- Calcular e interpretar medidas de tendencia central, posición, dispersión y forma para variables discretas y continuas en contextos empresariales, contables y públicos.
- Aplicar modelos elementales de regresión, correlación, series de tiempo y números índice para analizar la evolución y relación de variables económicas, financieras y sociales.

- Utilizar los resultados del análisis estadístico para sustentar decisiones y propuestas de mejora en proyectos de diagnóstico empresarial, comportamiento organizacional y normativa del sector público

Contribución de la Asignatura

La asignatura de Estadística Descriptiva contribuye a que el estudiantado integre el razonamiento cuantitativo en la solución de problemas propios de cada campo profesional, articulando la lectura crítica de datos con la toma de decisiones responsables. En la carrera de Contabilidad y Auditoría, se convierte en un soporte técnico para el análisis sistemático de la información financiera, la evaluación de la razonabilidad de saldos y tendencias, y el sustento numérico de informes contables y de auditoría. En Administración de Empresas, potencia la capacidad para diagnosticar el desempeño organizacional, valorar escenarios de riesgo, comparar alternativas estratégicas y justificar inversiones con base en evidencia estadística. En Administración Pública, permite fundamentar el diseño, implementación y evaluación de programas y políticas mediante el uso de indicadores sociales, económicos y territoriales, favoreciendo una gestión estatal orientada a resultados y al buen vivir de la ciudadanía.

Relación de la Asignatura con otras Materias

Aspecto	Contabilidad y Auditoría	Administración de Empresas	Administración Pública
Código	DDA02BEM04	DED02BEM02	DPA03BFT02
Semestre	Segundo	Segundo	Tercero
Total horas	120 horas	160 horas	120 horas
Prerrequisitos	Matemática (DDA01BFT03).	Matemática (DED01BFT05).	Matemática (DPA01BFT02).
Materias con las que se articula	Contabilidad General, Costos, Auditoría, Tributación, Finanzas, Sistemas de Información Contable.	Fundamentos de Administración, Comportamiento Organizacional, Finanzas, Producción y Operaciones, Investigación de Mercados.	Teoría del Estado, Economía Pública, Planificación y Presupuesto, Gestión de Políticas Públicas, Diseño de la Investigación.
Énfasis profesional	Análisis de datos contables y financieros para el control y la toma de decisiones.	Análisis estadístico para la gestión empresarial, evaluación de riesgos y mejora de procesos.	Análisis de información socioeconómica para la gestión pública, gobernabilidad e interculturalidad

Bibliografía y Netgrafía

Bibliografía básica

- Lind, D. A., Marchal, W. G., & Wathen, S. A. (2012). *Estadística aplicada a los negocios y a la economía* (15.^a ed.). McGraw-Hill.
- Marchal, W. G., & Wathen, S. A. (2012). *Estadística aplicada a los negocios y la economía*. McGraw-Hill.
- Rustom, A. (2012). *Estadística descriptiva, probabilidad e inferencia: Una visión conceptual y aplicada*. Universidad de Chile.
- Díaz Mata, A. (2013). *Estadística aplicada a la administración y la economía*. McGraw-Hill.

Bibliografía complementaria

- Webster, A. L. (2000). *Estadística aplicada a los negocios y a la economía* (3.^a ed.). McGraw-Hill.
- Salvador, L. (2022). *Estadística descriptiva y probabilidades*. LEOSalvador.
- Salvador, L., Revelo, E., Jurado, R., Zurita, N., Ruiz, J., & Andrade, M. (2017). *Análisis estadístico básico para negocios*.
- Estévez Echanique, R. (2024). *Análisis estadístico*.

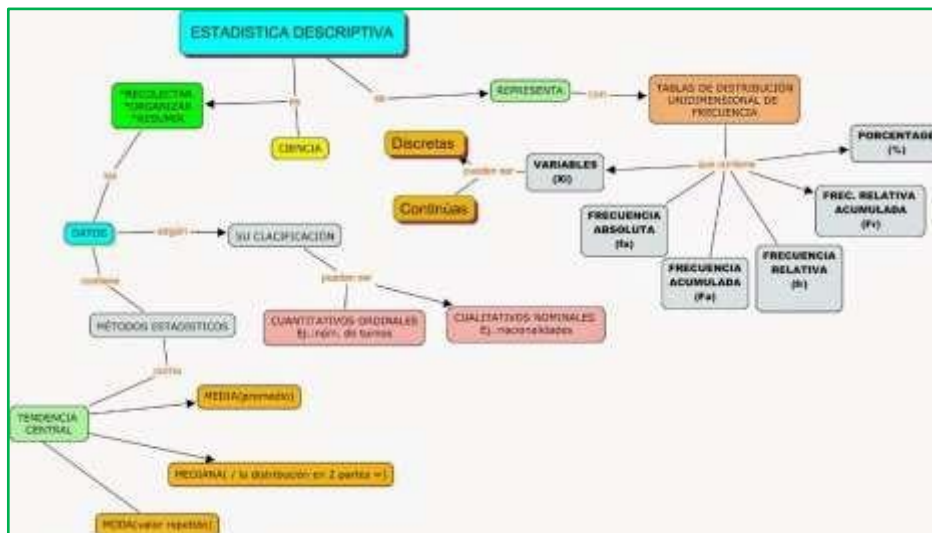
Netgrafía y recursos digitales

- Repositorio Digital UCE. <http://www.dspace.uce.edu.ec>
- Catálogo en línea BFCA – Biblioteca Facultad de Ciencias Administrativas. <https://catalogosib.uce.edu.ec>
- Plataformas y software estadístico (Excel, PSPP, R) utilizados para el procesamiento de datos en las actividades de aula y proyectos integradores.

Desarrollo de las Unidades de Estudio

1. UNIDAD I

1. Definición y organización de los datos



1.1 Introducción

La estadística descriptiva es una parte fundamental de la estadística y juega un papel crucial en la comprensión y el análisis de los datos. Su importancia radica en varios aspectos clave:

1. **Resumen de datos:** La estadística descriptiva proporciona herramientas para resumir datos de manera concisa y comprensible. Esto facilita la presentación de información compleja en forma de tablas, gráficos y medidas resumen, lo que permite a las personas obtener una visión general de los datos sin necesidad de examinar cada observación individual.
2. **Simplifica datos complejos:** En situaciones en las que se manejan grandes conjuntos de datos, la estadística descriptiva ayuda a simplificar la información para que sea más manejable. Esto es esencial para la toma de decisiones informadas en diversos campos, como la investigación, los negocios y la ciencia.
3. **Identificación de patrones y tendencias:** Mediante el uso de técnicas de estadística descriptiva, es posible identificar patrones, tendencias y características clave en los datos. Esto es crucial para comprender el comportamiento de las variables y tomar decisiones basadas en evidencia.
4. **Comparación de datos:** La estadística descriptiva permite la comparación de diferentes conjuntos de datos, lo que es fundamental en la investigación científica, la evaluación de

políticas públicas y la toma de decisiones empresariales. Permite determinar si existen diferencias significativas entre grupos y, en caso afirmativo, analizar la naturaleza de esas diferencias.

5. **Presentación visual:** La representación gráfica de los datos, como histogramas, gráficos de barras y diagramas de dispersión, es una parte esencial de la estadística descriptiva. Estos gráficos proporcionan una representación visual de los datos, lo que facilita la comunicación de resultados y conclusiones a una audiencia no especializada.
6. **Establece una base para la estadística inferencial:** La estadística descriptiva es la base sobre la cual se construye la estadística inferencial. Los resúmenes y las medidas de tendencia central (como la media y la mediana) se utilizan en la inferencia estadística para tomar decisiones y hacer predicciones sobre poblaciones más amplias a partir de muestras más pequeñas.
7. **En resumen,** la estadística descriptiva es esencial para la organización, resumen y visualización de datos, lo que a su vez facilita la toma de decisiones informadas en una amplia variedad de campos, desde la investigación científica hasta la toma de decisiones empresariales. Sin una comprensión sólida de la estadística descriptiva, sería difícil interpretar y utilizar eficazmente la información contenida en los datos.

Todos los días usted hace acopio de noticias e información que sirven de guía para su vida; la misma que le conduce a tomar decisiones, por ejemplo:

- Escuchar el informe meteorológico, le ayuda a decidir qué tipo de ropa usar.
- Administradores y Auditores deben tomar decisiones todos los días. Aunque en ocasiones evidencian “impulsos viscerales” al tomar decisiones (formalmente conocida como decisión no estructurada). La mayoría de veces deciden a partir de hechos concretos y de la experiencia.

Sin embargo, pueden aprender procedimientos y métodos que le ayudaran a tomar decisiones basadas en hechos concretos. Cuando comience a familiarizarse con los procedimientos y métodos implicados en la recolección, presentación y elaboración de resúmenes de un conjunto de datos, o a obtener conclusiones acerca de tales datos, entonces habrá descubierto la estadística.

- La palabra estadística la encontramos frecuentemente en nuestro lenguaje diario, en el uso más común, la estadística se refiere a una información numérica.

1.2 Definición

David M. Levine “Estadística es la rama de la Matemática que examina las formas de procesar y analizar datos”.

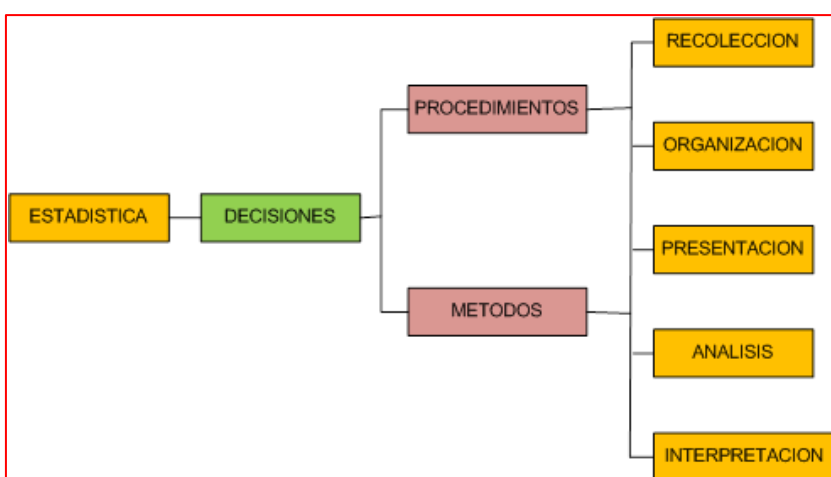
La estadística ofrece los procedimientos para recolectar y transformar los datos de manera que sean útiles a quienes toman decisiones en los negocios.

Mason/Lind/Marchal/ “Estadística es la ciencia de recolectar, organizar, presentar, analizar e interpretar datos para ayudar en una toma de decisiones más efectiva”

Como indica la definición, el primer paso en la investigación de un problema, es la recolección de datos importantes, la cual debe organizarse de cierta manera y presentarse en un gráfico (barras, histogramas, pastel). Podemos analizar e interpretar la información únicamente hasta cuando los datos hayan sido organizados.

Hace más de 100 años H.G. Wells, historiador y autor inglés señaló que: “Algún día el conocimiento estadístico será tan necesario para ser un buen y eficiente ciudadano, como la aptitud para leer”.

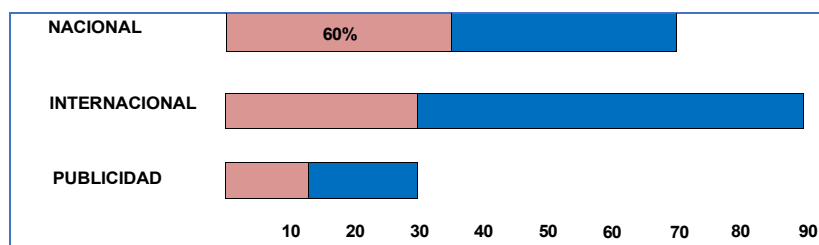
Adaptando ese señalamiento a la actualidad podríamos decir: “El conocimiento estadístico es necesario no solo para ser un buen ciudadano, sino también para una toma de decisiones efectivas en varias áreas de los negocios”



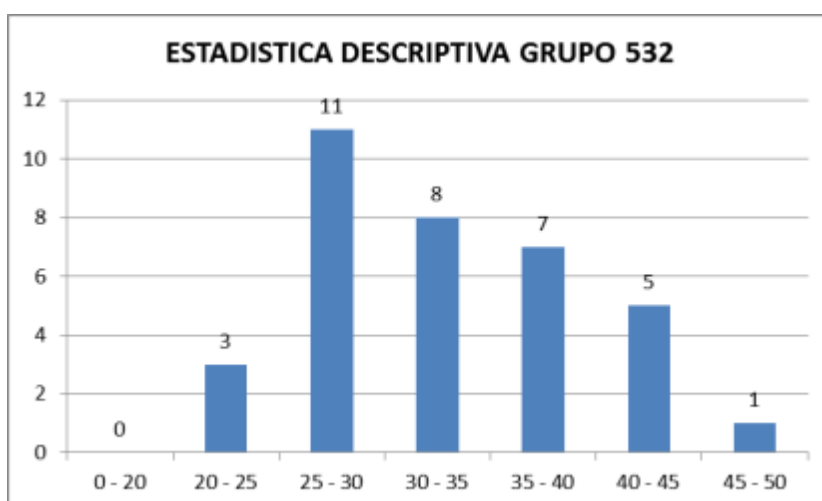
De acuerdo a la definición los pasos que sigue la estadística son:

- El primer paso en la investigación de un problema es la recolección de datos importantes.

- Estos datos deben organizarse de cierta manera y usualmente presentarse en forma de gráfico, como el presentado que muestra el porcentaje de ingresos del Club Barcelona-Ecuador, en las distintas competiciones del año 2015-2016, del total de ingresos a nivel nacional.



- El promedio de notas de los estudiantes del periodo 52 en la asignatura de Estadística Descriptiva en el primer hemisemestre es otro ejemplo de organización de datos.



- Solo una vez que los datos hayan sido organizados podemos analizarlos e interpretarlos. En el ejemplo de los ingresos del Club Barcelona, se nota claramente que los ingresos a nivel internacional de un total de 90 millones de dólares, únicamente participo en un 30% de ese valor, lo que indicaría que debería aumentar esa participación.
- Del análisis de notas del grupo 532 de la Asignatura de Estadística Descriptiva, se puede colegir que existe un número exagerado de estudiantes con notas de 25-30, los estudiantes y el profesor deben tomar planes remediales, para solucionar el problema.

Para una correcta toma de decisiones basada en una información se requiere:

1. Determinar si la información existente es adecuada o si se requiere información adicional.
2. Reunir información adicional, si es necesario, de tal forma que no haya resultados erróneos.

3. Resumir la información de modo útil o informativo.
4. Analizar la información disponible.
5. Sacar conclusiones y realizar las inferencias necesarias al tiempo que evalúa el riesgo de llegar a una conclusión incorrecta.

Razones para estudiar estadística

1. Los datos se encuentran en todos lados.
2. Las técnicas estadísticas se utilizan para la toma de muchas decisiones que afectan a nuestra vida.
3. Sin importar su línea de trabajo futura, usted tendrá que tomar decisiones que involucran datos.

1.3 Clases de estadística

Generalmente el estudio de la Estadística se divide en dos categorías: **la Estadística Descriptiva** y la Estadística Inferencial.

Estadística descriptiva

Conjunto de métodos para organizar, resumir y presentar los datos de manera informativa.

Ejemplo: Población del Ecuador por décadas.

Año	Número de Habitantes	Porcentaje de crecimiento
1960	6'520.000	
1970	7'850.000	16,94
1980	9'425.000	16,71
1990	10'326.000	8,73
2000	12'327.000	16,23

$$\Delta = 7'850\,000 - 6'520\,000 = 1'330\,000$$

$$\% \Delta = \frac{1'330\,000}{7'850\,000} * 100 = 16.94\%$$

Estadística inferencial, inferencia estadística o estadística inductiva

Conjunto de métodos utilizados para saber algo acerca de una población, basándose en una muestra.

Población

Conjunto de todos los posibles individuos, objetos o medidas de interés. Los habitantes de la ciudad de Quito-Ecuador es un ejemplo de población.

Muestra

Una porción, o parte, de la población de interés. Los habitantes del barrio de “La Floresta” de la ciudad de Quito, es una parte de todos los barrios de la ciudad.

Unidad experimental

Es la unidad más pequeña de la cual se obtiene una medida o característica. Por ejemplo, a un investigador le puede interesar: las familias, las personas, las empresas de un cierto tipo, etc.

Variable de respuesta

Es una característica individual de cada elemento de la población o de una muestra; por ejemplo: la edad de un alumno al ingresar a la universidad, el color del pelo, su estatura, su peso, etc.

Pieza de datos

El valor de la variable de respuesta asociado con cada elemento, será la pieza de datos, por ejemplo: el coche es verde, Jorge ingresó a la universidad a la edad de 23 años. Si nos damos cuenta estamos asignando a cada elemento su variable o característica.

Parámetro

Es una característica medida de una población completa, por ejemplo: la proporción de alumnos de más de 21 años que ingresan a la universidad. En estadística se asignan símbolos del alfabeto griego para designar un parámetro. $\mu = 1.63 \text{ metros}$

Experimento

Es una actividad planificada cuyos resultados nos producen un conjunto de datos.

Un experimento es un procedimiento mediante el cual se trata de comprobar (confirmar o verificar) una o varias hipótesis relacionadas con un determinado fenómeno, mediante la manipulación y el estudio de las correlaciones de las variables que presumiblemente son su causa

Estadístico

Es la medida de una característica relativa a la muestra, al valor promedio de los datos y la imagen de éstos; la mayoría de los estadísticos muestrales se encuentran por medio de fórmulas y suelen asignárseles símbolos del alfabeto latino. $\bar{X} = 1.67 \text{ metros}$

1.4 Tipos de variables

Existen dos tipos básicos de datos:

1. Los obtenidos a partir de una población cualitativa.
2. Los datos que resultan de una población cuantitativa.

Variable cualitativa

Una variable es cualitativa cuando las características de los datos o variable en estudio son no numéricas, se la denomina variable cualitativa o atributo. Ejemplos: Genero sexual, religión, tipo de automóvil, lugar de nacimiento, color de ojos.

Si la información estudiada es cualitativa, generalmente interesa saber cuántas o que proporción entra en cada categoría; ejemplo: ¿Qué porcentaje de los estudiantes de la UCE son de Quito? ¿Qué porcentaje de católicos o evangélicos hay en Guayaquil? Usualmente los datos cualitativos se resumen en diagramas o gráficos de barras, pastel.

Variable cuantitativa

Cuando la variable estudiada se puede expresar numéricamente, se denomina variable cuantitativa. Ejemplos: saldo en una cuenta de ahorros, el número de hijos en una familia, el número de dormitorios en una casa.

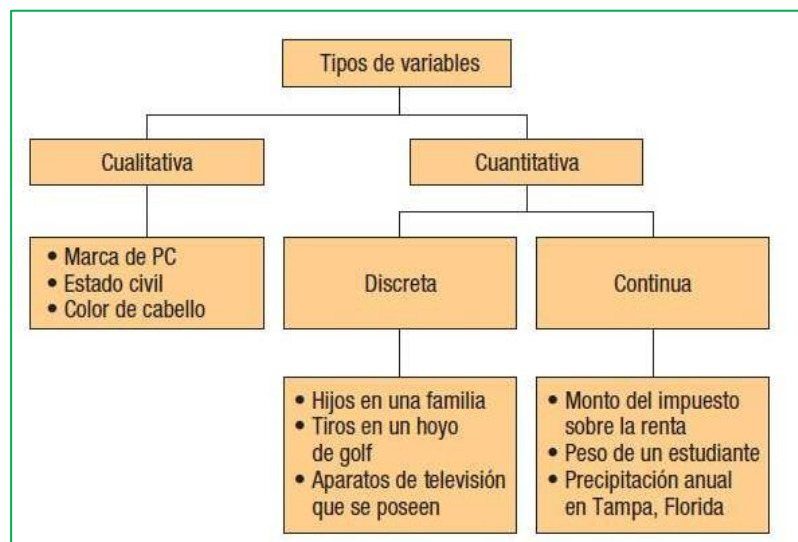
Las variables cuantitativas se clasifican en discretas y continuas.

Variables discretas

Solo pueden tomar ciertos valores (enteros) y generalmente existen brechas entre ellas. Ejemplo: El número de habitaciones en una casa (1, 2, 3...); el número de autos que pasan por el peaje de la autopista del Valle de los Chillos. Por lo general las variables discretas son el resultado del conteo.

Variables continuas

Pueden tomar cualquier valor dentro de un intervalo determinado. Ejemplos: Presión de autos de los neumáticos de un automóvil, tiempo de vuelo Quito-Guayaquil, precio de una bebida refrescante. Las variables continuas son el resultado de la medición de algo de interés.



1.5 Fuentes de datos estadísticos

1. **Estadísticas o fuentes primarias:** Son datos obtenidos ya sea por encuestas directas, mediante la utilización de cuestionarios, o como resultado de la observación directa.
2. **Estadísticas o fuentes secundarias:** los datos se obtienen de publicaciones, las cuales pueden ser reproducciones totales o parciales. Son fuentes secundarias: bibliotecas, centros de documentación, folletos, revistas, archivos, etc.
3. **Estadísticas o series temporales:** denominadas series de tiempo o series cronológicas. Son las obtenidas en forma cronológica, siendo el resultado de la investigación y observaciones periódicas: días, meses o años.

1.6 Niveles de medición

Los niveles de medición se utilizan para describir la naturaleza de los datos recopilados y determinar qué operaciones matemáticas y estadísticas son apropiadas para analizarlos.

Existen 4 niveles de medición: Nominal, Ordinal, de Intervalo y de Razón.

El nivel “más bajo” o más primitivo es el nominal. El “más alto” o el que proporciona la mayor cantidad de información acerca de la observación, es la medición de nivel de razón.

Datos de nivel nominal

En este nivel, las categorías son simplemente etiquetas o nombres que se utilizan para distinguir entre diferentes grupos o categorías. No hay un orden específico entre las categorías, y las únicas operaciones que se pueden realizar son la igualdad y la desigualdad. Ejemplos de datos nominales son el género (masculino, femenino, otro), la nacionalidad (estadounidense, mexicano, canadiense), o el color de ojos (azul, verde, marrón).

En el nivel nominal de la medición, las observaciones solo pueden clasificarse o contarse. No existe algún orden específico para las marcas o categorías.

Ejemplo: Análisis de llamadas en un día del mes de marzo-2018 de las operadoras telefónicas a nivel nacional.

Nº	Compañía	Nº de llamadas	Porcentaje
1	Claro	4'350.000	45,31
2	Movi	3'750.000	39,06
3	CNT	1'150.000	11,98
4	Otros	350.000	3,65
TOTAL		9'600.000	100

Los datos tienen las siguientes características:

Mutualmente excluyentes: cuando una persona, objeto o medición se ha de incluir en tan solo una categoría.

Exhaustivo: cada individuo, objeto o medición debe aparecer en una categoría.

En resumen, los datos de nivel nominal cuentan con las siguientes propiedades:

- Las categorías de datos son mutuamente excluyentes, por lo tanto, un objeto pertenece a una sola categoría.
- Las categorías de datos no cuentan con un orden lógico.

Datos de nivel ordinal.

En este nivel, las categorías tienen un orden específico, pero la distancia entre ellas no es uniforme ni significativa. Puedes comparar y ordenar los datos, pero no puedes realizar operaciones matemáticas significativas entre ellos. Ejemplos de datos ordinales incluyen las calificaciones (A, B, C, D, F), el nivel de satisfacción (muy insatisfecho, insatisfecho, neutral, satisfecho, muy satisfecho) o la clasificación social (baja, media, alta).

Ejemplo: Las calificaciones dadas por los estudiantes a un profesor de la asignatura de Estadística Descriptiva. Cada estudiante respondió a la siguiente pregunta: ¿Cómo califica al profesor de esta asignatura?, de acuerdo a las categorías siguientes:

CALIFICACION	FRECUENCIA
Superior	6
Bien	28
Regular	25
Malo	12
Muy malo	3

No podemos establecer con seguridad la diferencia entre Superior y Bien, o si esa diferencia es igual entre Regular y Malo. Las características de este nivel son:

1. Las categorías de datos son mutuamente excluyentes y exhaustivas.
2. Dichas categorías se clasifican u ordenan de acuerdo a las características particulares que poseen.

Datos de nivel de intervalo.

Nivel de intervalo incluye todas las características del nivel ordinal, pero, además, la diferencia entre valores es un tamaño constante.

En este nivel, las categorías tienen un orden específico, y la distancia entre ellas es uniforme y significativa, pero no existe un punto cero absoluto. Puedes realizar operaciones de suma y resta, pero no de multiplicación o división. Ejemplos de datos de intervalo incluyen la temperatura en grados Celsius, la puntuación en una escala de 1 a 10 en una encuesta de satisfacción.

Supongamos que medimos la temperatura de Quito en dos horas diferentes, el primer valor sea de 12 grados centígrados y el segundo de 20; podemos afirmar que entre esos valores existe una diferencia de 8 grados y podemos establecer porque el grado centígrado es un tamaño constante. En esta escala el cero, es un punto más de la escala, en el ejemplo el cero grados indica ausencia de calor, o que hace mucho frio; en otras palabras, no se pierde la esencia de la propiedad de la investigación.

ESCALA	CALIFICACION
0 - 5	Muy frio
5 - 10	Frio
10 - 15	Poco caluroso
15 - 20	Caluroso
20 - 25	Muy caluroso
25 - 30	Exc. Caluroso

Propiedades del nivel de intervalo son:

- Las categorías de datos son mutuamente excluyentes y exhaustivos.

- Las citadas categorías se clasifican de acuerdo con la cantidad de la característica que posean.
- Las diferencias iguales en la característica son representadas por iguales diferencias en los números asignados a las categorías.

Datos de nivel de razón.

Este es el nivel más avanzado, donde las categorías tienen un orden específico, la distancia entre ellas es uniforme y significativa, y hay un punto cero absoluto que implica la ausencia de la característica medida. En este nivel, se pueden realizar todas las operaciones matemáticas, incluyendo suma, resta, multiplicación y división. Ejemplos de datos de razón incluyen la edad, el peso, la altura, el ingreso y el tiempo transcurrido.

Es el nivel más alto, cuenta con todas las características del nivel de intervalo, incluyendo el hecho de que el punto cero es importante y la razón (cociente) entre dos números también es importante. Ejemplos: los salarios, la producción, el peso y la estatura.

Cuando hablamos de dinero, el tener cero dólares implica que no se tiene dinero, además indica la ausencia de la propiedad de la observación al igual que una báscula marque cero implica la ausencia total de peso.

La razón entre dos números también es importante; por ejemplo, una persona gano en la anterior semana \$1500 y en la actual \$3000, podemos establecer que esta semana se ganó el doble que la anterior.

Las propiedades del nivel de razón son:

- Las categorías de datos son mutuamente excluyentes y exhaustivas.
- Dichas categorías se clasifican de acuerdo con la cantidad de la característica que posean.
- Las diferencias iguales en las características son representadas por diferencias iguales en los números asignados a las categorías.
- El punto (valor) cero representa la ausencia total de la característica.

Ejemplo: En la tabla se presenta la investigación de los ingresos anuales de padre e hijo, de una muestra de grupos familiares.

NOMBRE DE FAMILIA	PADRE	HIJO
Jiménez	\$ 80.000	\$ 40.000
Blanco	\$ 90.000	\$ 30.000
Ruiz	\$ 60.000	\$ 120.000
Sánchez	\$ 75.000	\$ 130.000

Tenga en cuenta que el cero es importante y la razón entre dos números también es importante, afirmamos que en la familia Jiménez el padre gana el doble que el hijo por ejemplo.

Resumen



Los datos también son clasificados en cualitativos y cuantitativos. Los datos cualitativos comprenden etiquetas o nombres que se usan para identificar un atributo de cada elemento. Los datos cualitativos emplean la escala nominal o la ordinal y pueden ser numéricos o no. Los datos cuantitativos requieren valores numéricos que indiquen cuánto o cuántos. Los datos cuantitativos se obtienen usando las escalas de medición de intervalo o de razón.

Ejemplos:

1. ¿Cuál es el nivel de medición de cada una de las siguientes variables?

a) Coeficientes intelectuales de los estudiantes.

Cuantitativo-De razón

b) La distancia que viajan los estudiantes para llegar a clases.

Cuantitativa-De razón

c) Los números en los uniformes de un equipo universitario femenino de fútbol.

Cuantitativo- De intervalo

d) Una clasificación de estudiantes por fecha de nacimiento.

Cualitativa Ordinal

e) Una clasificación de estudiantes que cursan primero, segundo, tercero o último grados.

Cualitativa-Ordinal

f) Número de horas que los alumnos estudian a la semana.

De razón

Ejemplo: La Columbia House vende discos compactos a los miembros de su club de venta por correo. En una encuesta sobre música se les pidió a los nuevos miembros del club que llenaran un cuestionario con 11 preguntas. Algunas de las preguntas eran:

- a. ¿Cuántos discos compactos has comprado en los últimos 12 meses?
- b. ¿Eres miembro de algún club de venta de libros por correo (Sí o No)?
- c. ¿Cuál es tu edad?
- d. Incluyéndote a ti, de cuántas personas (adultos y niños) consta tu familia.
- e. ¿Qué tipo de música te interesa comprar? Se presentaban quince categorías entre las que se encontraban rock pesado, rock ligero, música contemporánea para adultos, rap y rancheras.

Responda si los datos que se obtienen con cada pregunta son cualitativos o cuantitativos.

Ejemplo: En una encuesta de “El Comercio” (13 de octubre de 2013) se les hacen a los suscriptores 46 preguntas acerca de sus características e intereses. De cada una de las preguntas siguientes, ¿cuál proporciona datos cualitativos o cuantitativos e indica la escala de medición apropiada?

- a. ¿Cuál es su edad?
- b. ¿Es usted hombre o mujer?
- c. ¿Cuándo empezó a leer el Comercio? Preparatoria, universidad al comienzo de la carrera, a la mitad de la carrera, al final de la carrera o ya retirado.
- d. ¿Cuánto tiempo hace que tiene su trabajo o cargo actual?

- e. ¿Qué tipo de automóvil piensa comprarse la próxima vez que compre uno? Ocho categorías para las respuestas, entre las que se encontraban sedán, automóvil deportivo, minivan, etcétera.

Ejemplo: Diga de cada una de las variables siguientes si es cualitativa o cuantitativa e indique la escala de medición a la que pertenece.

- a. Ventas anuales.

Cuantitativa-de razón

- b. Tamaño de los refrescos (pequeño, mediano, grande).

Cualitativa-nominal

- c. Clasificación como empleado (GS 1 a GS 18).

Cualitativa-ordinal

- d. Ganancia por acción.

Cuantitativa-razón

- e. Modo de pago (al contado, cheque, tarjeta de crédito).

Cualitativa-Nominal

1.7 Descripción de los datos

A través de la descripción de los datos podemos apreciar algunas características de ellos mediante gráficas y tablas, que permitan observar la distribución de ellos y así poder tomar decisiones acertadas.

Recuerde la Estadística Descriptiva permite organizar los datos de diferentes maneras y señalar donde los valores de los datos tienden a concentrarse, y ayudar a distinguir los valores mayores y menores. **Un método que se utiliza para describir un conjunto de datos es la distribución de frecuencias.**

1.7.1 Distribución de frecuencias.

Agrupamiento de datos en categorías que muestran el número de observaciones en cada categoría mutuamente excluyente.

Frecuencias relativas de clase

Es posible convertir las frecuencias de clase en frecuencias relativas de clase para mostrar la fracción del número total de observaciones en cada una de ellas. Así, una frecuencia relativa

capta la relación entre la totalidad de elementos de una clase y el número total de observaciones.

Ejemplo 1: La tabla que se indica muestra el número de autos vendidos en los locales de ciudades del país.

Autos vendidos por concesionaria		
Ciudad	Número de autos vendidos	Frecuencia relativa
Quito	52	0,29
Guayaquil	40	0,22
Cuenca	45	0,25
Ambato	43	0,24
Total	180	1,00

1.7.2 Distribuciones de frecuencias de datos cualitativos

Una distribución de frecuencias para datos cualitativos, también conocida como tabla de frecuencias o distribución de frecuencias nominal, se utiliza para resumir y organizar datos que son categorías o etiquetas sin un orden numérico. En lugar de contar valores numéricos, contamos cuántas veces aparece cada categoría en el conjunto de datos.

Ejemplo: Supongamos que tienes un conjunto de datos cualitativos que representan preferencias de color en una muestra de 50 personas. Las categorías podrían ser "Rojo", "Verde", "Azul", "Amarillo" y "Naranja". Para crear una distribución de frecuencias, siga estos pasos:

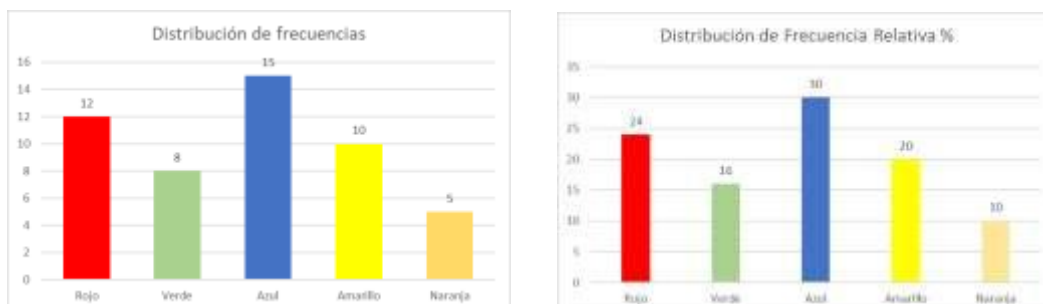
- Enumera todas las categorías cualitativas únicas que aparecen en tus datos. En este caso, serían las cinco opciones de colores mencionadas.
- Cuenta cuántas veces aparece cada categoría en tus datos y registra estas frecuencias en una tabla.

Color	Frecuencia	Frecuencia Relativa %
Rojo	12	24
Verde	8	16
Azul	15	30
Amarillo	10	20
Naranja	5	10
Total	50	100

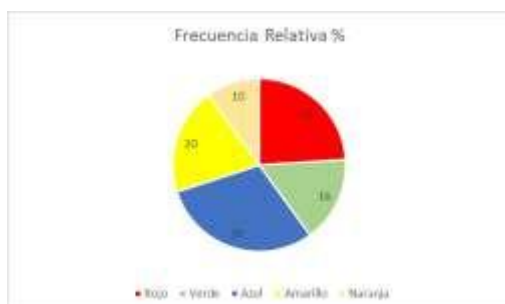
En esta tabla, se muestra cuántas veces cada color aparece en el conjunto de datos. La columna "Color" contiene las categorías cualitativas únicas, y la columna "Frecuencia" indica cuántas veces cada categoría específica fue seleccionada.

1.7.2.1 Representación gráfica de datos cualitativos

El instrumento más común para representar una variable cualitativa en forma gráfica es la Gráfica de barras. En la mayoría de los casos, el eje horizontal muestra la variable de interés y el eje vertical la frecuencia o fracción de cada uno de los posibles resultados. Una característica distintiva de esta herramienta es que existe una distancia o espacio entre las barras. Es decir, que como la variable de interés es de naturaleza cualitativa, las barras no son adyacentes

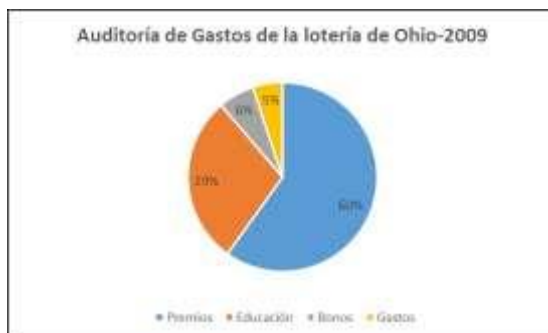


Gráfica de pastel: Gráfica que muestra la parte o porcentaje que representa cada clase del total de números de frecuencia.



Ejemplo: La tabla muestra una auditoria de gastos inversiones de una lotería del estado de Ohio en 2009

Uso del dinero de las ventas	Cantidad (millones de dólares)	Porcentaje de ventas
Premios	1 460.0	60
Educación	702.3	29
Bonos	150.0	6
Gastos	124.3	5
Total	2 436.6	100



1.7.3 Distribuciones de frecuencias de datos cuantitativos

Distribución de frecuencias

Agrupación de datos en clases mutuamente excluyentes, que muestra el número de observaciones que hay en cada clase.

Ejemplo: El Gerente de una distribuidora de autos está interesado en reunir información sobre los precios de venta de los vehículos vendidos en su agencia. Quiere contestar los siguientes interrogantes:

- ¿Cuál es el precio de venta típico?
- ¿Cuál es el mayor precio de venta?
- ¿Cuál es el menor precio de venta?
- ¿Alrededor de qué valor tienden acumularse los precios de venta?

El registro de venta del mes pasado muestra que se vendieron 80 vehículos los mismos que se resume en la tabla siguiente:

Precios de venta de vehículos vendidos en cierto mes									
20197	20372	17454	20591	23651	24453	14266	15021	25683	27872
16587	20169	32851	16251	17047	21285	21324	21609	25670	12546
12935	16873	22251	22277	25034	21533	24443	16889	17004	14357
17155	16688	20657	23613	17895	17203	20765	22783	23661	29277
17642	18981	21052	22799	12794	15263	33625	14399	14968	17356
18442	18722	16331	19817	16766	17633	17962	19845	23285	24896
26076	29492	15890	18740	19374	21571	22449	25337	17642	20613
21220	27655	19442	14891	17818	23237	17445	18556	18639	21296

Solución:

A la información desorganizada de la tabla, nos referiremos como **datos simples o datos no agrupados**. De la observación de los datos lo único que podemos contestar es que el menor precio (\$12 546) y el mayor precio de venta (\$33 625).

$n = 80$ *número de datos*

Valor máximo = \$33 625

Valor mínimo = \$12546

Los datos simples se interpretan más fácilmente si se organizan en una distribución de frecuencias. Para lo cual se sigue los pasos siguientes:

1. Categorías o clases

En número de categorías o clases, se puede determinar con la regla $2^k \geq n$; donde:

$K = \text{número de categorías}$

$n = \text{número de datos u observaciones}$

$$2^7 > 80 \quad K = 7$$

2. Intervalo de clase

Una categoría o clase está compuesta por el límite inferior (Li) y el límite superior (Ls), existiendo entre ellos un valor cuantificable, llamado intervalo de clase, existiendo algunos criterios para su determinación:

$$i = \frac{\text{Valor más alto} - \text{Valor más bajo}}{\text{Número de clase}}$$

$$i = \frac{\text{Valor más alto} - \text{Valor más bajo}}{1 + 3.322(\log. \text{del total de frecuencias})}$$

$$i = \frac{33625 - 12546}{7} = 3011$$

$$i = \frac{33625 - 12546}{1 + 3.322 * (\log 80)} = 2879$$

Resultaría molesto trabajar con los valores encontrados como intervalo de clase, por lo que se redondea a 3000.

3. Primera clase o categoría (Li – Ls)

Para el límite inferior (Li), se puede establecer bajo 2 criterios:

- a) Un valor algo menor que el menor de los datos, en nuestro caso el menor dato es \$12 546, se lo podría establecer en \$12 000
- b) El límite inferior de la primera clase debe ser un múltiplo par del intervalo de clase. En nuestro caso $i = 3000 \times 4 = 12000$ $Li = 12000$ de la primera clase.

El límite superior (Ls) es límite inferior (Li) más el intervalo de clase (i).

$$Ls = Li + i$$

$$Ls = 12\,000 + 3\,000 = \$15\,000$$

Primera categoría o clase: 12 000 a 15 000

4. Clases o categorías

Establecida la primera categoría organizamos la tabla de distribución de frecuencias, tomando en cuenta las consideraciones siguientes:

- a) **Las clases son mutuamente excluyentes** esto quiere decir que cualquier observación únicamente puede estar incluida en una sola categoría.
- b) Evite la superposición de límites de clase establecidos. Aunque al establecer categorías por convención de presentación en la distribución de frecuencias no se cumpliría que las clases sean mutuamente excluyentes, en dos categorías consecutivas el dato del límite superior de la primera clase, corresponde al límite inferior de la segunda categoría.

Supongamos dos categorías consecutivas:

Categoría o clase	Frecuencia
$\geq 1\,300$ a $< 1\,400$	
$\geq 1\,400$ a $< 1\,500$	

Si en el conjunto de datos de existe 1 400, el dato corresponde a la segunda categoría.

- c) Trate de evitar clases abiertas. Como por ejemplo mayor que 50 000 o menor que 20 000, este tipo de clases ocasionan problemas al hacer las gráficas.
- d) Los intervalos de clase de una distribución de frecuencias deben ser iguales. Los intervalos de clase desiguales ofrecen problemas al representar en forma gráfica la distribución.

- e) Las clases se van creando sumando el intervalo de clase, hasta llegar a la última que debe contener el mayor de los datos.

Clases o categorías	Frecuencias
12000 a 15000	
15000 a 18000	
18000 a 21000	
21000 a 24000	
24000 a 27000	
27000 a 30000	
30000 a 33000	
33000 a 36000	

Por conveniencia y para que la tabla de frecuencias este mejor presentada podemos expresarla en miles de dólares y escogemos los datos que corresponden a cada categoría. Completamos la tabla de frecuencias con el punto medio de clases y las frecuencias relativas.

Punto medio de clase (marca de clase)

Es equidistante o igual entre el límite inferior y superior de la clase. Puede ser calculado sumando el límite inferior y el límite superior y ese resultado dividido para dos.

$$\text{Marca de clase} = \frac{Li + Ls}{2}$$

En el ejemplo 1: Tenemos para la primera clase:

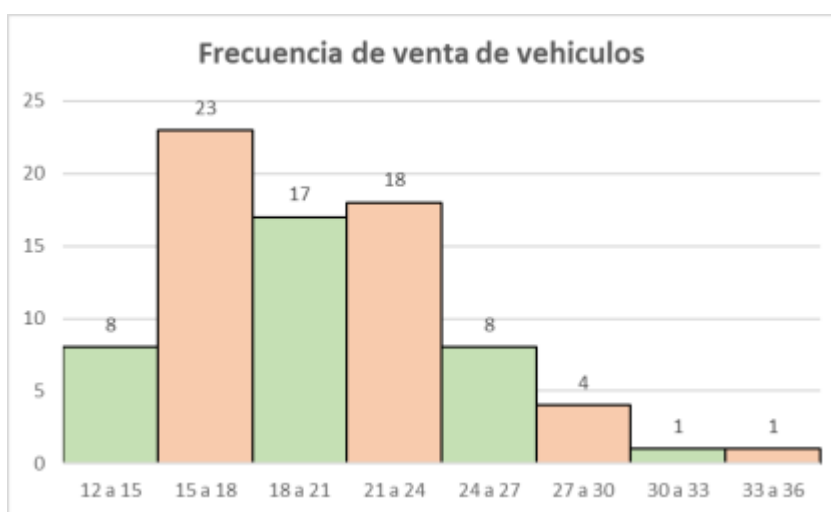
$$\text{Marca de clase} = \frac{12000 + 15000}{2} = 13500$$

Que viene a constituir el más representativo, o típico del precio de venta de los vehículos de esa clase.

Distribución de frecuencias relativas

Las frecuencias de clase relativas se obtienen del cociente entre la frecuencia de cada clase respecto al total de las frecuencias u observaciones. Nos permite visualizar el porcentaje del número de observaciones de cada clase.

Precios de venta (en miles de \$)	Frecuencia	Marca de clase	Frecuencia relativa
12 hasta 15	8	13,5	0,1
15 hasta 18	23	16,5	0,2875
18 hasta 21	17	19,5	0,2125
21 hasta 24	18	22,5	0,225
24 hasta 27	8	25,5	0,1
27 hasta 30	4	28,5	0,05
30 hasta 33	1	31,5	0,0125
33 hasta 36	1	34,5	0,0125
Total	80		1



Ahora que tenemos organizada la información podemos contestar los interrogantes planteados por la Gerencia y otras adicionales.

1. Los precios van desde alrededor de \$12000 a \$36000
2. Los precios de venta se concentran entre \$15000 y \$24000
3. Un total de 58 datos, o sea un 72.5%, de los vehículos se vendieron dentro este intervalo.
4. La concentración más grande está en la clase de \$15000 hasta \$18000. El valor central de la clase se encuentra en \$16500, por lo que decimos que un **precio típico** es de \$16500.
5. Dos de los vehículos se vendieron en \$30000 o más, y 8 se vendieron por menos de \$15000.

EJERCICIO 2. El director de programas de honores en la Western University tiene 16 solicitudes para su admisión el próximo otoño. Las calificaciones de la prueba ACT de los solicitantes son:

27	27	27	28	27	25	25	28
----	----	----	----	----	----	----	----

26	28	26	28	31	30	26	26
----	----	----	----	----	----	----	----

Tales calificaciones deben organizarse en una distribución de frecuencias.

- ¿Cuántas clases recomendaría?
- ¿Qué intervalo de clase sugeriría?
- ¿Cuál es el límite inferior que recomendaría para la primera clase?
- Organice las calificaciones en una distribución de frecuencias y determine la frecuencia de clases relativas.
- Comente acerca de la forma de distribución.

SOLUCION.

- ¿Cuántas clases recomendaría?

Número de clases: $2^k \geq n$

$$2^3 = 8 < 16 \quad 2^4 = 16 \geq 16$$

Recomendación número de clases: $k = 4$

- ¿Qué intervalo de clase sugeriría?

$$i = \frac{\text{Valor más alto} - \text{Valor más bajo}}{\text{Número de clase}}$$

$$i = \frac{31 - 25}{4} = 1.5 \quad \text{Se adopta: } i = 2$$

$$i = \frac{\text{Valor más alto} - \text{Valor menos alto}}{1 + 3.322(\log. \text{del total de frecuencias})}$$

$$i = \frac{31 - 25}{1 + 3.322 * \log 16} = 1.20 \quad \text{Se adopta: } i = 2$$

- ¿Cuál es el límite inferior que recomendaría para la primera clase?

Categoría: LI hasta LS

Recomendación: el límite inferior debe ser: $Li = i * \text{par}$

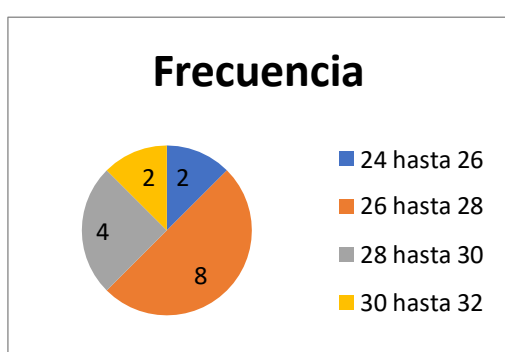
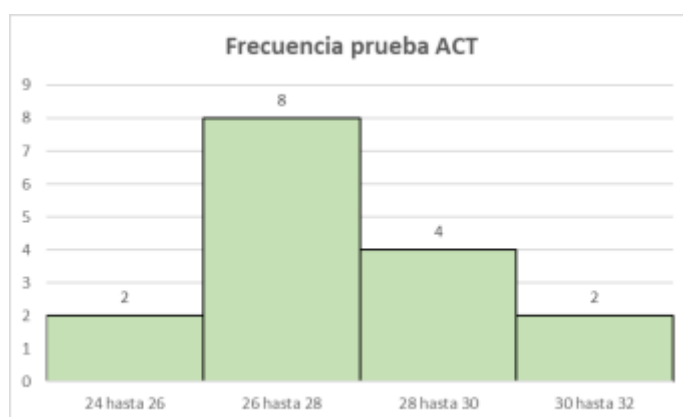
$$Li = 2 * 12 = 24$$

Valor más bajo=25

- d) Organice las calificaciones en una distribución de frecuencias y determine la frecuencia de clases relativas.

Primera categoría: 24 a 24+i

Calificaciones prueba ACT	Frecuencia	Frecuencia relativa	Marca de clase
24 hasta 26	2	0,125	25
26 hasta 28	8	0,5	27
28 hasta 30	4	0,25	29
30 hasta 32	2	0,125	31
	16	1	



- e) Comente acerca de la forma de distribución.

1.7.4 Representaciones de tallo y hoja.

Al organizar los datos en la distribución de frecuencias se consigue un cuadro visual rápido de la forma de la distribución sin hacer cálculos adicionales. Podemos ver dónde se concentran los

datos y determinar si hay valores sumamente grandes o pequeños y su valor típico. Teniendo como desventajas:

- Se pierde la identidad exacta de cada valor
- No estamos seguros de cómo se distribuyen los valores dentro de cada clase.

Analicemos una distribución de frecuencias que muestra el número de anuncios publicitarios adquiridos por los 45 miembros de la asociación de vendedores de automóviles.

Numero de anuncios adquiridos	Frecuencia
80 a 90	2
90 a 100	7
100 a 110	6
110 a 120	9
120 a 130	8
130 a 140	7
140 a 150	3
150 a 160	3

Para explicar lo anotado, analicemos la clase que va de 90 a 100 la frecuencia es 7, no podemos asegurar que datos dentro de ese intervalo es más recurrente, por ejemplo el 95 se repite una o tres de los siete.

Una técnica utilizada para mostrar información cuantitativa en forma condensada es la representación tallo y hoja. Una ventaja de esta representación sobre la distribución de frecuencias es que no perdemos identidad de los valores en la clase.

Supongamos que las observaciones en la clase 90 hasta 100 son: 96, 94, 93, 94, 95, 96 y 97. El tallo es el valor del dígito y se coloca a la izquierda de la barra vertical y las hojas son los dígitos siguientes y se colocan a la derecha así:

Tallo	Hoja
9	6 4 3 4 5 6 7

Luego ordenamos los valores de manera ascendente. Teniendo:

Tallo	Hoja
9	3 4 4 5 6 6 7

Representación de tallo y hoja. Es una técnica estadística para mostrar un conjunto de datos. Cada valor numérico se divide en 2 partes: los dígitos principales se toman como tallo, y el dígito siguiente es la hoja. Los tallos se ubican a lo largo del eje vertical principal y las hojas para cada observación, a lo largo del eje horizontal.

EJEMPLO 4. Una encuesta de llamadas recibidas por una muestra de suscriptores de una compañía telefónica, dio a conocer la siguiente información. Elabore una representación tallo y hoja. Resuma los datos sobre el número de llamadas recibidas. ¿Cuántas llamadas recibió el suscriptor típico? ¿Cuál fue el número más grande y el más pequeño de llamadas recibidas? ¿Alrededor de que valores tendieron a agruparse las llamadas?

52	43	30	38	30	42	12	46	39
37	34	46	32	18	41	05		

1. De la observación de datos tenemos que el menor valor es 5 y el mayor 52
2. Organizar datos en tallo de acuerdo a las recomendaciones.

Tallo	Hojas
0	5
1	2 8
2	
3	0 8 0 9 7 4 2
4	3 2 6 6 1
5	2

Organicemos la información

Tallo	Hojas
0	5
1	2 8
2	
3	0 0 2 4 7 8 9
4	1 2 3 6 6
5	2

Organizada la información conteste las preguntas.

a) ¿Cuántas llamadas recibió el suscriptor típico?

Se estudiaron 16 llamadas telefónicas.

b) ¿Cuál fue el número más grande y el más pequeño de llamadas recibidas?

El número de llamadas telefónicas esta entre 5 y 52

c) ¿Alrededor de que valores tendieron a agruparse las llamadas?

7 de los 16 suscriptores recibieron entre 30 y 39 llamadas.

En la distribución tallo hojas es posible presentar información adicional a la izquierda del tallo que representa valores acumulados, para lo cual sume los datos en forma descendente y ascendente (valores 2, 9, 21, etc); y en el valor que se cruce corresponde al valor medio que representa el valor por abajo del cual se presentaron la mitad de las observaciones.

Ejemplo: Se presenta una distribución tallo-hoja, en un total de 45 observaciones. La propuesta es encontrar el valor medio.

Pasos:

- Sumar los datos en forma descendente, llegamos hasta 15 y en forma ascendente a 21; llegando en los dos casos a una clase de cruce que corresponde a la 11 en donde se encuentra el valor medio.
- El valor medio corresponde al valor medio de datos en nuestro caso al tener 45 datos será alrededor de 23; entonces ese valor corresponde a la observación número 8 de la fila en este caso 118; o si vamos de manera ascendente de a 23 leemos el segundo dato de derecha a izquierda en nuestro caso 118.

Acumulados	Tallo	Hoja
2	8	8 9
9	9	3 4 4 5 6 6 7
15	10	3 3 4 6 7 8
(9)	11	1 2 2 3 3 7 7 8 9
21	12	0 0 4 5 5 5 7 7
13	13	2 4 5 6 8 9 9
6	14	2 3 8
3	15	5 5 6

EJEMPLO: La siguiente representación de tallo y hoja muestra el número de unidades producidas al día en una fábrica.

Acumulado	Tallo	Hoja
1	3	8
1	4	
2	5	6
9	6	0 1 3 3 5 5 9
(7)	7	0 2 3 6 7 7 8
9	8	5 9
7	9	0 0 1 5 6
2	10	3 6

a) ¿Cuántos días se estudiaron?

25

b) ¿Cuántas observaciones están en la primera clase?

1

c) ¿Cuál es el valor más pequeño y cuál es el valor más grande?

38, 106

d) Enumere los valores reales de la cuarta fila

60, 61, 63, 63, 65, 65, 69

e) Enuncie los valores reales de la segunda fila

Sin valores

f) ¿Cuántos valores son menores que 70?

9

g) ¿Cuántos valores son 80 o mayores?

9

h) ¿Cuál es el valor medio?

76

i) ¿Cuántos valores hay entre 60 y 89?

16

1.7.5 Representación gráfica de una distribución de frecuencias

Tres diagramas que representan gráficamente de manera adecuada una distribución de frecuencias son el histograma, el polígono de frecuencias y el polígono de frecuencias acumuladas (ojiva).

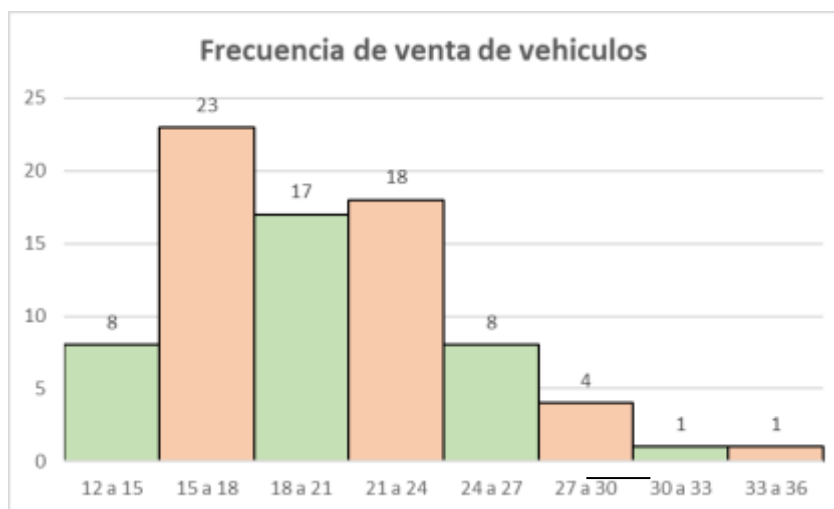
HISTOGRAMA: Es un diagrama en el que se marcan las clases en el eje horizontal y las frecuencias de clase en el eje vertical. Las frecuencias de clase quedan representadas por las alturas de las barras, y estas se colocan adyacentes.

Consideremos la distribución de frecuencias del ejemplo 1, que se describe a continuación.

Precios de venta (en miles de \$)	Frecuencia
12 hasta 15	8
15 hasta 18	23
18 hasta 21	17
21 hasta 24	18
24 hasta 27	8
27 hasta 30	4
30 hasta 33	1
33 hasta 36	1
Total	80

RECOMENDACIONES PARA GRAFICAR

1. Escoja una escala adecuada tanto para las clases como para las frecuencias, las escalas pueden ser diferentes de acuerdo a los datos.
2. La calidad de la imagen que debemos proyectar es importante, escoja colores atractivos para la presentación.



En base al histograma a simple vista podemos realizar algunos comentarios como:

- La frecuencia de clase mayor es la clase de 15000 a 18000. Un total de 23 de los 80 vehículos vendidos se encuentra dentro de ese intervalo de precios.
- La mayor concentración de autos vendidos esta entre 15000 y 24000 en alrededor del $(58/80)$ 72.5% del total de autos vendidos.
- Apenas 2 autos son vendidos entre 30000 a 36000.
- Los precios van desde alrededor de 12000 hasta 360000

Polígono de frecuencias.

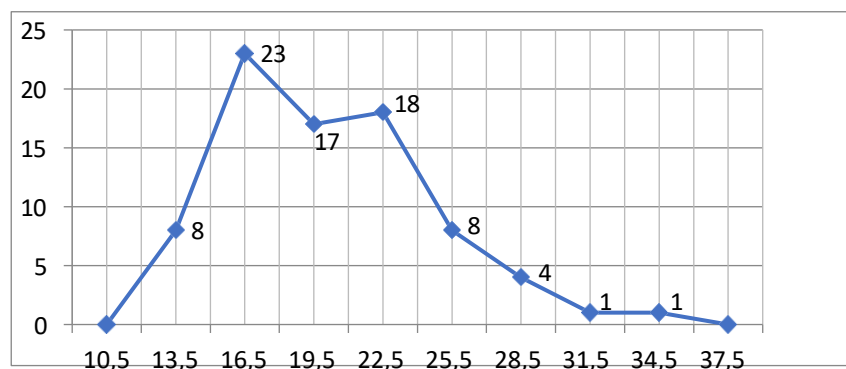
Es similar al histograma, consiste en segmentos de recta que unen los puntos determinados por la intersección de los puntos medios de la clase y las frecuencias de clase.

RECOMENDACIONES PARA GRAFICAR.

1. Escoja una escala adecuada tanto para los puntos medios (marca de clase), así como para las frecuencias.
2. Marque cada punto correspondiente $(13.5, 8)$, $(16.5, 23)$... y así sucesivamente.

3. Se recomienda cerrar el polígono de frecuencias, restando del primer punto medio el intervalo de clase (3) y obtenemos 10.5; también en el extremo sumando al último punto medio (3) y obtenemos 34.5.
4. Una con rectas los puntos obtenidos.

Precios de venta (en miles de \$)	Punto medio (PM)	Frecuencia
12 - 15	13,5	8
15 - 18	16,5	23
18 - 21	19,5	17
21 - 24	22,5	18
24 - 27	25,5	8
27 - 30	28,5	4
30 - 33	31,5	1
33 - 36	34,5	1
Total		80

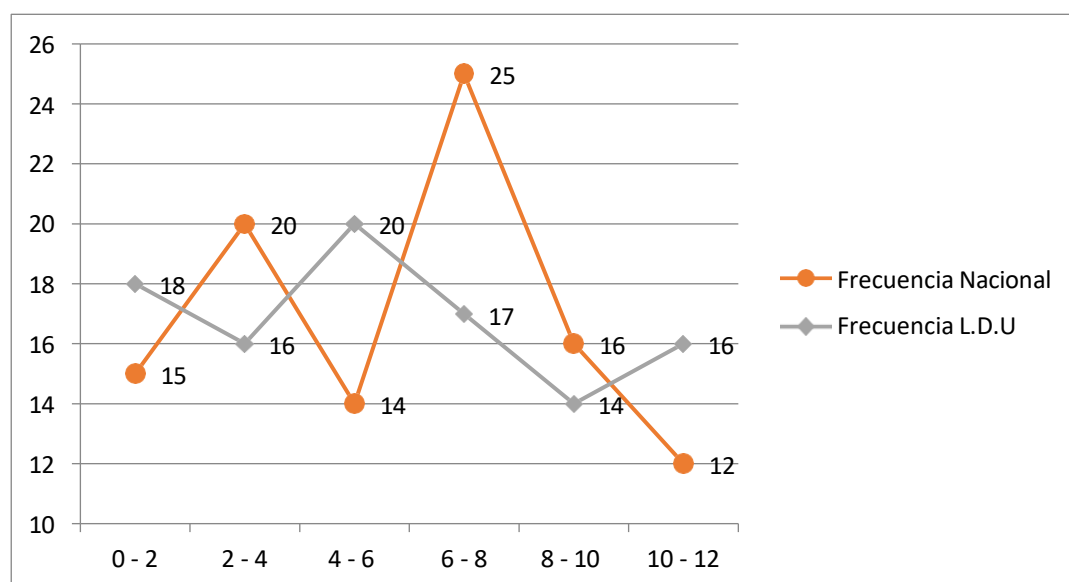


CONCLUSIONES:

- Tanto el histograma como el polígono de frecuencias, nos permiten tener una imagen clara y rápida de las principales características de los datos (altos, bajos, puntos de concentración, etc)
- El histograma presenta la facilidad que permite visualizar cada clase con un rectángulo y su correspondiente frecuencia.
- En cambio, el polígono de frecuencias nos permite realizar comparaciones entre dos o más distribuciones de frecuencias. Para visualizar lo expresado, utilicemos el ejemplo siguiente.

EJEMPLO 5: Supongamos que se ha recolectado información del número de goles anotados por el Nacional y Liga en el año y por cada bimestre obteniendo la siguiente información.

Periodo	PM	Frecuencia Nacional	Frecuencia L.D.U
0 - 2	1	15	18
2 - 4	3	20	16
4 - 6	5	14	20
6 - 8	7	25	17
8 - 10	9	16	14
10 - 12	11	12	16
Total		102	101



Polígono de frecuencias acumuladas “menor que” “ojiva”

Algunos interrogantes a la serie de datos y distribuciones de frecuencias estudiadas, se hacen más evidentes si realizamos una distribución de frecuencias acumuladas y mostrándola gráficamente en un polígono de frecuencias acumuladas.

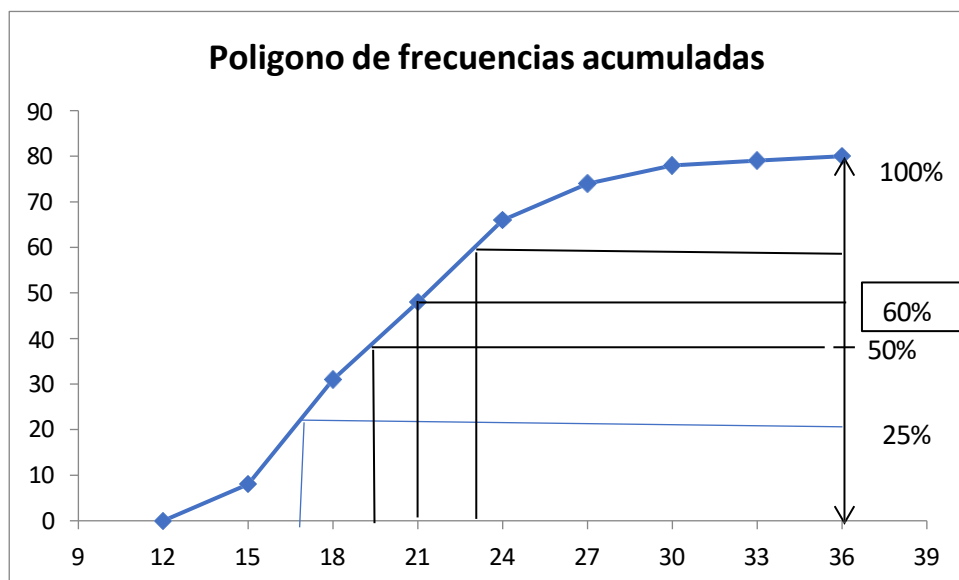
Utilicemos para el análisis el ejemplo 1, con su respectiva distribución de frecuencias y contestemos las preguntas siguientes:

- ¿Un 50% de los vehículos se vendieron por menos de qué cantidad?
- ¿Veinticinco de los vehículos se vendieron por menos de qué cantidad?

Para construir la tabla de frecuencias acumuladas a la primera sumar la siguiente y ese resultado adicionar la siguiente y así sucesivamente.

Para construir el diagrama de frecuencias acumuladas menor que, localice el límite superior de la clase en el eje x y las frecuencias acumuladas en el eje y, teniendo los puntos siguientes (0, 12) (15, 8) (18, 31).....

Precios de venta (en miles de \$)	Frecuencia	Frecuencia acumulada
12 -15	8	8
15 - 18	23	31
18 - 21	17	48
21 - 24	18	66
24 - 27	8	74
27 - 30	4	78
30 - 33	1	79
33 - 36	1	80
Total	80	



Para graficar realice el mismo procedimiento que el polígono de frecuencias, adicionalmente a la derecha del grafico establezca una escala adicional con %, siendo 0 el 0% y 80 el 100%, gradúe la escala.

Solución.

- ¿Un 50% de los vehículos se vendieron por menos de qué cantidad?

Trace una línea horizontal tomando como referencia la escala de la derecha, intercepte la curva y proyecte ese punto al eje horizontal, se lee aproximadamente 19.5; por lo que se estima que el 50% de vehículos se vendieron por menos de \$19.500.

- ¿Veinticinco de los vehículos se vendieron por menos de qué cantidad?

Estime el valor de 25 en el eje vertical de la izquierda, intercepte la curva y proyecte al eje horizontal, se lee aproximadamente 17.5; se estima que 25 vehículos se vendieron en menos de \$17.500

- ¿Cuál es el % de vehículos que se vendieron por menos de \$26.000?

Ubique en el eje horizontal 26, proyecte con una línea vertical e intercepte la curva, proyecte con una línea vertical a la escala de la derecha, el valor aproximado está en alrededor del 88%; se estima que el 88% de los vehículos se vendieron menos que \$26.000.

1.7.6 Otras representaciones gráficas.

Diagramas lineales:

Estas representaciones son efectivas en negocios porque puede mostrarse el cambio de una variable a través del tiempo. La variable se la ubica en el eje vertical y el tiempo en el eje horizontal.

A menudo, en la misma grafica se trazan dos o más series de cifras. Por lo tanto, una gráfica puede mostrar la tendencia de varias series.

EJEMPLO: Se puede analizar la tasa de interés promedio de un país y la inflación correspondiente como los expuestos en la siguiente tabla y su correspondiente grafico lineal.

Periodo	Tasa de interés	Inflación
1995	12,50%	5%
1996	13,50%	4,50%
1997	14,90%	5,30%
1998	16,50%	6,90%
1999	13,70%	5,30%
2000	14,90%	6,20%

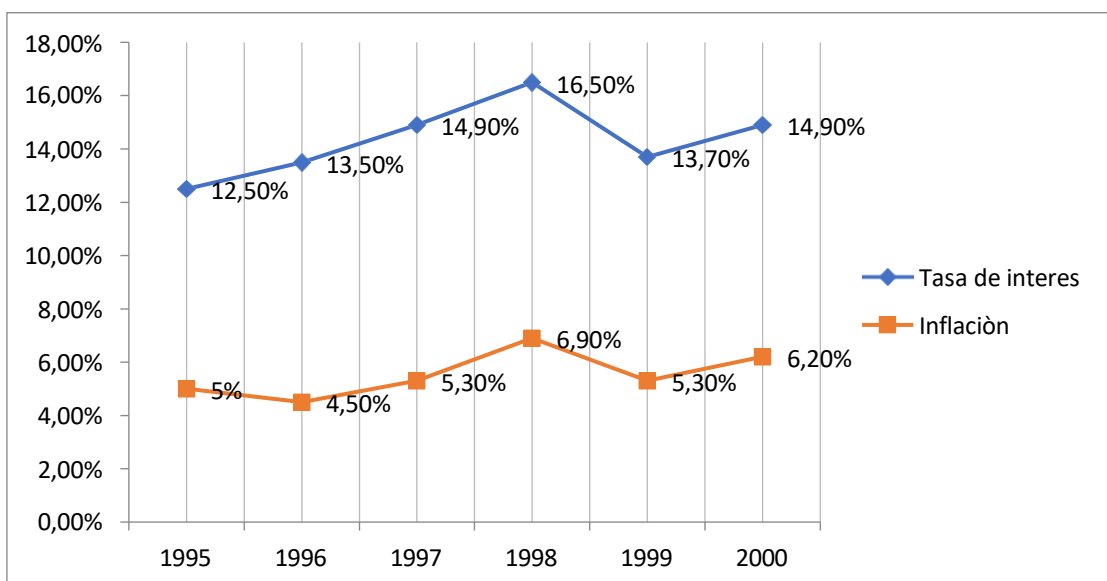


Diagrama de barras

Puede ser utilizado para representar cualquiera de los niveles de medición y pueden ser de barras horizontales y verticales; para lo cual revisemos los ejemplos siguientes:

Barras horizontales:

Ejemplo: El Municipio del D.M. de Quito en el año 2012, determino que para ejercer lícitamente la profesión debía adquirir o una inscripción de la patente municipal, teniendo como resultado la siguiente tabla:

OCUPACIÓN	% REGISTRADOS
MEDICOS	27%
ABOGADOS	15%
ING. CIVILES	12%
ING. COMERCIALES	18%
AUDITORES	14%
ARQUITECTOS	8%
OTROS	6%
	100%



Barras verticales:

Ejemplo: Se realiza una investigación de preferencia de los estudiantes de **UCE** acerca de la simpatía de los clubes deportivos que intervienen en el campeonato nacional, encontrándose la información siguiente:

Equipo Deportivo	Preferencia
LDU	800
Deportivo Quito	430
Nacional	120
Barcelona	500
Emelec	70
Otros	45
TOTAL	1965

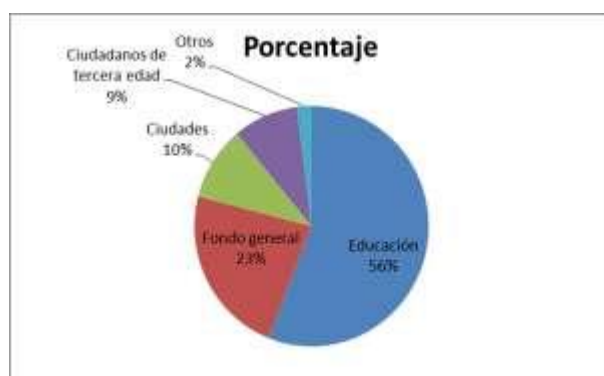


La diferencia con el histograma en su representación es que las barras no se encuentran unidas.

Grafica circular.

Es especialmente útil para mostrar datos de nivel ordinal, para lo cual utilizaremos los datos de las utilidades de una lotería estatal de EEUU.

Uso de las utilidades	Porcentaje
Educación	56%
Fondo general	23%
Ciudades	10%
Ciudadanos de tercera edad	9%
Otros	2%
Total	100%



CONCLUSIONES:

De todos los diagramas mostrados se pueden obtener conclusiones a simple vista, como por ejemplo en diagrama circular o de pastel podríamos decir que en educación se invierte en una cantidad en más del doble que su fondo general y otras, que nos servirían para crear políticas del uso de utilidades.

2. UNIDAD II

Medidas de tendencia central y números índices



2. Medidas de tendencia central

Hemos visto que la Estadística Descriptiva permite a través de una distribución de frecuencias organizar datos no agrupados o en vivo y su correspondiente representación a través de histogramas, polígono de frecuencias y frecuencias acumulados, que nos han permitido realizar observaciones de los datos mayores, menores y su valor medio.

Ahora desarrollaremos métodos para describir datos encontrando **un único valor** para describir un conjunto de ellos. Este único valor se le conoce como medida de tendencia central, por lo común se le denomina promedio.

Las medidas de tendencia central son estadísticas descriptivas utilizadas para resumir o representar el valor central o típico de un conjunto de datos. Estas medidas proporcionan información sobre la ubicación de los datos y son fundamentales para comprender la distribución de los mismos.

Estas medidas de tendencia central son fundamentales en estadística y son útiles para resumir y comprender la estructura de un conjunto de datos. Dependiendo de la naturaleza de los datos y el objetivo del análisis, una de estas medidas puede ser más apropiada que las otras.

Medida de Tendencia Central: un único valor que resume un conjunto de datos. Señala al centro de los valores.

2.1 Media de la población

A partir de los datos en vivo, los que no han sido agrupados en una distribución de frecuencias o en una representación de tallo y hoja; la media de una población está dada por:

$$\text{Media de una poblacion} = \frac{\text{Suma de todos los valores de la poblacion}}{\text{Numero de valores en la poblacion}}$$

$$\mu = \frac{\sum X}{N}$$

μ representa la media de la población. Letra griega

N es el numero total de elementos en la poblacion

X representa cualquier valor en particular

Cualquier característica medible de una población se denomina parámetro. La media de una población es un parámetro, así como la amplitud de variación (diferencia entre el valor más alto y más bajo de un conjunto de datos).

Parámetro: una característica de la población.

2.2 Media de la muestra

Se ha definido lo que es una muestra de una población y el porqué de su utilización.

$$\text{Media de una muestra} = \frac{\text{Suma de todos los valores de la muestra}}{\text{Numero de valores en la muestra}}$$

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

$\bar{X}$ representa la media de la muestra. Letra latina

n es el numero total de elementos en la muestra

X representa cualquier valor en particular

La media de una muestra o cualquier otra medida basada en datos muestrales, se denomina un dato estadístico.

Dato estadístico: Una característica de una muestra.

EJEMPLO: La empresa “MLGF” se especializa en tratos a largo plazo de países extranjeros. Interesa saber la tasa de interés de estos acuerdos financieros. Una muestra aleatoria de seis bonos presento lo siguiente:

Articulo	Tasa de interés
Bonos del Gobierno de Australia	9,50%
Bonos del Gobierno de Bélgica	7,25
Bonos del Gobierno de Canadá	6,5
Bonos del Gobierno de Francia	4,75
Bonos del Gobierno de Italia	12
Bonos del Gobierno de España	8,3

¿Cuál es la media de las tasas de interés en esta muestra de tratos de finanzas a largo plazo?

Solución: Como se trata de una muestra se puede utilizar la fórmula de media muestra:

$$\bar{X} = \frac{\sum X}{n} \quad \bar{X} = \frac{9.50 + 7.25 + 6.5 + 4.75 + 12 + 8.3}{6} = 8.05\%$$

La tasa de interés media (aritmética) de la muestra es igual al valor de 8.05%

Propiedades de la media aritmética

1. Todo conjunto de datos de nivel de intervalo y de nivel de razón tiene un valor medio.
2. Al evaluar la media se incluyen todos los valores.
3. Un conjunto de datos solo tiene una media. Este es un valor único.
4. La media es una medida muy útil para comparar dos o más poblaciones. Ejemplo se puede comparar el promedio de calificaciones en el primer hemisemestre de la asignatura de Estadística Descriptiva de los paralelos A y B de UCE.
5. La media es la única medida de ubicación donde la suma de las desviaciones de cada valor respecto a la media siempre será cero.

LAS DESVIACIONES RESPECTO A LA MEDIA SUMAN CERO: $\sum(X - \bar{X}) = 0$

$$\begin{aligned} \sum(X - \bar{X}) &= (9.50 - 8.05) + (7.25 - 8.05) + (6.50 - 8.05) + (4.75 - 8.05) + \\ &(12.00 - 8.05) + (8.30 - 8.05) = 0 \end{aligned}$$

Por lo tanto, se puede considerar la media como un punto de equilibrio para un conjunto de datos.

Sin embargo, la media tiene las siguientes desventajas:

1. Si uno o dos datos es muy grande o pequeño; por ejemplo, se toma la estatura a 5 estudiantes de la UCE y se tiene los siguientes datos: 1.65, 1.70, 1.72, 1.63, 2.15; su media es de $\bar{X} = 1.77cm$; nótese que el valor de estatura de 2.15 afecta considerablemente a la media pues la mayoría de los datos no están dentro de esa media.
2. La media es inadecuada si hay clase de extremos abiertos para datos agrupados en una distribución de frecuencias.

Categorías	Frec.
15 a 20	6
20 a 25	10
Mayor de 25	15

2.2.1 Media ponderada.

Es un caso especial de la media aritmética; se presenta cuando hay varias observaciones del mismo valor que pueden ocurrir si los datos se han agrupado en una distribución de frecuencias.

$$\text{MEDIA PONDERADA } \bar{X}_w = \frac{\sum(wX)}{\sum w}$$

EJEMPLO: La constructora “CCT” paga a sus empleados \$6.50, \$7.50, o bien \$8.50 por hora. Hay 26 empleados por hora; 14 reciben la tarifa de \$6.50, 10 la de \$7.50 y 2 la de \$8.50. ¿Cuál es la media de la tarifa por hora pagada a los 26 trabajadores?

$$\bar{X}_w = \frac{14 * 6.50 + 10 * 7.50 + 2 * 8.50}{26} = \$7.04$$

La media ponderada de los sueldos por hora es \$7.04

2.3 Mediana

Se señaló que para datos que contienen uno o dos valores sumamente grandes o pequeños, la media aritmética puede no ser representativa. El punto central de tales datos puede ser descrito utilizando como medida de tendencia central la mediana.

Mediana, es el punto medio de los valores después de ordenarlos de menor a mayor, o de mayor a menor. Se tiene que el 50% de las observaciones se encuentran por arriba de la mediana y 50% por debajo de ella.

Para aclarar lo expuesto; supongamos que el precio de ciertos departamentos es de \$60.000; \$65.000; \$70.000; \$80.000 y un penthouse de super lujo cuesta \$275.000. La media aritmética es de \$110.000, pero la media aritmética está haciendo que los precios se inclinen hacia arriba, por lo que es un promedio no representativo. Parecería ser que un precio entre \$65.000 y \$75.000 es un promedio más típico y representativo, y de hecho lo es. En casos, como este la mediana proporciona una medida más exacta de tendencia central.

Para determinar esta medida ordenemos los datos de manera ascendente y descendente.

Precios de menor a mayor		Precios de mayor a menor
\$ 60.000		\$ 275.000
65000		80000
70000	Mediana	70000
80000		65000
275000		60000

Obsérvese que hay el mismo número de precios por debajo de \$70.000 como por arriba. Por lo tanto, la mediana no se ve afectada por observaciones muy bajas o muy altas.

Cuando se tiene un **número par de observaciones**, la mediana la encontramos de manera similar al del número impar de observaciones; esto es ordenamos de mayor a menor o de menor a mayor y escogemos el promedio de los dos valores centrales; no pudiendo ser algún valor de las observaciones dadas.

EJEMPLO: Se presentan la edad de 10 personas en la tienda de videos en un Centro comercial a las 10 a.m. Determine la mediana.

12 8 17 6 11 14 8 17 10 8

$$10/2=5$$

SOLUCION: Ordenamos de menor a mayor y tenemos:

6

8

8

8

10

11

12

14

17

17

$$\text{Mediana} = (10+11)/2 = 10.5 \text{ años}$$

Propiedades de la mediana:

1. Es única, solo existe una mediana para un conjunto de datos.
2. No se ve afectada por valores muy grandes o muy pequeños.
3. Puede calcularse para una distribución de frecuencias con una clase de extremo abierto.
4. Puede utilizarse para datos de nivel de razón, de intervalo y ordinal.

2.4 Moda

Es el valor de la observación que aparece con mayor frecuencia.

En el caso del ejemplo anterior la moda es 8; pues es el que se repite 3 veces.

Propiedades de la moda:

1. Puede utilizarse para niveles de datos nominal, ordinal, de intervalo y razón.
2. No se ve afectada por valores muy altos o muy pequeños.
3. Puede utilizarse para distribuciones con clases de extremo abierto.

Tiene algunas desventajas, por lo que es **menos utilizada** con menos frecuencia que la media o la mediana, podemos citar:

1. Para muchos conjuntos de datos no hay moda; por ejemplo: Para las observaciones de datos: **2, 4, 6, 7, 9**.
2. Por lo contrario, para otros conjuntos de datos, esta medida no es única; ejemplo: Para la observación de las edades de un conjunto de personas se tiene: 24, **25, 25**, 26, 28, **31, 31**; tanto las edades de 25 y 31 son modas: por tanto, esta agrupación de edades se conoce como bimodal. Es cuestionable utilizar 2 modas para representar la tendencia central de este conjunto de datos de edades.

2.5 Media geométrica

La media geométrica es útil para encontrar el promedio de porcentajes, razones, índices o tasas de crecimiento. Muy utilizada en los negocios, porque usualmente interesa encontrar el cambio porcentual en las ventas, sueldos y cifras económicas como PNB.

MEDIA GEOMETRICA $MG = \sqrt[n]{(x_1)(x_2) \dots (x_n)}$

La media geométrica siempre será menor o igual que la media aritmética y el valor del conjunto de datos siempre debe ser positivos.

EJEMPLO: Suponga que recibe un aumento del 8% de su sueldo este año, y recibirá uno de 16% el próximo año. Encuentre la media geométrica.

Suposición: $10 + 0.08 * 10 = 10(1 + 0.08)$

$$MG = \sqrt{(1.08)(1.16)} = 1.1193$$

Lo anterior podemos verificar con el ejemplo siguiente: Supongamos, que el salario actual es de \$2.000; tendríamos:

Aumento 1: $\$2.0000 * 1.08 = \$2.160$

Aumento 2: $\$2.160 * 1.16 = \$2.505,60$

Que es equivalente:

$$\$2.000 * 1.1193 == \$2.238,60$$

$$\$2.238,60 * 1.1193 = \$2.505,66$$

EJEMPLO: Las ganancias obtenidas por una Consultora de Auditoria en cuatro proyectos son: 12%, 14%, 16%, 21%. ¿Cuál es la media geométrica de la ganancia?

$$MG = \sqrt[n]{(x_1)(x_2) \dots (x_n)}$$

$$MG = \sqrt[4]{(12)(14)(16)(21)} = 15.41\% \quad < \quad X = 15.75\%$$

La media aritmética: $X = \frac{12+14+16+21}{4} = 15.75\%$

$MG < X$ "Valor más conservador"

Otra aplicación de la media geométrica es encontrar un aumento porcentual promedio en un intervalo de tiempo.

2.5.1 Media geométrica en un promedio en un periodo dado

$$MG = \sqrt[n]{\frac{\text{Valor al final del periodo}}{\text{Valor al principio del periodo}}} - 1$$

EJEMPLO: Supóngase que el sueldo mensual de un trabajador de la construcción ha cambiado de \$250 en el año 2.000 a \$480 en el 2.013. ¿Cuál es la tasa del incremento porcentual anual promedio para el periodo?

SOLUCION: Hay 13 años de diferencia por lo tanto $n = 13$.

$$MG = \sqrt[13]{\frac{480}{250}} - 1 = 0.0515$$

La tasa de aumento anual es de 5.15% anual para el indicado periodo.

2.6 Media mediana y moda para datos agrupados

2.6.1 Media

Para evaluar la media (aritmética) de datos organizados en una distribución de frecuencias, suponemos que las observaciones en cada clase son representadas por el punto medio de esta.

Media aritmética de datos agrupados:

$$\bar{x} = \frac{\sum(f * X)}{n}$$

Consideremos el ejemplo anterior de la venta de autos:

Precios de venta (en miles de \$)	Frecuencia (f)
12 a 15	8
15 a 18	23
18 a 21	17
21a 24	18
24 a 27	8
27 a 30	4
30 a 33	1
33 a 36	1
Total	80

Precios de venta (en miles de \$)	Frecuencia (f)	Punto medio (X)	f*X
12 a 15	8	13,5	108
15 a 18	23	16,5	379,5
18 a 21	17	19,5	331,5
21a 24	18	22,5	405
24 a 27	8	25,5	204
27 a 30	4	28,5	114
30 a 33	1	31,5	31,5
33 a 36	1	34,5	34,5
Total	80		1608

$$\bar{x} = \frac{\sum fX}{n} = \frac{1608}{80} = 20.1$$

Se concluye que la media del precio de venta de los vehículos es aproximadamente de \$20.100

2.6.2 Mediana

La mediana se define como el valor abajo del cual se encuentran la mitad de los valores, y arriba del cual se halla la otra mitad. Una vez que los datos en vivo (originales) se han organizado en una distribución de frecuencias, parte de la información no es identificable; por lo tanto, no es posible identificar la mediana exacta.

Sin embargo, puede estimarse de la forma siguiente:

1. Localizando la clase donde está la mediana y;
2. Interpolando dentro de esa clase para obtener el valor.

La razón para este enfoque es que los elementos de la clase en que se encuentra la mediana están espaciados de manera uniforme por toda la clase. La fórmula es:

$$Mediana = Li + \frac{\frac{n}{2} - FA}{f}(i)$$

Li es el límite inferior de la clase que contiene a la mediana

n es el número total de frecuencias

f es la frecuencia de la clase antes mencionada

FA es el numero acumulativo de frecuencias en todas que preceden inmediatamente a la clase en cuestión (con la mediana)

i es el ancho de la clase en que se encuentra la mediana

Consideremos el ejemplo de la venta de autos para determinar la media de los datos agrupados. ¿Cuál es la mediana del precio de venta?

Precios de venta (en miles de \$)	Frecuencia f	Frecuencia acumulada FA
12 a 15	8	8
15 a 18	23	31
18 a 21	17	48
21 a 24	18	66
24 a 27	8	74
27 a 30	4	78
30 a 33	1	79
33 a 36	1	80
Total	80	



Escriba aquí la ecuación.

$$80/2=40$$

$$Mediana = Li + \frac{\frac{n}{2} - FA}{f}(i)$$

$$Mediana = 18 + \frac{\frac{80}{2} - 31}{17}(3) = \$19,59 \quad Mediana = \$19 590$$

SOLUCION: Pasos a seguir.

1. Determinamos la frecuencia acumulada (FA)
2. Ubicamos la clase en la que se encuentra la mediana de los datos; $(n + 1)/2$ por ser número par de observaciones; pero es suficiente con encontrar con $n/2$; en nuestro caso $\frac{80}{2} = 40$. La observación 40 se encuentra en la clase 18 a 21.
3. Además, la observación 40 se encuentra en el intervalo de 31 a 48; entonces suponiendo que los precios de venta se distribuyen uniformemente entre 18 a 21 tenemos, utilizando la interpolación lineal:

$$= \frac{9}{17}(\$3000) = \$1588$$

$$Mediana = 18.000 + 1.588 = \$19.588$$

La mediana estimada del precio de venta de los vehículos es de \$19.588

Podemos encontrar la mediana de datos agrupados utilizando la formula descrita anteriormente; donde:

$$n = 80$$

$FA = 31$ *numero acumulativo de precios que preceden a la clase de la mediana*

$i = \$3000$ *intervalo de clase*

$$Mediana = L + \frac{\frac{n}{2} - FA}{f}(i)$$

$$Mediana = \$18.000 + \frac{\frac{80}{2} - 31}{17}(\$3.000) = \$19.588$$

OBSERVACION:

- La consideración en que se basa la aproximación de la mediana; que las frecuencias en la clase que la contiene se distribuyen en forma uniforme \$18.000 a \$21.000 puede no ser correcta. Por tanto, es más seguro decir que aproximadamente la mitad de precios de venta son menores de \$19.588 y la otra es mayor.
- Los valores calculados de media y mediana de datos agrupados pueden diferir de los valores calculados para los mismos para datos en vivo, lo que podemos realizarlos utilizando Excel, obteniéndose los resultados siguientes:

Precios de venta de vehículos vendidos en cierto mes									
20197	20372	17454	20591	23651	24453	14266	15021	25683	27872
16587	20169	32851	16251	17047	21285	21324	21609	25670	12546
12935	16873	22251	22277	25034	21533	24443	16889	17004	14357
17155	16688	20657	23613	17895	17203	20765	22783	23661	29277
17642	18981	21052	22799	12794	15263	33625	14399	14968	17356
18442	18722	16331	19817	16766	17633	17962	19845	23285	24896
26076	29492	15890	18740	19374	21571	22449	25337	17642	20613
21220	27655	19442	14891	17818	23237	17445	18556	18639	21296

Mediana	19831	19588
Media	20227	20100

- Los valores obtenidos, como se muestran en el resumen, no difieren en más del 1% y de 0.67%; por lo que podemos concluir que los métodos analizados para su cálculo son totalmente válidos.

2.6.3 Moda

La moda es el valor que ocurre con mayor frecuencia. Para datos agrupados en una distribución de frecuencias, es posible aproximar la moda como el punto medio de la clase que contiene el mayor número de frecuencias.

EJEMPLO: Las ventas netas de una muestra de pequeñas plantas de estampado se organizaron en la siguiente tabla de distribución de frecuencias. ¿Cuál es la moda y mediana?

Ventas netas (millones de \$)	Porcentaje del total
1 a 4	13
4 a 7	14
7 a 10	40
10 a 13	23
13 y mayores	10
Total	100

SOLUCION: Para el ejemplo seleccionamos la clase que tiene mayor frecuencia esto es: 7 a 10 cuya frecuencia es de 40. La moda estará representada por el punto medio que es 8.5 millones de dólares. Esto indica que las plantas de estampado tuvieron ventas netas de 8.5 millones de dólares más que otros valores.

Cuando un conjunto de datos tiene más de dos modas se les llama bimodales, y si tiene más de dos modas se llama multimodales. Esto ocurre cuando el conjunto de datos no es homogéneo. Por ejemplo, en una fábrica en la que se reclute personal nuevo y se mantenga personal antiguo. Tenemos 22, 25, 31, 31, 31, 35, 36, 53, 60, 60,60.

Ventas netas (millones de \$)	Porcentaje del total	FA
----------------------------------	-------------------------	----

1 a 4	13	13
4 a 7	14	27
7 a 10	40	67
10 a 13	23	90
13 y mayores	10	100
Total	100	

Clase donde se encuentra la mediana: $\frac{100}{2} = 50$ (7 a 10)

$$\text{Mediana} = Li + \frac{\frac{n}{2} - FA}{f}(i)$$

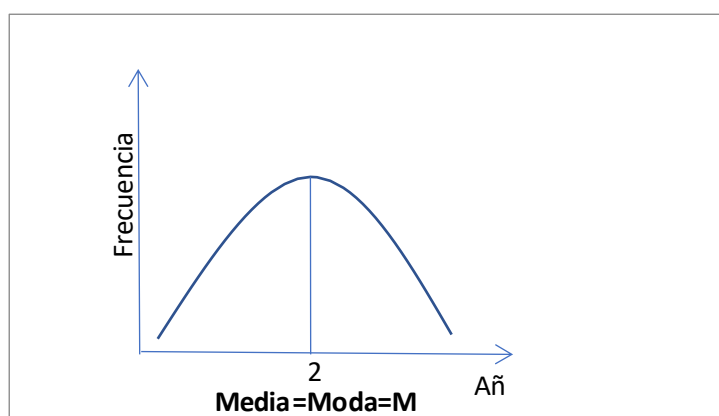
$$\frac{100}{2} - 27$$

$$\text{Mediana} = 7 + \frac{2}{40}(3) = \$8.725 \text{ en millones de \$}$$

$$\text{Moda} = \$8.5 \text{ en millones}$$

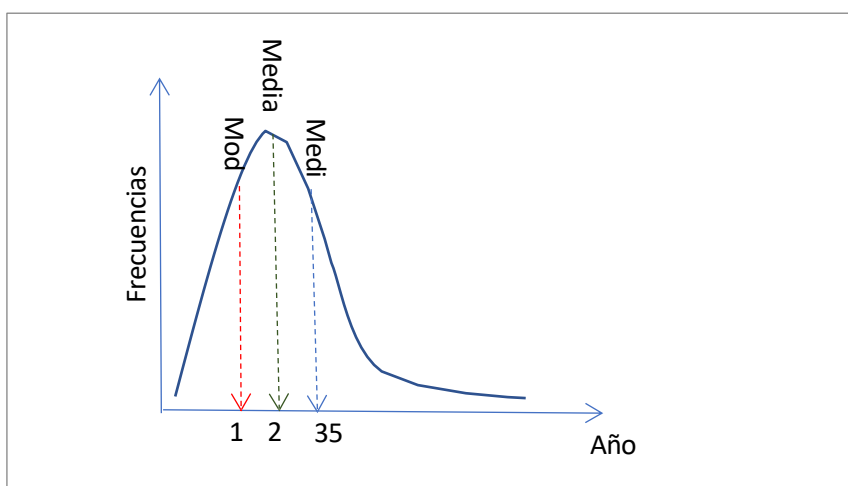
2.6.4 Selección de un promedio para datos de una distribución de frecuencias.

El polígono de frecuencias es simétrico, significa que la distribución de frecuencias tiene la misma forma a ambos lados del centro. Para esta distribución la moda, la mediana y la media se localizan al centro y son iguales.



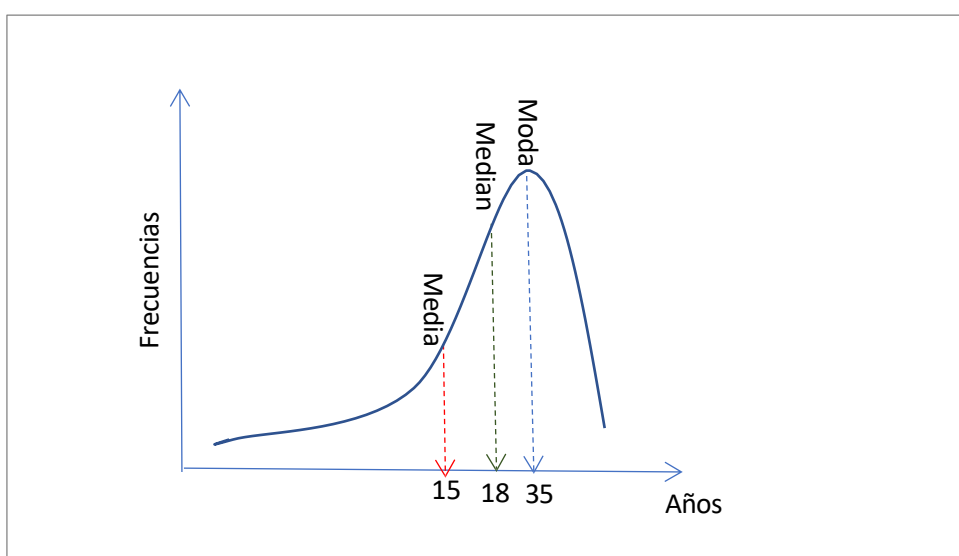
Para el caso de la figura cualquiera de las medidas de tendencia central son las adecuadas para representar esta distribución de frecuencias.

Cuando la distribución de frecuencias se vuelve **asimétrica o sesgada**, la relación entre los promedios cambia. Es una **distribución con asimetría positiva**, la media aritmética es el mayor de los tres promedios. Esto ocurre porque en la media influyen valores sumamente altos, más que en los otros dos promedios. Por lo general, el siguiente promedio mayor es la mediana y la moda es el menor de los tres.



En este caso la media no sería un promedio útil; la mediana y la moda son más representativas.

En una **distribución negativamente asimétrica**, la media es el menor de los tres promedios, que se ve afectada por valores muy pequeños; la mediana es mayor que la media aritmética, y la moda es el mayor de los promedios. En este caso al igual que el anterior la media no debe ser utilizada para representar los datos.



2.7 Medidas de dispersión

Un promedio, como la media y la mediana, solo localizan el centro de los datos y es importante desde un punto de vista, pero no dice nada acerca de la diseminación de los datos. Por ejemplo, si una guía de información turística nos indica que la profundidad promedio de un río es de 1.80 metros. ¿Lo cruzaría sin información adicional? Probablemente no, sin una información adicional; porque la profundidad podría variar de 1.60 a 2.00 metros o podría suceder que varíe de 0.40 a 3.20. Esta información adicional es la que nos permite, tomar decisiones adecuadas.

En estadística, las medidas de dispersión, también conocidas como medidas de variabilidad o medidas de dispersión de datos, son estadísticas que proporcionan información sobre cuánto se extienden o dispersan los valores en un conjunto de datos. Estas medidas ayudan a comprender

la variabilidad de los datos y a evaluar qué tan dispersos o agrupados están alrededor de la media o valor central.

Un valor pequeño para una medida de dispersión indica que los datos se encuentran acumulados cercanamente alrededor de la media aritmética; por lo tanto, esta se considera representativa de los datos, por lo tanto, este valor es un promedio confiable. Lo contrario un valor alto de una medida de dispersión, indica que la media es poco representativa del conjunto de datos.

MEDIDAS DE DISPERSION.

Se consideran varias medidas de dispersión: La amplitud de variación o rango, la desviación media, la varianza y la desviación estándar se basan en desviaciones respecto a la media.

2.7.1 Amplitud de variación o rango

La medida de dispersión más sencilla. Se trata de la diferencia entre el valor más grande y el más pequeño de un conjunto de datos.

$$\text{Amplitud de Variación o Rango} = \text{Valor más grande} - \text{Valor más pequeño}$$

2.7.2 Desviación media

El defecto de la amplitud es que solo se basa en dos valores el más grande y el más pequeño, no considera todos los datos. La desviación media si lo hace y mide el valor promedio en los que varían los valores de una población o muestra respecto a su media.

Desviación Media: Es la media aritmética de los valores absolutos de las **desviaciones respecto a su media aritmética.**

$$DM = \frac{\sum |X - \bar{X}|}{n}$$

X es el valor de cada observación

$\bar{X}$ media aritmetica de los valores

n es el numero de observaciones de la muestra

EJEMPLO: El número de pacientes atendidos en la sala de urgencias en un hospital XXX para una muestra de 5 días el año pasado fueron: 103, 97, 101, 106 y 103. Determine la amplitud de variación y la desviación media.

$$\text{Amplitud de Variación o rango} = 106 - 97 = 9$$

Numero de datos	Número de casos	$ X - \bar{X} $
1	103	1
2	97	5
3	101	1
4	106	4
5	103	1
TOTAL	510	12

$$\bar{X} = \frac{510}{5} = 102$$

$$DM = \frac{\sum |X - \bar{X}|}{n} = \frac{12}{5} = 2.4$$

La desviación media es 2.4 pacientes por día. El número de esta varia, en promedio, en 2.4 pacientes por día respecto de la media 102 enfermos por día.

CONCLUSIONES:

- La desviación media tiene dos ventajas. Primero utiliza todos los valores de la muestra. Segundo, es fácil comprender, pues es el promedio en que los valores se desvían respecto a la media.
- Su desventaja es el uso de valores absolutos.
- La desviación media no se usa frecuentemente como otras medidas de dispersión, tales como la desviación estándar.

2.7.3 Varianza y desviación estándar

La varianza y la desviación estándar se basan en las desviaciones respecto a la media.

VARIANZA: La media aritmética de las desviaciones cuadráticas con respecto a la media.

DESVIACION ESTANDAR. La raíz cuadrada positiva de la varianza.

Varianza poblacional

Las fórmulas para la varianza poblacional y muestral difieren un poco. La varianza poblacional para datos no agrupados, o para datos no tabulados en una distribución de frecuencias, se obtiene:

$$\sigma^2 = \frac{\sum (X - \mu)^2}{N}$$

σ^2 es el símbolo para la varianza de la población. Se expresa comúnmente como “sigma cuadrada”

X valor de una observación en la población.

μ Media aritmética de la población.

N número total de observaciones de la población.

EJEMPLO: El reporte de la empresa Dennis Industries dio los siguientes rendimientos de capital para los accionistas, en el mismo periodo de cinco años (en %): 13.2, 5.0, 10.2, 17.5, y 12.9

a) ¿Cuál es la media aritmética de los valores?

b) ¿La varianza?

SOLUCION:

Datos	Rendimiento capital (%) X	X-μ	(X-μ) ²
1	13,20	1,44	2,07
2	5,00	-6,76	45,70
3	10,20	-1,56	2,43
4	17,50	5,74	32,95
5	12,90	1,14	1,30
TOTAL	58,80		84,45

Media:	11,76
Varianza:	16,89

$$\mu = \frac{\sum X}{N} = \frac{58.80}{5} = 11.76\%$$

$$\sigma^2 = \frac{\sum (X - \mu)^2}{N} = \frac{84.45}{5} = 16.89 \%^2$$

Desviación estándar poblacional

El valor obtenido de la varianza $\sigma^2 = 16.89\%^2$ resulta difícil de interpretar, no se encuentra en términos de porcentaje, sino de porcentaje al cuadrado.

Esta dificultad, obtenemos como la raíz cuadrada de la varianza. Tenemos:

$$\sigma = \sqrt{\frac{\sum (X - \mu)^2}{N}}$$

EJEMPLO: Del caso anterior, se tiene:

$$\sigma = \sqrt{16.89} = 4.11\% . \text{ Este resultado ya lo tenemos como \% y podemos interpretar.}$$

VARIANZA MUESTRAL: La fórmula de la varianza muestral, cambia:

Varianza muestral, formula conceptual:

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

s^2 representa la varianza muestral

- X observaciones en la muestra
- $\bar{X}$ media de la muestra
- n total de observaciones en la muestra

Puede demostrarse que:

$$\sum (X - \bar{X})^2 = \sum X^2 - \frac{(\sum X)^2}{n}$$

Varianza muestral, formula operativa:

$$s^2 = \frac{\sum X^2 - \frac{(\sum X)^2}{n}}{n - 1}$$

Desviación estándar muestral

Es un estimador de la desviación estándar poblacional. Y es la raíz cuadrada de la varianza muestral.

$$s = \sqrt{\frac{\sum X^2 - \frac{(\sum X)^2}{n}}{n - 1}}$$

EJEMPLO: La empresa DAD, una empresa de equipos que instala abridores automáticos de puertas de garaje. Con base, en una muestra, los siguientes son los tiempos, en minutos requeridos para instalar 10 puertas: 28, 32, 24, 46, 44, 40, 54, 38, 32 y 42. Determine:

- Calcule la varianza elevando al cuadrado las desviaciones individuales respecto de la media.
- Evalué la varianza elevando también al cuadrado los valores originales.
- Determine la desviación estándar muestral.

SOLUCION:

Datos	Tiempo (min) X	$X - \bar{X}$	$(X - \bar{X})^2$	X^2
1	28	-10	100	784
2	32	-6	36	1024
3	24	-14	196	576
4	46	8	64	2116
5	44	6	36	1936
6	40	2	4	1600
7	54	16	256	2916
8	38	0	0	1444
9	32	-6	36	1024
10	42	4	16	1764
TOTAL	380	0	744	15184

$$\bar{X} = \frac{380}{10} = 38 \text{ min}$$

$$\text{a) } s^2 = \frac{\sum (X - \bar{X})^2}{n-1} = \frac{744}{10-1} = 82.67 \text{ minutos}^2$$

$$\text{b) } s^2 = \frac{\sum X^2 - \frac{(\sum X)^2}{n}}{n-1} = \frac{15184 - \frac{(380)^2}{10}}{10-1} = 82.67 \text{ minutos}^2$$

$$\text{c) } s = \sqrt{82.67} = 9.09 \text{ minutos}$$

2.8 Medidas de dispersión para datos agrupados en una distribución de frecuencias

Cuando se trabaja con datos agrupados, es decir, cuando los datos se han organizado en intervalos o clases en lugar de observaciones individuales, se utilizan medidas de dispersión específicas para este tipo de datos.

Estas medidas de dispersión son útiles cuando se trabaja con datos agrupados en histogramas o tablas de frecuencia. Sin embargo, son aproximaciones y pueden no capturar completamente la variabilidad de los datos, especialmente si los intervalos son muy amplios. La elección de la medida de dispersión adecuada dependerá de las necesidades y de cómo se hayan agrupado los datos.

2.8.1 Amplitud de variación

Para datos agrupados en una distribución de frecuencias, se resta el límite inferior de la clase más pequeña del límite superior de la clase mayor.

2.8.2 Varianza y desviación estándar de una muestra

$$s^2 = \frac{\sum fX^2 - \frac{(\sum fX)^2}{n}}{n-1} \quad \text{Varianza de una muestra}$$

$$\sigma^2 = \frac{\sum fX^2 - \frac{(\sum fX)^2}{N}}{N} \quad \text{Varianza de una población}$$

$$s = \sqrt{\frac{\sum fX^2 - \frac{(\sum fX)^2}{n}}{n-1}}$$

s desviación estándar muestral

X punto medio de clase

f frecuencia de clase

n número total de observaciones en la muestra

EJEMPLO: Una muestra de las cantidades que los empleados de DPC invierten quincenalmente en el plan de participación de utilidades, se organizó en una distribución de frecuencias. ¿Cuál es la amplitud de la muestra, la desviación estándar de estos datos? ¿Cuál es la varianza muestral?

Cantidad invertida	Número de empleados
\$30 a \$ 35	3
35 a 40	7
40 a 45	11
45 a 50	22
50 a 55	40
55 a 60	24
60 a 65	9
65 a 70	4

SOLUCION:

a) Amplitud o rango de la muestra = $70 - 30 = \$40$

b) Desviación estándar

$$s^2 = \frac{\sum fX^2 - \frac{(\sum fX)^2}{n}}{n - 1}$$

Cantidad invertida	Número de empleados f	PUNTO MEDIO X	f*X	f*X ²
\$30 a \$ 35	3	32,5	97,50	3168,75
35 a 40	7	37,5	262,50	9843,75
40 a 45	11	42,5	467,50	19868,75
45 a 50	22	47,5	1045,00	49637,50
50 a 55	40	52,5	2100,00	110250,00
55 a 60	24	57,5	1380,00	79350,00
60 a 65	9	62,5	562,50	35156,25
65 a 70	4	67,5	270,00	18225,00
Total	120		6185,00	325500,00

$$s^2 = \frac{325\,500 - \frac{(6185)^2}{120}}{120 - 1} = 56.43 \$^2$$

$$s = \sqrt{56.43} = \$7.51$$

$$s = \sqrt{\frac{\sum fX^2 - \frac{(\sum fX)^2}{n}}{n - 1}}$$

$$s = \sqrt{\frac{325.500 - \frac{(6185)^2}{120}}{120 - 1}} = \$7.51$$

c) Varianza muestral.

$$s^2 = (7.51)^2 = 56.40 \text{ dolares}^2$$

Interpretación y uso de la desviación estándar.

Por lo común la desviación estándar se emplea como una medida para comparar la dispersión en dos o más conjuntos de datos. Ejemplo:

Supóngase que los datos que se presentan a continuación corresponden al salario de grupos de trabajadores de Sierra y Amazonia.

Grupo	Media $\bar{X}$	Desviación estándar (s)
Sierra	\$952,35	\$23.47
Amazonia	\$955,27	\$93.45

Se puede analizar que las medias son aproximadamente iguales, pero las desviaciones estándar difieren en mayor medida en la Amazonia, pudiéndose concluir que la media para este grupo de trabajadores no es una medida muy confiable porque sus valores se concentran más dispersos en relación a la media.

2.8.2 Teorema de Chebyshev

Una desviación estándar pequeña para un conjunto de datos indica que estos se encuentran localizados cerca de la media, y cuando esta es grande sucede lo contrario.

El matemático ruso P.L Chebyshev (1821 – 1894) desarrollo un teorema que permite determinar la proporción mínima de los valores que se encuentran dentro de un número especificado de desviaciones estándares con respecto a la media. De acuerdo a Chebyshev se establece que sin importar la forma de la distribución (simetría y asimetría):

- Al menos de 3 a 4 valores o sea el 75% se encuentra entre la media ± 2 desviaciones estándar.
- Al menos 8 de cada 9 valores o sea 88.9%, estarán entre la media ± 3 desviaciones estándar
- Al menos 24 de 25 valores, es decir el 96%, se encuentran entre la media ± 5 desviaciones estándar.

En términos generales el teorema de Chebyshev establece que:

Teorema de Chebyshev: Para un conjunto cualquiera de observaciones (muestra o población), la proporción mínima de los valores que se encuentran dentro de k desviaciones estándar desde la media es al menos $(1 - 1/k^2)$, donde k es una constante mayor que 1.

$$\% \text{ de valores} = 1 - \frac{1}{k^2}$$

Ejemplo: Determinar para $\bar{X} = 51.54$ y $s = 7.51$ qué porcentaje de valores está a una distancia de ± 3.5 desviaciones estándar respecto a la media.

$$\% \text{ de valores} = 1 - \frac{1}{k^2} = 1 - \frac{1}{(3.5)^2} = 0.92$$

$$\% \text{ de valores} = 92\%$$

$$\text{Intervalo de valores: } 51.54 \pm 3.5(7.51)$$

$$\text{Intervalo de valores: } 25.26 \text{ a } 77.83$$

¿Qué porcentaje de valores está entre la media y ± 2 desviaciones estándar y cuáles son esos valores?

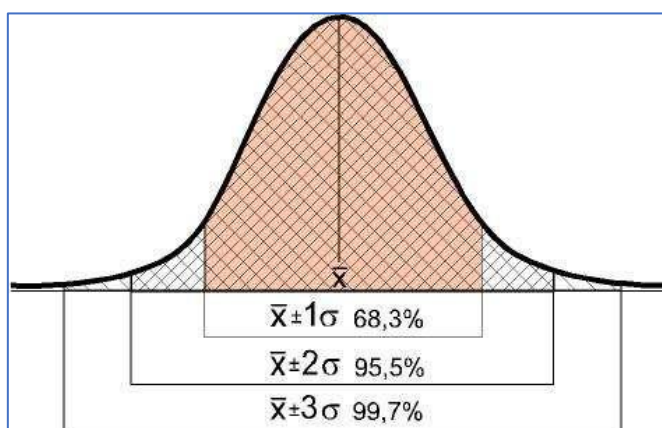
$$\% \text{ de valores} = 1 - \frac{1}{k^2} = 1 - \frac{1}{(2)^2} = 0.75 = 75\%$$

$$\text{Valores} = 51.54 \pm 2(7.51)$$

$$\text{Intervalo de valores} = 36.5 \text{ a } 66.56$$

2.8.3 Regla Empírica

Para una distribución de frecuencias simétrica de campana; aproximadamente 68% de las observaciones estará a $\pm$ una desviación estándar desde la media; aproximadamente el 95% entre ± 2 desviaciones estándar y prácticamente todas las observaciones el 99.7% estarán a ± 3 desviaciones estándar con respecto a la media.



EJEMPLO: Si $\bar{X} = 120$ y $s = 15$, prácticamente todas las observaciones se encuentran entre $120 + 3(15)$ y $120 - 3(15)$; o sea 75 y 165. Por lo tanto, la amplitud de variación es 90 que se obtiene de $165 - 75 = 90$.

De igual forma, si se conoce que la amplitud es 90 para una distribución de frecuencias simétrica de campana, podemos aproximar la desviación estándar dividiendo entre 6 el valor de la amplitud; para el ejemplo: $s = \frac{90}{6} = 15$

Ejemplo: Una muestra de la cantidad de dinero mensual destinados a alimentación por un ciudadano de la tercera edad que vive solo, sigue aproximadamente una distribución de frecuencia simétrica de campana. La media muestral es \$190 y desviación estándar es \$25. Utilizando la regla empírica. Determinar:

1. Aproximadamente, ¿entre cuales dos cantidades está el 68% de los gastos mensuales en alimentos?

$$X \pm s = 190 \pm 25 = 165 \text{ a } 215$$

2. Aproximadamente, ¿entre cuales dos cantidades se hallan el 95% de los gastos mensuales por alimentación?

$$X \pm 2s = 190 \pm 2(25) = 140 \text{ a } 240$$

3. Aproximadamente, ¿entre cuales dos cantidades se hallan todos de los gastos mensuales por alimentación?

$$X \pm 3s = 190 \pm 3(25) = 115 \text{ a } 265$$

2.8.4 Dispersión relativa

Resulta imposible una comparación directa de dos o más medidas de dispersión; por ejemplo, la desviación estándar para un ingreso mensual de un grupo de empleados ($s=\$1.150$) y la desviación estándar de la inasistencia para el mismo grupo de empleados ($s=5.3$ días); obviamente no podemos comparar directamente dólares con días de inasistencia.

A fin de realizar una comparación significativa de la distribución de ingresos y faltas, se puede convertir estas medidas a un valor relativo, es decir, a un porcentaje. El mismo es conocido como coeficiente de variación (CV), que es una medida muy útil cuando:

1. Los datos están en unidades diferentes (como dólares y días de inasistencia).
2. Los datos están en las mismas unidades, pero las medias muy distantes (como salarios de ejecutivos y salarios de empleados no calificados).

Coeficiente de variación: Es la razón de la desviación estándar a la media aritmética, expresada en porcentaje.

$$\text{Coeficiente de variación} \quad CV = \frac{s}{X} * 100$$

EJEMPLO 1: Un estudio de las calificaciones de un curso de principios de Administración y los años de servicio de los empleados en una Agencia Bancaria, arrojaron los resultados siguientes:

DESCRIPCION	$\bar{X}$	s
Calificaciones	120	20
Años de servicio	20	2

Comparar la dispersión relativa de las dos distribuciones empleando el coeficiente de variación.

$$\text{Calificaciones:} \quad CV = \frac{20}{120} * 100 = 16.67\%$$

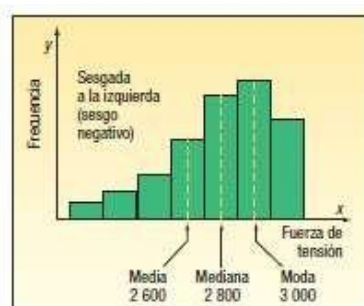
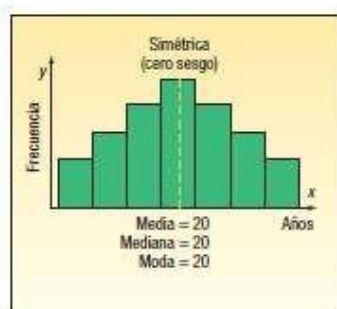
Años de servicio: $CV = \frac{2}{20} * 100 = 10\%$

Se puede apreciar que existe una mayor dispersión relativa respecto a la media en la distribución de calificaciones en relación a los años de servicio.

2.8.5 Posiciones relativas de la media, la mediana y la moda

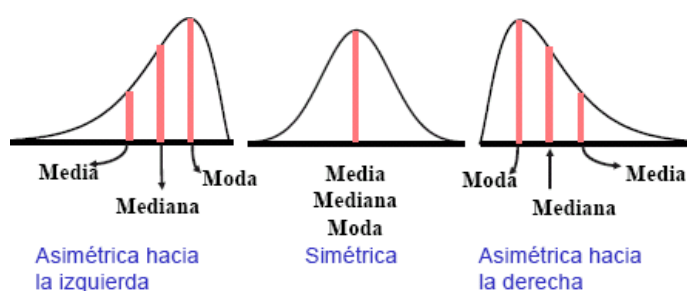
Se trata de una distribución simétrica que también tiene forma de campana. Esta distribución posee la misma forma a cualquier lado del centro. Si el polígono estuviera doblado a la mitad, las dos mitades serían idénticas. En cualquier distribución simétrica, la moda, la mediana y la media siempre son iguales. Cualquiera frecuencia con sesgo positivo. La moda es la menor de las tres medidas. La media no sería una medida adecuada. La mediana y la moda serían más representativas.

Por el contrario, si una distribución tiene un **sesgo negativo**, la media es la menor medida de las tres. Por supuesto, la media es sensible a la influencia de una cantidad extremadamente pequeña de observaciones. La mediana es mayor que la media aritmética y la moda es la más grande de las tres medidas. De nuevo, si la distribución tiene un sesgo muy pronunciado, la media no se utilizaría para representar a los datos.



2.8.6 Asimetría o Curtosis

Para evaluar la tendencia central de un conjunto de observaciones se la utilizó la media, mediana y moda; además de las medidas de dispersión de las observaciones respecto a la media. Además, se ha analizado otra característica que se refiere al grado de asimetría de una distribución que se resume en el gráfico siguiente:



Recuerde que una distribución de frecuencias es simétrica cuando media mediana y moda son iguales, es decir no tiene sesgo, es decir se asimetría es nula. Así mismo si una o más observaciones son sumamente altas, la media es mayor que las otras medidas y es asimétrica con sesgo positivo y cuando las observaciones una o más son sumamente pequeñas la media es menor que las otras medidas y se produce una asimetría con sesgo negativo.

Karl Pearson desarrollo una medida para evaluar el grado de asimetría en lo referente a su sesgo denominado coeficiente de asimetría.

$$CV = \frac{3(\text{media} - \text{mediana})}{\text{Desviación estándar}}$$

EJEMPLO: Una muestra de las casas que se ofrecen en venta en la ciudad de Quito, revelo que el precio medio solicitado es de \$75.900, la mediana \$70.100 y el precio modal es \$67.200. La desviación estándar de la distribución es \$5.900

Media=\$75 900

Mediana=\$70 100

Moda=\$67 200

a) La distribución de precios ¿es simétrica, o asimétrica con sesgo positivo o negativo?

Es asimétrica con sesgo positivo pues la media es mayor que las otras medidas de tendencia central.

b) ¿Cuál es el coeficiente de asimetría? Interpretélo.

$$CV = \frac{3(\text{media} - \text{mediana})}{\text{Desviación estándar}}$$

$$CV = \frac{3(75.900 - 70.100)}{5.900} = 2.95$$

Interpretación: El coeficiente de asimetría por lo general se encuentra entre -3 y +3, por lo que el valor 2.95 indica un grado importante con sesgo positivo. Lo que indica que un alto número de las casas ofrecidas en venta tendrán dificultad de ser vendidas.

2.8.7 Cuartiles deciles y centiles (o percentiles)

La varianza y la desviación estándar son las medidas de dispersión más ampliamente utilizadas. Además, existen otras medidas para determinar la diseminación de un conjunto de datos; un método es determinar la ubicación de los valores que dividen un conjunto de observaciones en partes iguales. Estas medidas son los cuartiles, deciles y los centiles.

Los cuartiles dividen un conjunto de observaciones en cuatro partes iguales. Para aclarar lo expuesto, recordemos que la mediana representaba el valor central de un conjunto de

observaciones ordenadas de menor a mayor; esto es, el 50% de las observaciones son mayores que la mediana y el 50% son menores. De manera similar, los cuartiles dividen un conjunto de observaciones en 4 partes iguales. El primer cuartil simbolizado como Q_1 , es el valor abajo del cual se encuentran el 25% de las observaciones y el tercer cuartil Q_3 , constituye el valor abajo del cual se encuentra el 75% de las observaciones. Q_2 viene a representar la mediana.

Se tiene que Q_1 puede ser considerada como la “mediana” de la mitad inferior de las observaciones y Q_3 como la mediana de la mitad superior de los datos.



Los deciles dividen a un conjunto de observaciones en 10 partes iguales; así como los centiles en 100 partes iguales. Por lo tanto, si una persona se entera que su promedio de calificaciones se encuentra en el octavo decil, se puede concluir que el 80% de los estudiantes tiene un promedio menor que uno (persona) y que el 20% tiene un promedio mayor.

Para encontrar la ubicación del centil deseado utilizamos la fórmula:

$$UBICACION DE UN PORCENTIL: L_P = (n + 1) \frac{P}{100}$$

Donde:

n =número de observaciones.

P = porcientil deseado.

EJEMPLO: A continuación, se presenta las comisiones obtenidas el mes pasado por una muestra de 15 corredores de la empresa “CCA”.

\$ 2.038	1758	1721	1637	2097	2047	2205	1787	2287
\$ 1.940	2311	2054	2406	1471	1460			

Localice la mediana, el primero y el tercer cuartil de las comisiones en cuestión.

- Organizar los datos de menor a mayor.

1460	1471	1637	1721	1758	1787	1940	2038
2047	2054	2097	2205	2287	2311	2406	

- Mediana: se encuentra en el centro de los valores y se localiza: $\frac{(n+1)}{2} = \frac{(15+1)}{2} = 8$; la posición 8 corresponde al valor de Mediana=\$2038

$$L_P = (n + 1) \frac{P}{100}$$

$$L_P = (15 + 1) \frac{50}{100} = 8$$

Primer cuartil:

|

$$L_P = (n + 1) \frac{P}{100}$$

$$L_P = (15 + 1) \frac{25}{100} = 4$$

$$Q_1 = \$1721$$

- Tercer Cuartil

$$L_P = (n + 1) \frac{P}{100}$$

$$L_P = (15 + 1) \frac{75}{100} = 12$$

$$Q_3 = \$2205$$

- Segundo decil

$$L_P = (n + 1) \frac{P}{100}$$

$$L_P = (15 + 1) \frac{20}{100} = 3.2$$

$$Q_2 = 1637 + 0.2(1721 - 1637) = \$1\,653.80$$

- Octavo decil

$$L_P = (n + 1) \frac{P}{100}$$

$$L_P = (15 + 1) \frac{80}{100} = 12.8$$

$$Q_8 = 2\,205 + 0.80(2\,287 - 2\,205) = \$2\,270.60$$

- Vigésimo cuarto centil

$$L_P = (n + 1) \frac{P}{100}$$

$$L_P = (15 + 1) \frac{24}{100} = 3.84$$

$$Q_{24} = 1637 + 0.84(1721 - 1637) = \$1\,707.56$$

- El tercer quintil

$$L_P = (n + 1) \frac{P}{100}$$

$$L_P = (15 + 1) \frac{60}{100} = 9.60$$

$$Q_3 = 2047 + 0.60(2054 - 2047) = \$2\,051.20$$

EJEMPLO 2: Supongamos que un conjunto de datos incluye los siguientes valores: 91, 75, 61, 101, 43 y 104. Determinar el primer y tercer cuartil.

- Organizar datos de menor a mayor.

43, 61, 75, 91, 101 y 104

$$L_P = (n + 1) \frac{P}{100} \qquad L_P = (6 + 1) \frac{25}{100} = 1.75$$

Para determinar Q_1 tome la posición 1 y el 0.75 de la distancia entre 61 y 43.

$$Q_1 = 43 + 0.75(61 - 43) = 56.5$$

Para determinar Q_3 determinemos la posición del tercer cuartil tenemos:

$$L_P = (n + 1) \frac{P}{100} \qquad L_P = (6 + 1) \frac{75}{100} = 5.25$$

$$Q_3 = 101 + 0.25(104 - 101) = 101.75$$

Podemos extender lo anterior tanto a los deciles como a los centiles. Si deseáramos encontrar el centil 23 en una muestra de 80 observaciones; buscaríamos la posición:

$$L_P = (n + 1) \frac{P}{100} \qquad L_P = (80 + 1) \frac{23}{100} = 18.63$$

$$Q_{23} = (\text{valor de posición } 18) + 0.63(\text{valor de posición } 19 - \text{valor de posición } 18)$$

2.9 Diagrama de caja.

Un diagrama de caja es una representación gráfica, basada en cuartiles, que ayuda a presentar un conjunto de datos. Para construir un diagrama de caja, sólo necesita cinco estadísticos: el valor mínimo, Q_1 (primer cuartil), la mediana, Q_3 (tercer cuartil) y el valor máximo. Un ejemplo ayudará a explicarlo.

Ejemplo: Pizza Hot; ofrece entregas gratuitas de pizza a 15 Km a la redonda. Alex, el propietario, desea información relacionada con el tiempo de entrega. ¿Cuánto tiempo tarda una entrega típica? ¿En qué margen de tiempo deben completarse la mayoría de las entregas? En el caso de una muestra de 20 entregas, Alex recopiló la siguiente información:

Valor mínimo = 13 minutos

Q_1 = 15 minutos

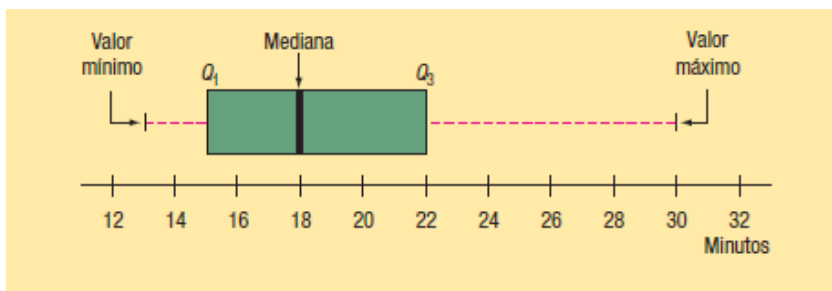
Mediana = 18 minutos

Q_3 = 22 minutos

Valor máximo = 30 minutos

Elabore un diagrama de caja de los tiempos de entrega. ¿Qué conclusiones deduce sobre los tiempos de entrega?

Solución El primer paso para elaborar un diagrama de caja consiste en crear una escala adecuada a lo largo del eje horizontal. Luego, se debe dibujar una caja que inicie en Q_1 (15 minutos) y termine en Q_3 (22 minutos). Dentro de la caja trazamos una línea vertical para representar a la mediana (18 minutos). Por último, prolongamos líneas horizontales a partir de la caja dirigida al valor mínimo (13 minutos) y al valor máximo (30 minutos). Estas líneas horizontales que salen de la caja, a veces reciben el nombre de bigotes, en virtud de que se asemejan a los bigotes de un gato.



El diagrama de caja muestra que el valor medio de las entregas, 50%, consume entre 15 y 22 minutos. La distancia entre los extremos de la caja, 7 minutos, es el rango intercuartil. Este rango, que es la distancia entre el primer y el tercer cuartil, muestra la propagación o dispersión de la mayoría de las entregas.

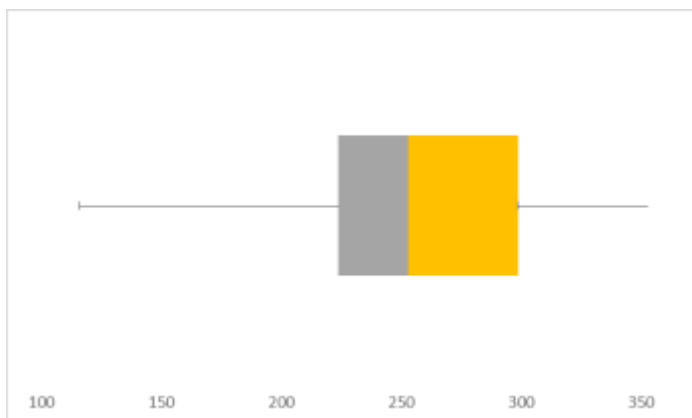
Ejemplo 2: Una muestra de 28 departamentos de tiempo compartido en el área de Orlando, Florida, reveló las siguientes tarifas diarias de una suite con una recámara. Por comodidad, los datos se encuentran ordenados de menor a mayor. Construya un diagrama de caja para representar los datos. Haga algún comentario sobre la distribución. Identifique el primer y tercer cuartiles, y la mediana.

\$116	\$121	\$157	\$192	\$207	\$209	\$209
229	232	236	236	239	243	246
260	264	276	281	283	289	296
307	309	312	317	324	341	353

$$\text{Primer cuartil: } Ln = (28 + 1) \frac{25}{100} = 7.25 \quad Q_1 = 209 + 0.25(229 - 209) = 214$$

$$\text{Segundo cuartil: } Ln = (28 + 1) \frac{50}{100} = 14.5 \quad Q_2 = 246 + 0.50(260 - 246) = 253$$

$$\text{Tercer cuartil: } Ln = (28 + 1) \frac{75}{100} = 21.75 \quad Q_3 = 296 + 0.75(307 - 296) = 304.25$$



2.10 Números índices

Número índice es una medida estadística diseñada para poner de relieve cambios en una variable o en un grupo de variables relacionadas con respecto al tiempo, situación geográfica, ingresos, o cualquier otra característica. (Spiegel, 1997)

Número índice es un número que expresa el cambio relativo en precio, cantidad o valor comparado con un periodo base. (Lind, Marchal, & Wathen, 2006).

- Es un porcentaje, pero generalmente se omite el signo porcentual.
- Tiene un período base.
- La mayor parte de los índices se aproximan al décimo más próximo de un porcentaje .
- La base de la mayor parte de los índices es 100.

2.10.1 Clasificación de los números índices

Números índices de precios

Es un número que expresa el cambio relativo de precio comparando con un periodo base. El índice de este tipo más conocido es Índice de precios al consumidor (IPC), el cual mide el costo de vida en los países.

Números índices de cantidad

Es un número que expresa el cambio relativo de cantidad de una variable comparando con un periodo base.

Números índices de valor

Es un número que expresa el cambio relativo de valor comparando con un periodo base. Es decir, mide los cambios en el valor monetario total. (Levin & Rubin, 2004)

2.10.2 Números índices simples

Son los que se refieren a una sola magnitud, y por lo tanto nos proporcionan la variación que ha sufrido esa magnitud en dos periodos distintos.

Se calcula hallando el cociente del valor del año determinado entre el valor del año base por 100, así:

$$P = \frac{pt}{po} * 100$$

Donde:

P = índice simple

pt = valor del año determinado

po = valor del año base

Ejemplo de aplicación 1

El salario básico unificado en enero del 2012 fue de \$292, el salario básico unificado en enero del 2015 fue de \$354. ¿Cuál es el índice correspondiente para los trabajadores en enero del 2015, con base en los datos de enero del 2012?

$$P = \frac{pt}{po} * 100 = \frac{354}{292} 100 = 121.23$$

Durante este periodo aumento en $121.23 - 100 = 21.23$ el salario básico unificado.

Ejemplo de aplicación 2

En el 2013 la población de hombres en Ecuador fue de 7,868,368 y el de mujeres fue de 7,869,510 ¿Cuál fue la proporción de la población de hombres comparada con la de mujeres?

$$P = \frac{pt}{po} * 100 = \frac{7,868,368}{7,869,510} 100 = 99.98$$

$$99.98 - 100 = -0.02$$

¿Cuál fue la proporción de la población de mujeres comparada con la de hombres?

$$P = \frac{pt}{po} * 100 = \frac{7\,869\,510}{7\,868\,368} 100 = 100.01$$

$$100.01 - 100 = 0.01$$

El índice de hombres es de 99.98 de la población de mujeres o la población de hombres es $100 - 99.98 = -0.02$ menor que la de mujeres

Ejemplo de aplicación 3

Los siguientes datos, se tomaron de los informes anuales de la empresa Johnson & Johnson, de la misma que sus acciones comunes se enlistan en la Bolsa de Valores con el símbolo de JNJ. Tomando como base el año 1991, calcular el Índice Simple.

Años	Ventas nacionales en miles de dólares	Cálculo del Índice	Índice Simple %
1991	5,43	1,0000	100,0%
1992	6,25	1,1510	115,1%
1993	6,9	1,2707	127,1%
1994	7,2	1,3260	132,6%
1995	7,81	1,4383	143,8%
1996	9,19	1,6924	169,2%

2.10.3 Números índices no ponderados

2.10.3.1 Promedio simple de los índices de precios

Este índice se obtiene sumando los índices simples de cada producto y dividiendo para el número de productos.

Así:

$$P = \frac{\sum P_i}{n}$$

P_i = Índices simples de cada uno de los productos

n = número de productos

Ejemplo de aplicación 4

Tomando como base los siguientes datos se calcula el promedio simple de los índices de precios relativos.

$$P = \frac{100 + 115.1 + 127.07 + 132.60 + 143.83 + 169.24}{6}$$

$$P = \frac{787.85}{6}$$

$$P = 131.31$$

Ejemplo de aplicación 5

Un administrador estudia la evolución de los precios de un artículo que produce su empresa en los 5 últimos años, los valores se registran en la siguiente tabla. Calcule el promedio simple de los índices de precios, tomando como periodo de referencia el año 1.

Años	1	2	3	4	5
Precio del producto	4	5.5	6	5	8

Primero calculo los índices simples

Años	Precio del producto	Índice Simple
1	4	100
2	5,5	137,5
3	6	150
4	5	125
5	8	200
		712,5

$$P = \frac{\sum P_i}{n} = \frac{712.5}{5} = 142.5$$

Esto indica que el precio del producto a incrementado en 42.5

2.10.3. Índice agregado simple

Un índice agregado simple se calcula sumando todos los elementos de un periodo dado y luego dividiendo este resultado entre la suma de los mismos elementos durante el periodo base. Así:

$$P = \frac{\sum P_i}{\sum P_0} * 100$$

P_i : Cantidad de elementos del periodo que se desea el índice

P_0 : Cantidad de elementos en el año base

Ejemplo de aplicación 6

Determine el índice agregado simple de precios para el año 2015 y 2012 de cinco productos considerados, usando como año base 2012.

Producto	2012	2015
Leche \$/lt	0,84	0,86
Pan \$/und	0,14	0,16
Huevos\$/docena	1,44	1,5
Azucar \$/Kg	1,6	1,8
Arroz (Kg)	1,2	1,5
Total	5,22	5,82

$$P = \frac{\sum P_i}{\sum P_0} * 100 = \frac{5.82}{5.22} * 100$$

$$P = 111.49$$

2.11 Números índices ponderados

Dos métodos para calcular el índice de precios compuestos o ponderado son el método de Laspeyres y el de Paasche. Difieren sólo en el periodo para la ponderación. En el método de Laspeyres se utilizan ponderaciones en el periodo base; es decir, los precios y las cantidades originales de los artículos comprados se utilizan para encontrar el

según el problema. En el método de Paasche se utilizan ponderaciones en el año en curso. (Lind, Marchal, & Wathen, 2006)

2.11.1 Índice de Laspeyres

$$P = \frac{\sum p_t q_0}{\sum p_0 q_0} \times 100$$

Dónde:

P es el índice de precios.

p_t es el precio actual

p_0 es el precio del periodo base

q_0 es la cantidad en el periodo base

(Lind, Marchal, & Wathen, 2006)

Ejemplo de aplicación 7

Determinar un índice de precios ponderado con el método de Laspeyres, con los precios dados en la siguiente tabla tomando como referencia

2.11.2 Índice de Paasche

El cálculo es similar que el índice de Laspeyres, pero en lugar de emplear cantidades en el periodo base como ponderaciones, se utilizan cantidades en el período actual.

$$P = \frac{\sum p_t q_t}{\sum p_0 q_t} \times 100$$

Ejemplo de aplicación 8

En el ejercicio de aplicación 7, calculemos el índice de Paasche

TABLA 15-3 Precio y cantidad de artículos de alimentos en 1999 y en 2009

Artículo	Precio en 1999	Cantidad en 1999	Precio en 2009	Cantidad en 2009
Pan blanco, costo por libra	\$0.87	50	\$1.28	55
Huevos, docena	1.05	26	2.17	20
Leche blanca, galón	2.94	102	3.87	130
Manzanas, Red Delicious, 1 libra	0.86	30	1.16	40
Jugo de naranja, concentrado, 12 onzas	1.75	40	2.54	41
Café, 100% de grano tostado, 1 libra	3.43	12	3.68	12

Determine un índice de precios ponderado con el método de Laspeyres. Interprete el resultado.

Artículo	1999		2009		pt*qo	po*qt
	po	qo	pt	qt		
1	0,87	50	1,28	55	64	43,5
2	1,05	26	2,17	20	56,42	27,3
3	2,94	102	3,87	130	394,74	299,88
4	0,86	30	1,16	40	34,8	25,8
5	1,75	40	2,54	41	101,6	70
6	3,43	12	3,68	12	44,16	41,16
					695,72	507,64

$$P = \frac{\sum pt.qo}{\sum po.qt} = \frac{695.72}{507.64} = 137.05$$

Artículo	1999		2009		pt*qt	po*qt
	po	qo	pt	qt		
1	0,87	50	1,28	55	70,4	47,85
2	1,05	26	2,17	20	43,4	21
3	2,94	102	3,87	130	503,1	382,2
4	0,86	30	1,16	40	46,4	34,4
5	1,75	40	2,54	41	104,14	71,75
6	3,43	12	3,68	12	44,16	41,16
					811,6	598,36

$$P = \frac{\sum pt.qt}{\sum po.qt} = \frac{811.6}{598.36} = 135.64$$

Ejemplo de aplicación 9

Los precios, de cuatro artículos que se indican a continuación, en los años 2001 y 2003, permitan calcular el índice de Paasche.

Artículo	2001		2003		pt*qt	po*qt
	po	qo	pt	qt		
Pan	0,05	20	0,08	24	1,92	1,2
Huevos	0,06	30	0,09	36	3,24	2,16
Leche	0,35	10	0,48	15	7,2	5,25
Manzana	0,15	15	0,25	20	5	3
Σ	0,61	75	0,9	95	17,36	11,61

$$P = \frac{\sum p_t q_t}{\sum p_0 q_t} \times 100$$

$$P = \frac{17.36}{11.61} \times 100 = 149.52$$

Este resultado, indica que el precio de este grupo de alimentos, aumentó en el 49.52 %, en el período de dos años.

2.11.3 Índice de Fisher

Es la media geométrica de los índices de Laspeyres y Paasche.

$$\text{Índice ideal de Fisher} = \sqrt{(\text{Índice de Laspeyres})(\text{Índice de Paasche})}$$

Parece ser el ideal porque combina las mejores características del de Laspeyres y del de Paasche.

Ejemplo de aplicación 10

Al tener ya determinado el índice de Laspeyres y el índice de Paasche, en el ejemplo de aplicación 9 procedemos a calcular el índice ideal de Fisher.

$$\text{Índice ideal de Fisher} = \sqrt{(137)(135.6)}$$

$$\text{Índice ideal de Fisher} = 136.30$$

2.11.3 Índices de valor

Mide cambios tanto en los precios como en las cantidades que intervienen. Se usa precios y cantidades del período base y del período actual.

$$V = \frac{\sum pt * qt}{\sum po * qo}$$

donde:

V es el índice de valor

po es el precio en el período base

qo es la cantidad consumida en el período base

qt es la cantidad consumida en el período actual

pt es el precio actual

Ejemplo de aplicación 11

Los precios de cuatro artículos que se indican a continuación, en los años 2001 y 2003, permitirán calcular el índice de Valor.

Artículo	2001		2003		pt*qt	po*qo
	po	qo	pt	qt		
Pan	0,05	20	0,08	24	1,92	1
Huevos	0,06	30	0,09	36	3,24	1,8
Leche	0,35	10	0,48	15	7,2	3,5
Manzana	0,15	15	0,25	20	5	2,25
Σ	0,61	75	0,9	95	17,36	8,55

$$V = \frac{\sum pt * qt}{\sum po * qo}$$

$$V = \frac{17.36}{8.55} * 100 = 203.04$$

Con este resultado, miramos que el valor de este grupo de alimentos, aumentó en el 103.04 %, en el período de dos años.

3. UNIDAD III

Introducción a la teoría de probabilidades



Introducción

Se ha estudiado la Estadística Descriptiva para la organización de datos en una distribución de frecuencias para mostrar el valor más alto como el más bajo, así como la visualización de la concentración de datos. Además, se ha podido establecer medidas de tendencia central como la media, moda y mediana; así como medidas de dispersión respecto a la media. Todas estas con el objetivo de tomar las mejores decisiones.

La segunda faceta de la Estadística que se refiere a la posibilidad de que algo ocurrirá en el futuro, esta faceta se denomina Inferencia Estadística o Estadística Inferencial. Recordemos la definición hecha en la primera parte.

Estadística Inferencial, Inferencia Estadística o Estadística Inductiva

Conjunto de métodos utilizados para saber algo acerca de una población, basándose en una muestra.

Población

Conjunto de todos los posibles individuos, objetos o medidas de interés.

Muestra

Una porción, o parte, de la población de interés.

En pocas ocasiones el encargado de tomar decisiones dispone de toda la información a partir de la cual pueda tomar una determinación. Por ejemplo:

- Una empresa de cosméticos está empeñada en ofrecer un nuevo producto para la limpieza facial y desea conocer si este tendrá aceptación en el mercado. Para minimizar los riesgos de una decisión equivocada recurre a una empresa de encuestas para realizar la misma sobre una muestra de 1500 personas de una población y les preguntara sobre la aceptación del nuevo producto.
- Una fábrica que produce adoquines de cemento requiere conocer si su producto cumple las especificaciones de resistencia de 300kg/cm^2 ; solicitada por una gran empresa de construcción para lo cual recurre a un muestreo aleatorio de su producción diaria de 3500 unidades, consistente en 10 muestras para realizar un ensayo de compresión de las muestras, el mismo que es de carácter destructivo, el mismo que no podría ser realizado sobre toda la producción. En base de estos resultados estará en capacidad de ofrecer su producto a la constructora.

La inferencia estadística es la encargada de hacer deducciones de una población basada en una muestra tomada de una población; debido a que existe una incertidumbre considerable al tomar estas decisiones, es importante que todos los riesgos implícitos conocidos, se evalúen en forma científica; esta evaluación se puede realizar utilizando la teoría de probabilidades, a la que se le conoce como, "ciencia de la incertidumbre".

El empleo de la teoría de probabilidades permite a quien toma decisiones, analizar con información limitada, los riesgos y minimizar el azar inherente. Como en los ejemplos lanzar un nuevo producto o aceptar por parte de la constructora toda la producción de adoquines.

Nos estamos introduciendo en la utilización de los términos probabilidad, posibilidad y azar. Escuchamos con alguna frecuencia que existe posibilidad de un 80% de lluvia a través del pronóstico meteorológico o que existe la probabilidad de que un nuevo producto tiene el 0.15 de que, si se lanza al mercado los consumidores estén dispuestos a comprarlo. Estos ejemplos nos permiten hacer juicios sobre la decisión de llevar o no llevar paraguas respecto a la lluvia o si se lanza o no el nuevo producto.

3.1 Probabilidad

Valor entre cero y uno, inclusive, que describe la posibilidad relativa de que ocurrirá un evento.

$$P(E) = \frac{n(E)}{n(N)} \quad 0 \leq P(E) \leq 1$$

Ejemplo 1: En el lanzamiento de una moneda. ¿Cuál es la probabilidad que el resultado sea cara?

$$P(E) = \frac{n(E)}{n(N)} = \frac{1}{2} = 0.50 = 50\% \quad 0 \leq P(E) \leq 1$$

Ejemplo: Lanzar un dado. Probabilidad de que un lanzamiento sea mayor que 4

$$P(E) = \frac{n(E)}{n(N)} = \frac{2}{6} = 0.3333 = 33.33\%$$

La probabilidad puede expresarse como: $\frac{5}{13}$, $\frac{6}{100}$, 0.52, 32%.

Ejemplo 2: De una corrida de producción de 5 000 focos luminosos, 2% de los cuales están defectuosos, se selecciona uno de manera aleatoria. ¿Cuál es la probabilidad de que sea el foco defectuoso? ¿Cuál es la probabilidad de que no lo sea?

(a) ¿Cuál es la probabilidad de que sea el foco defectuoso?

$$P(E) = \frac{n(E)}{n(N)}$$

$$n(N) = 5\,000$$

$$n(E) = 0.02 * 5\,000 = 100$$

$$P(E) = \frac{100}{5\,000} = \frac{1}{50} = 0.02$$

(b) ¿Cuál es la probabilidad de que no lo sea?

$$P(E) + P(\sim E) = 1 \qquad P(\sim E) = 1 - P(E)$$

$$P(\sim E) = 1 - 0.02 = 0.98$$

Recuerde una probabilidad únicamente puede tomar valores entre cero y uno. La probabilidad 1 representa algo que seguramente va a suceder y cero algo que no puede ocurrir.

Se utilizan tres palabras clave en el estudio de la probabilidad: experimento, resultado y evento. Estos términos se los utiliza cotidianamente, pero en estadística tienen significados específicos.

3.1.1 Experimento

Proceso que conduce a la ocurrencia de una (y solamente una) de varias observaciones posibles.

Esta definición es la más general que se la utiliza en las ciencias físicas, donde uno se imagina a alguien manipulando microscopios o tubos de ensayo. En el caso de probabilidad, un experimento tiene dos o más resultados posibles, y es incierto cual habrá de ocurrir.

3.1.2 Resultado

Lo que resulta específicamente de un experimento.

- Lanzar una moneda al aire es un experimento, se puede observar el lanzamiento, pero no se sabe el resultado si caerá cara o cruz.
- Una encuesta a 200 personas para conocer el gusto o preferencia entre tres refrescos (Coca Cola, Pepsi Cola, Tesalia) también es un experimento, que no se sabe que resultado o preferencia nos dará.

En los dos casos se obtendrá un resultado particular, quedando el resto como resultados alternativos.

3.1.3 Evento

Conjunto de dos o más resultados de un experimento.

Ejemplos: Diferenciar los términos experimento, resultado y evento.

1. **Al lanzar un dado;** determine el cuál es el experimento, resultado y evento

Experimento: Lanzar el dado.

Resultado: existen 6 resultados posibles.

Evento: existe una variedad, que salga un número par o impar, un número mayor que 3 o menor que 3...

2. El Departamento de vía pública de la ciudad de Ambato, está considerando ampliar la Avenida Cevallos a 3 carriles. Antes de tomar la decisión, se preguntó a 500 ciudadanos si apoyaban la ampliación.

a) ¿Cuál es el experimento?

El experimento es preguntar a 500 ciudadanos si apoyan o no la ampliación a tres carriles.

b) ¿Cuáles son algunos de los eventos posibles?

- a favor.
- en contra

c) Mencione 2 resultados posibles

Las respuestas variaran, pero existen, pero 2 posibilidades son:

- Una mayoría está a favor de la ampliación, puede ser más de 300 están a favor de la ampliación, o
- Una mayoría está en contra de la ampliación, podría ser 251 o más.

3.2 Enfoques de probabilidad.

Se analizarán dos enfoques de probabilidad, específicamente, los puntos de vista objetivo y subjetivo. La probabilidad objetiva puede subdividirse en probabilidad clásica y probabilidad empírica.

3.2.1 Probabilidad clásica

Se basa en la consideración de que los resultados de un experimento son igualmente posibles. La probabilidad de que suceda un evento se calcula dividiendo el número de resultados favorables entre el número total de resultados posibles.

$$P(E) = \frac{\text{Número de resultados favorables}}{\text{Número total de resultados posibles}} \quad 0 \leq P(E) \leq 1$$

EJEMPLO: Considerar el experimento de extraer una carta del naipes español. ¿Cuál es la probabilidad del evento es una carta de corazones negros?



$$\text{Probabilidad una carta de corazones negros: } P(E) = \frac{13}{52} = 0.25 = 25\%$$

Si solo uno de varios eventos puede ocurrir cada vez, se dice que son mutuamente excluyentes.

Mutuamente excluyente

Se expresa esto **porque la ocurrencia de cualquier evento implica que ningún otro puede ocurrir al mismo tiempo**. Y son:

Colectivamente exhaustivo

Se señala esto porque al menos uno de los eventos debe ocurrir cuando se realiza un experimento.

En el experimento de extraer una carta, cada resultado será una carta de corazones negros o diamantes, o corazones rojos o trébol. Por lo que el conjunto es colectivamente exhaustivo.

Si el conjunto de eventos es colectivamente exhaustivo y los eventos son mutuamente excluyentes, la suma de probabilidades es igual a 1. Por ejemplo, la lanzar una moneda; la probabilidad de el evento de caer cara es 0.5, así como el evento de caer sello es igualmente 0.5 y la suma de probabilidades de los eventos posibles es igual a 1.

Recuerde: Suma de probabilidades = 1

NOTA: Para que se pueda aplicar el enfoque clásico, los eventos deben tener la misma posibilidad de ocurrir y el conjunto de eventos deben ser mutuamente excluyente y colectivamente exhaustivo (Suma de probabilidades igual a 1)

3.2.2 Probabilidad empírica

Otra manera de definir la probabilidad es con base a las frecuencias relativas. La probabilidad de que un evento ocurra a largo plazo se determina observando en que fracción de tiempo sucedieron eventos semejantes en el pasado.

$$P(E) = \frac{\text{Número de veces que el evento ocurrió en el pasado}}{\text{Número total de observaciones}}$$

El enfoque empírico de la probabilidad se basa en la llamada ley de los grandes números. La clave para determinar probabilidades de forma empírica consiste en que una mayor cantidad de observaciones proporcionarán un cálculo más preciso de la probabilidad.

LEY DE LOS GRANDES NÚMEROS En una gran cantidad de intentos, la probabilidad empírica de un evento se aproximará a su probabilidad real o clásica.

Supongamos que lanzamos una moneda común. El resultado de cada lanzamiento es cara o cruz. Si lanza la moneda una sola vez, la probabilidad empírica de las caras es cero o uno.

Número de ensayos	Número de caras	Frecuencia relativa de las caras
1	0	.00
10	3	.30
50	26	.52
100	52	.52
500	236	.472
1 000	494	.494
10 000	5 027	.5027

¿Qué hemos demostrado? A partir de la definición clásica de probabilidad, la posibilidad de obtener una cara en un solo lanzamiento de una moneda común es de 0.5. Según el enfoque empírico de la frecuencia relativa de la probabilidad, la probabilidad del evento se aproxima al mismo valor determinado de acuerdo con la definición clásica de probabilidad.

EJEMPLO: Se efectuó un estudio de 500 graduados en administración de empresas de la UCE. Este experimento reveló que 152 de los 500 no estaban empleados según su principal área de estudio en la universidad. Por ejemplo, una persona que se graduó en un área especializada en contaduría, ahora es gerente de mercadeo de una empresa de procesamiento de tomates. ¿Cuál es la probabilidad de que un graduado específico en administración este empleado en un área distinta a la principal de sus estudios en universidad?

$$P(E) = \frac{\text{Núm. de veces que el evento ocurrió en el pasado}}{\text{Número total de observaciones}}$$

$$P(E) = \frac{152}{500}$$

$$P(E) = 0.304$$

En conclusión, en base a la experiencia, existe la probabilidad de 0.304 de que un graduado en administración de empresas este empleado en un campo distinto al de su área principal de estudios.

3.2.3 Probabilidad subjetiva

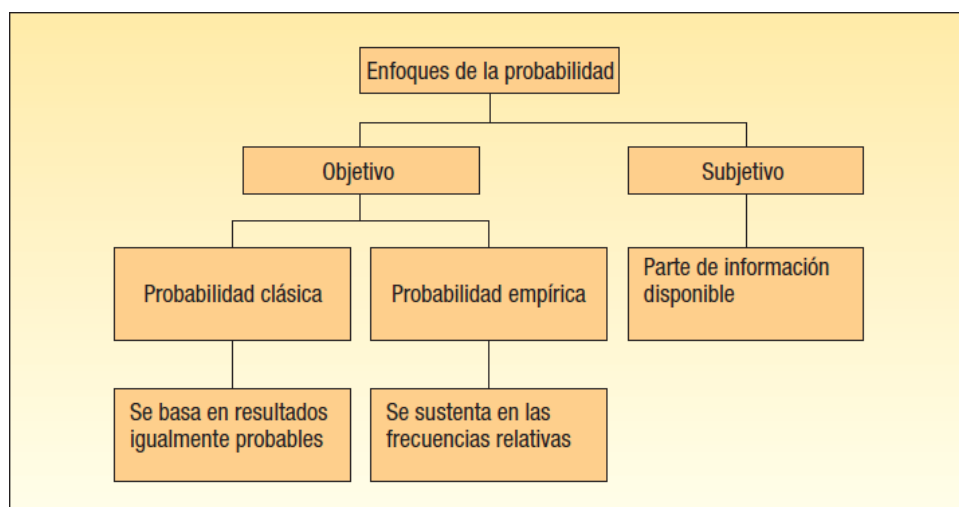
Si existe poca o ninguna experiencia en la cual se puede basar una probabilidad, de todas formas, se puede obtener una probabilidad en forma subjetiva. Se trata de evaluar opiniones disponibles y otra información subjetiva para después estimar o asignar la probabilidad; a este concepto se le denomina probabilidad subjetiva.

Probabilidad subjetiva

Posibilidad (probabilidad) de que suceda un evento específico, asignado por una persona con base en cualquier información que disponga.

Como ejemplos de probabilidad subjetiva tenemos:

- Valorar la posibilidad de que usted obtenga la calificación de 20/20 en el examen final en Estadística Descriptiva.
- Calcular la probabilidad de que Claro, pierda su lugar número 1 de aceptación respecto a su competidor principal Movi, dentro de 2 años.

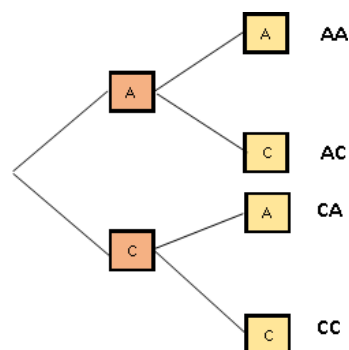


EJERCICIOS

1. Algunas personas están a favor de reducir los beneficios del Seguro Social a fin de lograr un presupuesto equilibrado, en tanto que otras están en contra. Se seleccionaron dos personas y se han de registrar sus opiniones. Mencione los resultados posibles.

Resultados posibles: A (a favor) C (en contra)

	Personas	
Resultados	1	2
1	A	A
2	A	C
3	C	A
4	C	C



- ¿Encuentre la probabilidad de que al menos una persona conteste en contra?

$$P(E) = \frac{n(E)}{n(N)}$$

$$P(\text{al menos una en contra}) = \frac{3}{4} = 0.75 = 75\%$$

- ¿Encuentre la probabilidad de que las 2 personas contesten a favor?

$$P(2 \text{ personas a favor}) = \frac{1}{4} = 0.25 = 25\%$$

- Una encuesta en una clase de 34 estudiantes de una escuela de Administración, revelo la siguiente selección de carreras:

Contaduría	10
Finanzas	5
Sistemas de información	3
Administración	6
Mercadotecnia	10

Suponga que selecciona a un (o una) estudiante y observa su opción profesional.

- ¿Cuál es la probabilidad de que él o ella estudie la carrera de administración?

$$P(A) = \frac{6}{34}$$

- ¿Qué concepto de probabilidad utilizo para hacer esta estimación?

SOLUCION:

$$a) P(E) = \frac{\text{Número de veces que el evento ocurrió en el pasado}}{\text{Número total de observaciones}}$$

$$P(E) = \frac{6}{34}$$

- Probabilidad Empírica

3. En cada uno de los siguientes casos indique si se utiliza la probabilidad clásica, la empírica o la subjetiva.
- a) Una jugadora de basquetbol realiza 30 canastas en 50 tiros de falta. La probabilidad que efectúe bien el próximo tiro es $30/50=0.60$
 - b) Se formó un comité de alumnos de siete miembros para estudiar asuntos ambientales. ¿Cuál es la probabilidad de que uno de ellos sea elegido como vocero? $=1/7$
 - c) Usted compra uno de los 5 millones de boletos que Lotto Canadá que vendió para un sorteo. ¿Cuál es la probabilidad de que gane el premio principal de un millón de dólares? $1/5\,000\,000$
 - d) La probabilidad de que ocurra un sismo en el Norte de California en los próximos 10 años, es de 0.80

SOLUCION

a) Estadística Empírica

b) Estadística Clásica

c) Estadística Clásica

d) Estadística Subjetiva

4. Antes de efectuar una encuesta en todo el país se seleccionaron 40 personas para probar el cuestionario. Una pregunta acerca de si debe o no legalizarse el aborto, requiere una respuesta de sí o no.
- a) ¿Cuál es el experimento?
 - b) Mencione un posible evento.
 - c) Diez de los 40 aspirantes se declararon a favor de la legalización. Con base de estas respuestas muestrales, ¿Cuál es la probabilidad de que una persona especifique este a favor de tal legalización?
 - d) ¿Qué concepto de probabilidad ilustra esto?
 - e) ¿Cada uno de los resultados posibles son igualmente probables y mutuamente excluyentes?

SOLUCION

- a) ¿Cuál es el experimento?

La encuesta de 40 personas acerca del aborto

- b) Mencione un posible evento.

Ejemplo. 20 personas respondieron que si están de acuerdo con la legalización del aborto

- c) Diez de los 40 aspirantes se declararon a favor de la legalización. Con base de estas respuestas muestrales, ¿Cuál es la probabilidad de que una persona específica este a favor de tal legalización?

$$P(A) = \frac{10}{40} = 0.25$$

- d) ¿Qué concepto de probabilidad ilustra esto?

Estadística Empírico

- e) ¿Cada uno de los resultados posibles son igualmente probables y mutuamente excluyentes?

Los eventos no son iguales, pero si mutuamente excluyentes.

Ejercicios propuestos

- Se eligió una muestra de 40 ejecutivos de la industria del petróleo para someter a prueba un cuestionario. Una pregunta relacionada con cuestiones ambientales requería un sí o un no.

- a) ¿En qué consiste el experimento?

- b) Indique un posible evento.

- c) Diez de los 40 ejecutivos respondieron que sí. Con base en estas respuestas de la muestra, ¿cuál es la probabilidad de que un ejecutivo de la industria del petróleo responda que sí?

- d) ¿Qué concepto de probabilidad se ilustra?

3.3 Reglas de probabilidad

Definida la probabilidad y descrito los diferentes enfoques, se estudiará la combinación de eventos con la aplicación de las reglas de adición y multiplicación.

$$P(E) = \frac{n(E)}{n(N)} \quad 0 \leq P(E) \leq 1$$

Reglas de adición

3.3.1 Regla especial de Adición

Para aplicar la regla especial de adición, los **eventos deben ser mutuamente excluyentes**; lo que significa que cuando ocurre un evento, ninguno de los otros puede ocurrir al mismo tiempo.

- Para explicar la regla consideremos el experimento de lanzar un dado y sean los evento A “un número 4 o mayor” y un B “número 2 o menor”; los resultados para el primer grupo son {4,5 y 6} no pueden estar en el segundo grupo {1 y 2}.
- A la vez que un producto que sale de un proceso industrial no puede ser satisfactorio y defectuoso a la vez.

Si dos eventos A y B son mutuamente excluyentes, la regla especial de adición indica que la probabilidad de que ocurra uno u otro de los eventos, es igual a la suma de sus probabilidades.

REGLA ESPECIAL DE ADICIÓN $P(A \text{ o } B) = P(A \cup B) = P(A) + P(B)$

Para tres eventos mutuamente excluyentes denotados por A, B y C se tiene:

$$P(A \text{ o } B \text{ o } C) = P(A \cup B \cup C) = P(A) + P(B) + P(C)$$

EJEMPLO 1: Una maquina automática llena bolsas de plástico con una mezcla de frijoles, brócoli y otras legumbres. La mayoría de las bolsas contiene el peso correcto, pero debido a ligeras variaciones en el tamaño de las verduras, un paquete puede tener un peso ligeramente menor o mayor. Una verificación de 4.000 paquetes llenados el mes pasado indico:

Peso	Evento	Número de paquetes	Probabilidad de ocurrencia
Con peso menor	A	100	$P(A)=100/4000=0.025$
Satisfactorio	B	3600	$P(B)=3600/4000=0.9$
Con peso mayor	C	300	$P(C)=300/4000=0.075$
		4000	1

$$P(E) = \frac{n(E)}{n(N)} \quad 0 \leq P(E) \leq 1$$

¿Cuál es la probabilidad de que un paquete especial tenga un **peso menor o mayor**?

SOLUCION:

$A = \text{evento peso menor}$

$C = \text{evento peso mayor}$

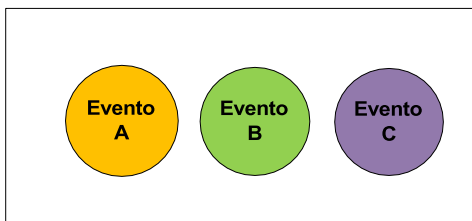
$$P(A \text{ o } C) = P(A \cup C) = P(A) + P(C) = 0.025 + 0.075 = 0.10$$

- Los eventos A, B y C son mutuamente excluyentes, es decir un paquete cualquiera únicamente puede tener peso menor, mayor o satisfactorio al mismo tiempo; además:
- $P(A \text{ o } B \text{ o } C) = P(A) + P(B) + P(C) = 1.000$

Diagrama de Venn

El concepto de mutuamente excluyente y otras reglas diversas para combinar probabilidades puede visualizarse empleando este dispositivo.

Mutuamente excluyente.



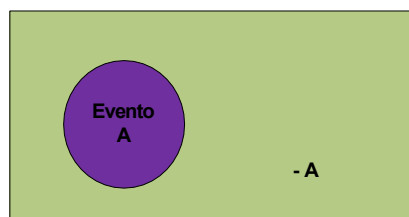
- No existe superposición de eventos.
- La probabilidad de que una bolsa de verduras mixta seleccionada sea de peso menor $P(A)$, más la probabilidad que no sea una bolsa con peso menor, que se escribe $P(\sim A)$ y se lee “no A” y es el **complemento de A**, debe ser lógicamente igual a 1; que se expresa:

$$P(A) + P(\sim A) = 1 \quad \text{expresada como} \quad P(A) = 1 - P(\sim A)$$

- ¿Cuál es la probabilidad de que un paquete cualquiera **no penga peso satisfactorio**?

$$P(B) + P(\sim B) = 1 \quad P(\sim B) = 1 - 0.90 = 0.10$$

A esto se le denomina **regla del complemento**; que se lo utiliza para determinar que ocurra un evento restando del número 1 la probabilidad de que no ocurra. El correspondiente diagrama de Venn del complemento se representa:

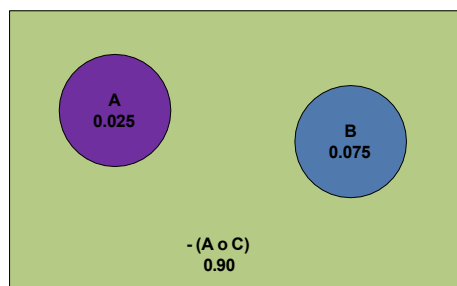


EJEMPLO: Demostrar utilizando el diagrama de Venn que una bolsa de legumbres mixtas tenga un peso satisfactorio.

$$\text{Recuerde: } P(A \text{ o } C) = P(A) + P(C) = 0.025 + 0.075 = 0.10$$

Entonces utilizando la regla del complemento tenemos que:

$$P(B) = 1 - P(A \text{ o } C) = 1 - 0.10 = 0.90$$



- La regla del complemento es muy importante en el estudio de la probabilidad. Resulta conveniente calcular la probabilidad de que ocurra un evento determinando la probabilidad de que no suceda y restando 1.

Ejercicios propuestos:

1. Los eventos A y B son mutuamente excluyentes. Suponga que $P(A)=0.30$ y $P(B)=0.20$. ¿Cuál es la probabilidad de que ocurran ya sea A o B? ¿Cuál es la probabilidad de que ni A ni B sucedan?

$$P(A \cup B) = P(A) + P(B)$$

$$P(A \cup B) = 0.30 + 0.20 = 0.50$$

$$P(A \cup B) + \sim P(A \cup B) = 1$$

$$\sim P(A \cup B) = 1 - P(A \cup B) = 1 - 0.50 = 0.50$$

2. El presidente de la junta directiva afirma: “Hay 50% de posibilidades de que esta compañía obtenga utilidades; 30% de que termine sin pérdidas ni ganancias y 20% de que pierda dinero durante el próximo trimestre.”

A=obtenga utilidades $P(A)=0.50$

B=sin pérdidas ni ganancias $P(B)=0.30$

C=que pierda dinero $P(C)=0.20$

a) Aplique una de las reglas de la adición para determinar la probabilidad de que la compañía no pierda dinero el siguiente trimestre.

$$P(A \cup B) = 0.50 + 0.30 = 0.80$$

b) Aplique la regla del complemento para determinar la probabilidad de que no pierda dinero el próximo trimestre.

$$P(C) + P(\sim C) = 1 \quad P(\sim C) = 1 - 0.20 = 0.80$$

EJEMPLO: Se ha de entrevistar a un grupo selecto de empleados de ACME con respecto a un plan de pensiones. Se efectuarán entrevistas detalladas a cada uno de los empleados seleccionados en la muestra. Estos se clasificaron como sigue:

Clasificación	Evento	Número de Empleados
Supervisores	A	120
De mantenimiento	B	50
De producción	C	1460
Gerencia	D	302
Secretaria	E	68

$$P(E) = \frac{n(E)}{n(N)} \quad 0 \leq P(E) \leq 1$$

- a)Cuál es la probabilidad de que la primera persona seleccionada:
- ¿Sea empleado de mantenimiento o una secretaria?
 - ¿No sea de gerencia?
- b) Elabore un diagrama de Venn mostrando sus respuestas a la parte (a)
- c) ¿Los eventos en la parte (a) son complementarios o mutuamente excluyentes, o bien ambas clases?

SOLUCION:

- a)Cuál es la probabilidad de que la primera persona seleccionada:

Clasificación	Evento	Número de Empleados	Probabilidad de ocurrencia
Supervisores	A	120	0,06
De mantenimiento	B	50	0,025
De producción	C	1460	0,73
Gerencia	D	302	0,151
Secretaria	E	68	0,034
		2000	1

$$P(E) = \frac{n(E)}{n(N)} \quad 0 \leq P(E) \leq 1$$

- i. ¿Sea empleado de mantenimiento o una secretaria?

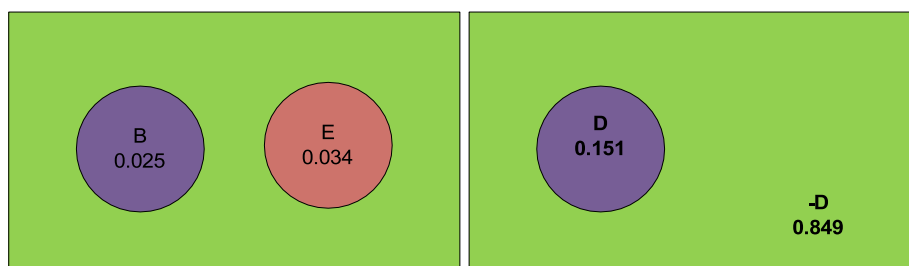
$$P(B \text{ o } E) = P(B) + P(E) = 0.025 + 0.034 = 0.059$$

- ii. ¿No sea de gerencia?

$$P(D) + P(\sim D) = 1 \quad P(\sim D) = 1 - P(D)$$

$$P(\sim D) = 1 - 0.151 = 0.849$$

- b) Elabore un diagrama de Venn mostrando sus respuestas a la parte (a)



- c) ¿Los eventos en la parte (a) son complementarios o mutuamente excluyentes, o bien ambas clases?

No son complementarios, si no mutuamente excluyentes.

3.3.2 Regla general de la adición

Los resultados de un experimento pueden no ser mutuamente excluyentes. Por ejemplo, suponga que el departamento de Educación Física de la UCE, selecciono una muestra de 400 estudiantes y se les indago sobre la preferencia del deporte entre futbol o básquet, la misma que revelo que 250 preferían el futbol (A) y 200 el básquet (B). ¿Cuál es la probabilidad de que un estudiante prefiera el futbol o el básquet?

B=evento que le guste el básquet (200)

A= evento que le guste el futbol (250)

Si utilizamos la regla especial de la adición tenemos:

$$P(A \text{ o } B) = P(A) + P(B) = \frac{250}{400} + \frac{200}{400} = 0.625 + 0.50 = 1.125 \text{ XXXXX}$$

Pero recuerde, la probabilidad no puede ser mayor que 1. La explicación es que en la encuesta 100 estudiantes expresaron tener preferencia por los dos deportes y la encuesta está registrando 2 veces; entonces la probabilidad de que un estudiante prefiera el futbol o el básquet está dada por:

$$P(A \text{ o } B) = P(A) + P(B) - P(A \cap B) \quad \text{Regla general de la adición}$$

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

$$P(A \text{ o } B) = \frac{250}{400} + \frac{200}{400} - \frac{100}{400} = 0.625 + 0.50 - 0.25 = 0.875$$

Cuando los eventos se superponen se llama probabilidad conjunta.

Probabilidad conjunta: $P(A \text{ y } B)$ $P(A \cap B)$

Es la probabilidad que mide la posibilidad de que dos o más eventos ocurran en forma simultánea. La regla general de la adición se le utiliza para combinar eventos que no son mutuamente excluyentes; teniendo:

REGLA GENERAL DE LA ADICION: $P(A \text{ o } B) = P(A) + P(B) - P(A \text{ y } B)$

REGLA GENERAL DE LA ADICION: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

En la expresión $P(A \text{ o } B)$, el término “o” indica que A puede ocurrir o que B también puede suceder; además incluye la posibilidad de que A y B pueden ocurrir. Al término “o” se lo llama **inclusivo**.

EJEMPLO

1. ¿Cuál es la probabilidad de que una carta elegida al azar de una baraja latina sea una K o una de corazones rojos?

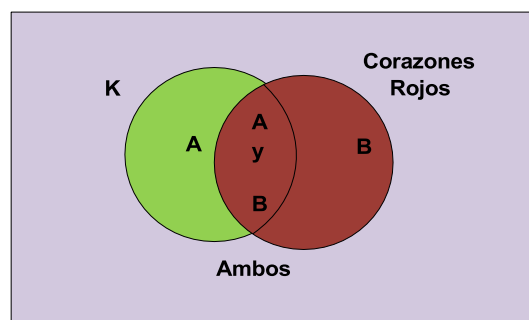
A = Evento que sea K

B = Evento que sea de corazones rojos

$$P(A \text{ o } B) = P(A) + P(B) - P(A \text{ y } B)$$

$$P(A \text{ o } B) = \frac{4}{52} + \frac{13}{52} - \frac{1}{52} = \frac{16}{52} = 0.3077$$

Diagrama de Venn:



2. Un estudiante toma dos cursos, historia y matemáticas. La probabilidad de que pase el curso de historia es de 0.60 y la de que apruebe el de matemáticas es de 0.70. La probabilidad de pasar ambos es de 0.50. ¿Cuál es la probabilidad de pasar por lo menos uno?

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

$$P(A \cup B) = 0.60 + 0.70 - 0.50 = 0.80$$

Ejercicios propuestos:

1. Un estudio de 200 empresas de publicidad reveló los siguientes ingresos después de impuestos:

Ingreso después de impuestos	Número de empresas
Menos de \$1 millón	102
De \$1 millón a \$20 millones	61
\$20 millones o más	37

- a) ¿Cuál es la probabilidad de que una empresa de publicidad seleccionada al azar tenga un ingreso después de impuestos menores a \$1 millón?
- b) ¿Cuál es la probabilidad de que una empresa de publicidad seleccionada al azar tenga un ingreso después de impuestos entre \$1 millón y \$20 millones o un ingreso de \$20 millones o más? ¿Qué regla de probabilidad aplicó?
2. Sean $P(X) = 0.55$ y $P(Y) = 0.35$. Suponga que la probabilidad de que ambos ocurran es de 0.20. ¿Cuál es la probabilidad de que X o Y ocurran?
- $$P(X \text{ o } Y) = P(X) + P(Y) - P(X \text{ y } Y) = 0.55 + 0.35 - 0.20 = 0.70$$
3. Las probabilidades de los eventos A y B son 0.20 y 0.30, respectivamente. La probabilidad de que A y B ocurran es de 0.15. ¿Cuál es la probabilidad de que A o B ocurran?
4. Una encuesta sobre tiendas de comestibles del sureste de Estados Unidos reveló que 40% tenían farmacia, 50% florería y 70% salchichonería. Suponga que 10% de las tiendas cuentan con los tres departamentos, 30% tienen tanto farmacia como salchichonería, 25% tienen florería y salchichonería y 20% tienen tanta farmacia como florería.

Denominación	Evento	Probabilidad
Farmacia	A	$P(A)=0.40$
Florería	B	$P(B)=0.50$
Salchichería	C	$P(C)=0.70$
Farmacia, florería, salchichería	AyByC	$P(AyByC)=0.10$
Farmacia, salchichería	AyC	$P(AyC)=0.30$
Florería, salchichería	ByC	$P(ByC)=0.25$
Farmacia, florería	AyB	$P(AyB)=0.20$

$$P(AyByC) = P(A \cap B \cap C) = 0.10$$

- a) ¿Cuál es la probabilidad de seleccionar una tienda de manera aleatoria y hallar que cuenta con farmacia y florería?

$$P(AyB) = 0.20$$

- b) ¿Cuál es la probabilidad de seleccionar una tienda de manera aleatoria y hallar que cuenta con farmacia y salchichonería?

$$P(AyC) = 0.30$$

c) Los eventos “seleccionar una tienda con salchichonería” y “seleccionar una tienda con farmacia”, ¿son mutuamente excluyentes?

No son mutuamente excluyentes

d) ¿Qué nombre se da al evento “seleccionar una tienda con farmacia, florería y salchichonería”?

$P(AyByC) = 0.10$ Son complementarios

e) ¿Cuál es la probabilidad de seleccionar una tienda que no incluya los tres departamentos?

$P(AyByC) + \sim P(AyByC) = 1$ $\sim P(AyByC) = 1 - 0.10 = 0.90$

f) ¿Cuál es la probabilidad de seleccionar una tienda que tenga farmacia o salchichonería?

$P(AoC) = P(A) + P(C) - P(AyC) = 0.40 + 0.70 - 0.30 = 0.80$

5. Un banco local informa que 80% de sus clientes tiene cuenta de cheques; 60% tiene cuenta de ahorros y 50% cuenta con ambas. Si se elige un cliente al azar, ¿cuál es la probabilidad de que el cliente tenga ya sea una cuenta de cheques o una cuenta de ahorros? ¿Cuál es la probabilidad de que el cliente no tenga una cuenta de cheques ni una de ahorros?

A = cuenta cheques $P(A) = 0.80$

B = cuenta de ahorros $P(B) = 0.60$

$A \cap B$ = cuenta de ahorros y cheques $P(AyB) = P(A \cap B) = 0.50$

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

$P(A \cup B) = 0.80 + 0.60 - 0.50 = 0.90$

$P(A \cup B) + \sim P(A \cup B) = 1$

$\sim P(A \cup B) = 1 - 0.90 = 0.10$

3.4 Reglas de multiplicación.

3.4.1 Regla especial de la multiplicación:

La regla especial de la multiplicación requiere que los eventos A y B sean independientes; que significa que la ocurrencia de uno no afecta la probabilidad del otro; o bien que, si los eventos A y B son independientes, la ocurrencia de A no altera la probabilidad de B.

Independiente

Se expresa esto cuando la ocurrencia de un evento no tiene efecto en la probabilidad de ocurrencia de cualquier otro.

Si A y B son independientes, la probabilidad de que ocurra A y B está dada por:

$$\text{REGLA ESPECIAL DE LA MULTIPLICACION} \quad P(A \text{ y } B) = P(A) \cdot P(B)$$

Por ejemplo, al lanzar dos monedas al aire el resultado de la primera moneda (cara o cruz) no afecta al resultado de la segunda moneda (cara o cruz); con lo que se evidencia la independencia de los eventos.

Para tres eventos independientes tendremos:

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

EJEMPLO: Si se lanzan tres monedas al aire. ¿Cuál es la probabilidad de que las tres caigan cruz?

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

$$P(A \text{ y } B \text{ y } C) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8} = 0.125$$

EJEMPLO: Se lanza una moneda y luego un dado ¿Cuál es la probabilidad de que salga una cara y un número mayor que 4 ?

A = moneda caiga cara

B=dado sea un número mayor que 4

$$P(A \text{ y } B) = P(A) * P(B) = \frac{1}{2} * \frac{2}{6} = \frac{1}{6} = 0.1667$$

3.5 Eventos dependientes

Si dos eventos no son independientes se dice que son dependientes. Suponga que tiene 15 transistores en una caja y se sabe que 5 están defectuosos; se selecciona uno. Resulta lógico que la probabilidad de escoger un transistor con defecto es 5/15 y uno no defectuoso es 10/15. Después se elige el segundo transistor sin devolver el primero a la caja. La probabilidad de que este esté defectuoso **depende** de que el primer transistor seleccionado este bueno o no. La probabilidad de que también el segundo transistor este defectuoso es:

- 4/14, si el primer transistor seleccionado fuera defectuoso;
- 5/14, si el primer transistor seleccionado no fuera defectuoso

A=evento no defectuoso (10)

B=evento sea defectuoso (5)

¿Cuál es la probabilidad de que el segundo transistor seleccionado sea defectuoso dado que el primero no es defectuoso?

$$P(B/A) = \frac{5}{14} \text{ Probabilidad condicional}$$

¿Cuál es la probabilidad de que el segundo transistor seleccionado no sea defectuoso dado que el primero es defectuoso?

$$P(B/A) = \frac{10}{14}$$

A la fracción 5/14, se le denomina probabilidad condicional porque su valor tiene tal característica (dependiente de estas) respecto de la primera selección de la caja; que se haya seleccionado un transistor defectuoso o satisfactorio.

3.5.1 Probabilidad condicional

Probabilidad de que un evento en particular ocurra, dado que otro evento haya acontecido.

Si se desea conocer la probabilidad de seleccionar 2 transistores defectuosos uno después de otro, se aplica la ley general de la multiplicación.

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

Ejercicios propuestos

1. Suponga que $P(A)=0.40$ y $P(B/A)=0.30$ ¿Cuál es la probabilidad conjunta de A y B?

$$P(A \text{ y } B) = P(A) \cdot P(B/A) = 0.40 \cdot 0.30 = 0.12$$

EJEMPLO: Consideremos el ejemplo de encontrar la probabilidad de extraer dos transistores de una caja que contiene 15 transistores de los cuales 3 están defectuosos. Considerar para el experimento la no reposición del transistor de la primera extracción.

$A = \text{no defectuoso}$

$B = \text{defectuoso}$

El experimento consiste en extraer dos transistores. Encontrar la probabilidad de que el segundo transistor sea defectuoso dado que el primer transistor no fue defectuoso.

$$P(B/A) = 3 / 14$$

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

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

$$P(A \text{ y } B) = \frac{12}{15} \cdot \frac{3}{14}$$

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

$$P(B \text{ y } A) = \frac{3}{15} \cdot \frac{12}{14}$$

Se puede aplicar la regla general de la multiplicación para tres eventos como sigue:

$$P(A \text{ y } B \text{ y } C) = P(A) \cdot P(B/A) \cdot P(C/A \text{ y } B)$$

EJEMPLO: Consideremos el mismo ejemplo para la probabilidad de que los 3 sean defectuosos.

$$P(A \text{ y } B \text{ y } C) = (3/15)(2/14)(1/13) = \frac{1}{455} = 0.022$$

EJEMPLO: La junta de directores de INDUTEX está formada por ocho hombres y cuatro mujeres. Se seleccionará un comité de cuatro miembros, en forma aleatoria, para recomendar a un nuevo presidente de la compañía.

- ¿Cuál es la probabilidad de que sean mujeres los cuatro miembros del comité de investigación?
- ¿Cuál es la probabilidad de que los cuatro miembros sean hombres?
- ¿La suma de las probabilidades para a y b es igual a 1? Explique su respuesta.

SOLUCIÓN.

A = evento que sea 4 mujeres.

B = evento que sea 4 hombres

- ¿Cuál es la probabilidad de que sean mujeres los cuatro miembros del comité de investigación?

$$P(A) = \left(\frac{4}{12}\right) * \left(\frac{3}{11}\right) * \left(\frac{2}{10}\right) * \left(\frac{1}{9}\right) = 0.002$$

- ¿Cuál es la probabilidad de que los cuatro miembros sean hombres?

$$P(B) = \left(\frac{8}{12}\right) \left(\frac{7}{11}\right) \left(\frac{6}{10}\right) \left(\frac{5}{9}\right) = 0.141$$

- ¿La suma de las probabilidades para a y b es igual a 1? Explique su respuesta.
- ¿Cuál es la probabilidad de que dos primeros sean hombres y los 2 siguientes sean mujeres?

$$P(AyB) = \left(\frac{8}{12}\right) \left(\frac{7}{11}\right) \left(\frac{4}{10}\right) \left(\frac{3}{9}\right) = 0.0566$$

- ¿Cuál es la probabilidad de que el primero sea hombre, segunda mujer, tercero hombre, cuarto mujer?

$$P(AyByCyD) = \frac{8}{12} * \frac{4}{11} * \frac{7}{10} * \frac{3}{9} = 0.0566$$

f) ¿La suma de las probabilidades para 1 y 2 es igual a 1? Explique su respuesta. 101

$$P(A) + P(B) = 0.002 + 0.141 = 0.143$$

Es diferente de 1 porque, existen otras probabilidades de formación del comité, por ejemplo 3 mujeres y un hombre.

Ejercicio propuesto

1. Un banco local informa que 80% de sus clientes tiene cuenta de cheques; 60% tiene cuenta de ahorros y 50% cuenta con ambas. Si se elige un cliente al azar, ¿cuál es la probabilidad de que el cliente tenga ya sea una cuenta de cheques o una cuenta de ahorros? ¿Cuál es la probabilidad de que el cliente no tenga una cuenta de cheques ni una de ahorros?
2. Un inversionista cuenta con tres acciones ordinarias. Cada una de ellas, independiente de las demás, tiene la misma probabilidad de: 1) incrementar su valor; 2) bajar su valor; 3) permanecer con el mismo valor. Elabore una lista de los posibles resultados de este experimento. Calcule la probabilidad de que por lo menos dos de las acciones aumenten de valor.

3.6 Tablas de contingencias

Una tabla de contingencia es una tabulación cruzada, que resume simultáneamente dos variables de interés. Por ejemplo:

- Los estudiantes en una universidad se clasifican por género y lugar en la clase.
- Un producto se clasifica como aceptable o inaceptable y de acuerdo con el turno (matutino, vespertino, nocturno) en el que se le fabrica.
- Un votante de una escuela que lleva a cabo un referendo para otorgar becas se clasifica de acuerdo con su afiliación partidista (demócrata, republicano u otro), y el número de hijos que asisten a la escuela del distrito (0, 1, 2, etcétera).

TABLA DE CONTINGENCIAS Tabla que se utiliza para clasificar observaciones de una muestra, de acuerdo con dos o más características identificables.

EJEMPLO: Una encuesta de ejecutivos se enfocó sobre la lealtad de la empresa. Una de las preguntas planteadas fue: "¿Si otra compañía le hiciera una oferta igual o ligeramente mejor que la de su puesto actual, permanecería con la empresa o tomaría el otro empleo?" Las respuestas de los 200 ejecutivos de la encuesta se clasificaron en forma cruzada con su tiempo de servicio a la compañía.

	Tiempo de servicio				
Lealtad	Menos de	1 a 5	6 a 10	Más de	Total
	1 año B1	Años B2	Años B3	10 años B4	
Se quedaría A1	10	30	5	75	120
No se quedaría A2	25	15	10	30	80
Total	35	45	15	105	200

- a) ¿Cuál es la probabilidad de seleccionar al azar un ejecutivo que es leal a la empresa (se quedaría) **y** que tiene más de 10 años de servicio?

$$P(A1 \text{ y } B4) = P(A1) * P\left(\frac{B4}{A1}\right) = \frac{120}{200} * \frac{75}{120}$$

- b) ¿Cuál es la probabilidad de seleccionar al azar un ejecutivo que no es leal a la empresa (no se quedaría) **y** que tiene 6 a 10 años de servicio?

$$P(A2 \text{ y } B3) = P(A2) * P\left(\frac{B3}{A2}\right) = \frac{80}{200} * \frac{10}{80}$$

- c) ¿Cuál es la probabilidad de seleccionar al azar un ejecutivo que no es leal a la empresa (no se quedaría) **o** que tiene 6 a 10 años de servicio?

$$P(A2 \text{ o } B3) = P(A2) + P(B3) - P(A2 \cap B3)$$

$$P(A2 \text{ o } B3) = \frac{80}{200} + \frac{15}{200} - \frac{10}{200}$$

SOLUCIÓN:

Analice tiene dos eventos simultáneos.

A1 = es leal a la empresa (se quedaría)

B4 = Tiene más de 10 años de servicio

P(B/A) = probabilidad condicional de que un ejecutivo de la empresa tenga más de 10 años **y** permanezca en la empresa.

$$P(A1 \text{ y } B4) = P(A1) * P(B4/A1)$$

$$P(A1 \text{ y } B4) = \left(\frac{120}{200}\right) * \left(\frac{75}{120}\right) = 0.375$$

- d) ¿Cuál es la probabilidad de seleccionar al azar un ejecutivo que tenga de 6 a 10 años de servicio **y** que no permanezca en la empresa al recibir una oferta igual o ligeramente mejor?

$$P(B3 \text{ y } A2) = P(B3) * P(A2/B3)$$

$$P(B3 \text{ y } A2) =$$

15 10
200 * 15

— —

- e) ¿Cuál es la probabilidad de seleccionar aleatoriamente un ejecutivo que no permanecería con la empresa **y** que tenga menos de un año de servicio?

$$P(A2 \text{ y } B1) = P(A2) * P(B1/A2) = \frac{80}{200} * \frac{25}{80}$$

- f) ¿Cuál es la probabilidad de seleccionar aleatoriamente un ejecutivo que no permanecería en la empresa **dado que** es un trabajador de 1 a 5 años en la empresa?

$$P(A2/B2) = \frac{15}{80}$$

- g) ¿Cuál es la probabilidad de seleccionar aleatoriamente un ejecutivo que no permanecería en la empresa **o** es un trabajador de 1 a 5 años en la empresa?

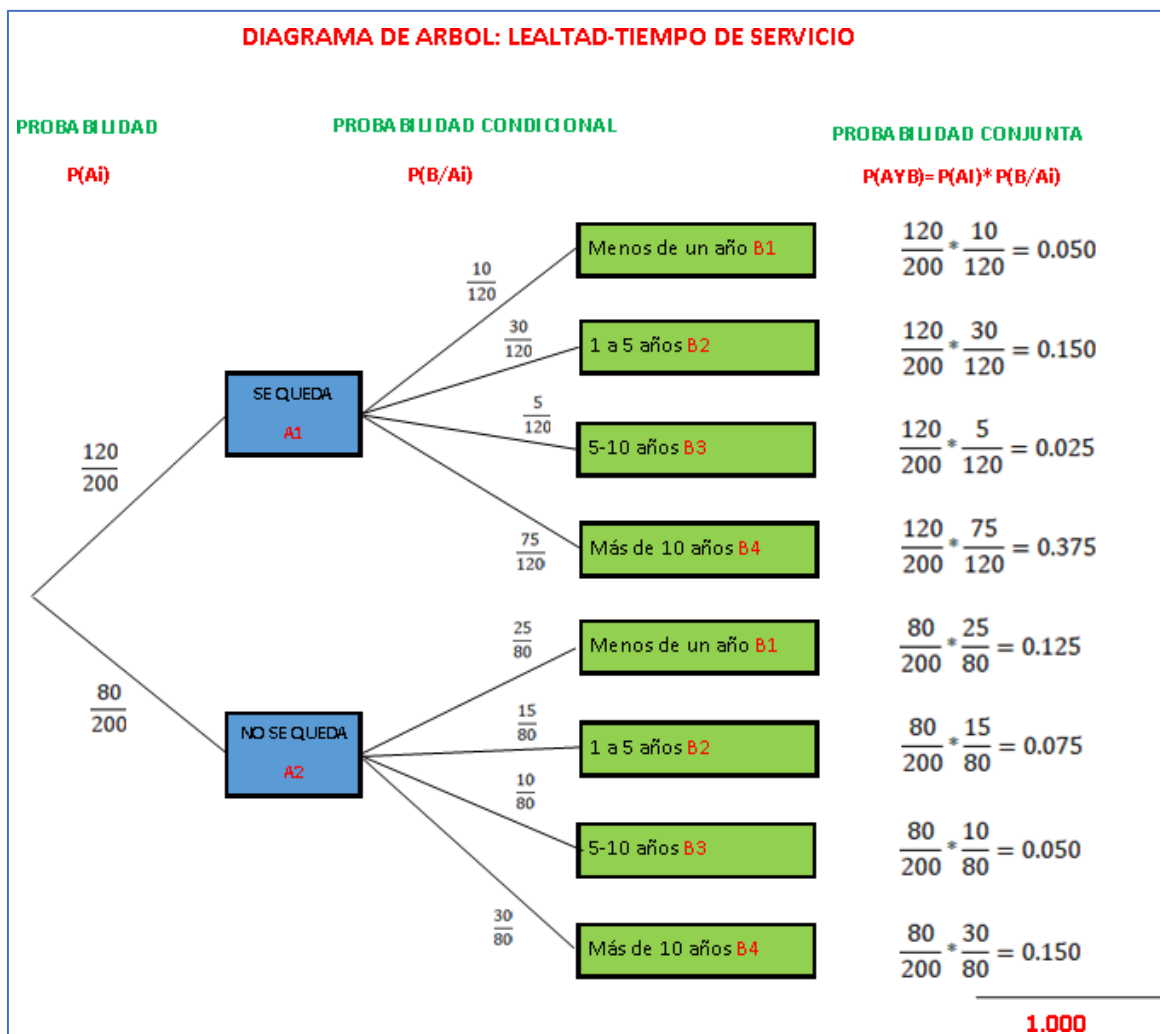
$$P(A2 \text{ o } B2) = P(A2) + P(B2) - P(A1 \cap B2)$$

$$P(A2 \text{ o } B2) = \frac{80}{200} + \frac{45}{200} - \frac{15}{200} = \frac{135}{200}$$

3.6.1 Diagramas de árbol (arborigramas)

Un diagrama de árbol o arborigrama es una representación gráfica útil para organizar cálculos que abarcan varias etapas. Cada segmento en el árbol es una etapa del problema. Las probabilidades escritas cerca de las ramas son las probabilidades condicionales del experimento.

Lealtad	Tiempo de servicio				Total
	Menos de	1 a 5	6 a 10	Más de	
	1 año(B1)	Años(B2)	Años(B3)	10 años(B4)	
Se quedaría (A1)	10	30	5	75	120
No se quedaría (A2)	25	15	10	30	80
	35	45	15	105	200



EJEMPLO PARA RESOLVER: Cada vendedor de QATARSA se califica “como abajo del promedio” o “arriba del promedio”, con respecto a su aptitud para las ventas. Además, cada uno también se clasifica respecto de su posibilidad de promoción en: regular, buena o excelente. En la tabla que sigue se presentan las clasificaciones de estos conceptos para los 500 vendedores.

Habilidad en ventas	Posibilidad de promoción			
	Regular B1	Buena B2	Excelente B3	Total
Por abajo del promedio A1	16	12	22	50
Promedio A2	45	60	45	150
Por arriba del promedio A3	93	72	135	300
Totales	154	144	202	500

a) ¿Cómo se denomina esta tabla?

Tabla de contingencia

- b) ¿Cuál es la probabilidad de que un vendedor seleccionado al azar tenga aptitud para las ventas por encima del promedio **y** excelente posibilidad de promoción?

$$P(A3 \text{ y } B3) = P(A3) * P(B3/A3) = \frac{300}{500} * \frac{135}{300}$$

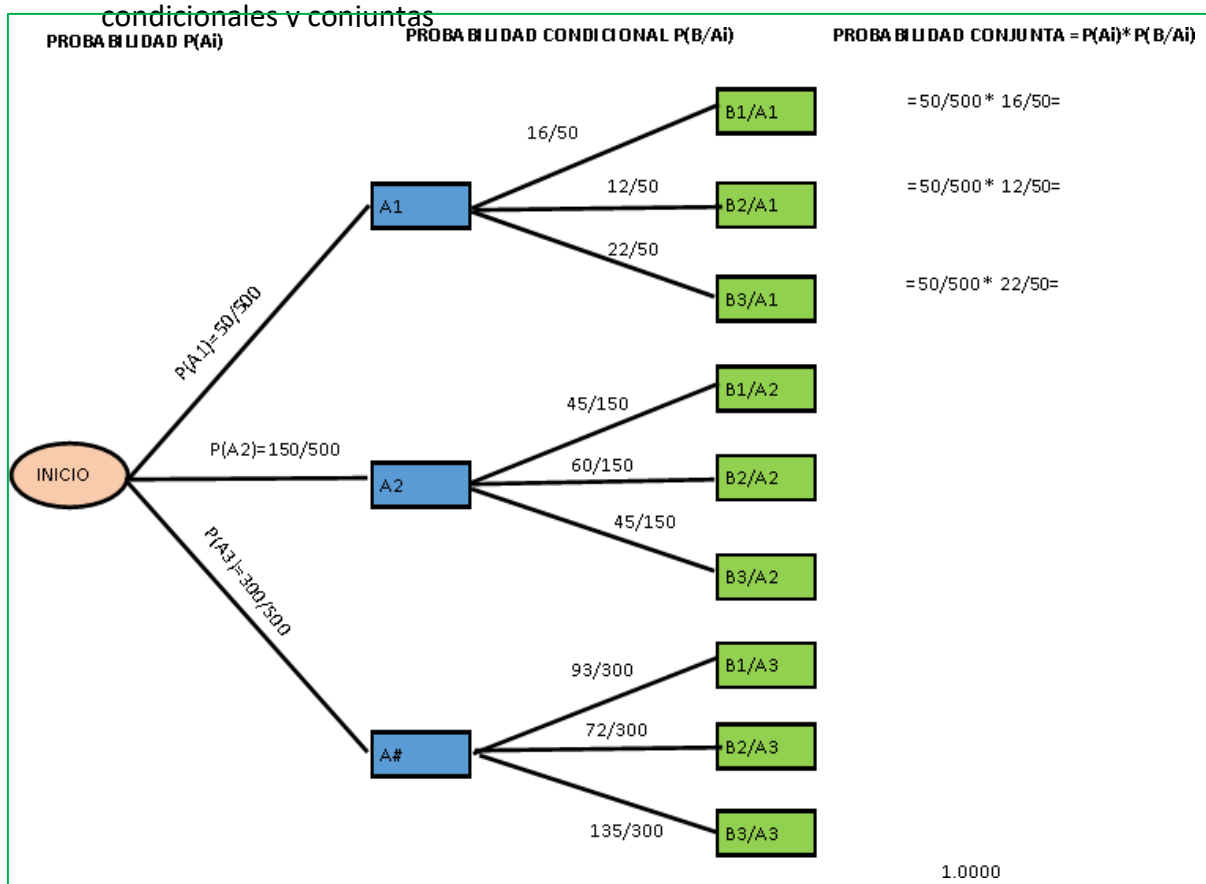
- c) ¿Cuál es la probabilidad de que un vendedor seleccionado al azar tenga aptitud para las ventas por debajo del promedio **y** buena posibilidad de promoción?

$$P(A1 \text{ y } B2) = P(A1) * P(B2/A1) = \frac{50}{500} * \frac{12}{50}$$

- d) ¿Cuál es la probabilidad de que un vendedor seleccionado al azar tenga aptitud para las ventas por debajo del promedio **dado** que tiene excelente posibilidad de promoción?

$$P(B3/A1) = \frac{22}{50}$$

- e) Trace un diagrama de árbol que muestre todas las probabilidades: normales, condicionales y conjuntas



3.6.2 Teorema de Bayes

En el siglo XVIII el reverendo Thomas Bayes, ministro presbiteriano inglés, planteo la siguiente cuestión: ¿Realmente existe Dios? Estando interesado en las matemáticas, intento desarrollar una fórmula para llegar a evaluar la probabilidad de que Dios exista, en base de la evidencia de la que disponía él en la Tierra. Más adelante Laplace afino el trabajo de Bayes y le dio el nombre de “Teorema de Bayes”. Que se puede expresar:

$$P(A_1/B) = \frac{P(A_1)P(B/A_1)}{P(A_1)P(B/A_1) + P(A_2)P(B/A_2)} \quad \textbf{Teorema de Bayes}$$

Explicación de Términos:

Ejemplo: Suponga que 5% de la población de XXX, un país del Tercer Mundo, padece una enfermedad que es originaria de ese lugar.

Sea A_1 , el “evento tiene la enfermedad”

A_2 el evento “no tiene la enfermedad”.

Por lo tanto, conocemos que, si seleccionamos una persona al azar del país XXX, la probabilidad de que tenga la enfermedad es 5%, o bien $P(A_1) = 0.05$ y la probabilidad que no tenga la enfermedad $P(A_2) = 1 - P(A_1) = 0.95$; a estas probabilidades se les **denomina probabilidad a priori**. Se les da este nombre porque la probabilidad se asigna antes de haber obtenido datos empíricos.

PROBABILIDAD A PRIORI: Es la probabilidad inicial con base en el nivel actual de información.

- En el país XXX, existe una técnica de diagnóstico para detectar la enfermedad, pero **no es muy exacta**. Sea B el evento “la prueba indica que la enfermedad está presente”. Considere que la evidencia histórica muestra que, si la persona realmente padece la enfermedad, la probabilidad de que la prueba indique la presencia de la misma es 0.90. Utilizando la definición de probabilidad condicional, podemos expresar: $P(B/A_1) = 0.90$

Considere ahora que la probabilidad de que una persona en realidad no tenga el padecimiento, pero que la prueba indique que el mismo está presente, es 0.15.
 $P(B/A_2) = 0.15$

Seleccionemos ahora en forma aleatoria a un habitante de XXX, al que se le aplica la prueba. Los resultados indican que la enfermedad está presente. ¿Cuál es la probabilidad de que realmente tenga la enfermedad? En forma simbólica, se desea determinar: $P(A_1/B)$, que se interpreta como:

P(tiene la enfermedad/los resultados son positivos). A esta se le denomina **probabilidad a posteriori**.

PROBABILIDAD A POSTERIORI: Es una probabilidad revisada con base en información adicional.

Utilizamos el Teorema de Bayes para determinar: $P(A_1/B)$

$$P(A_1/B) = \frac{P(A_1) * P(B/A_1)}{P(A_1) * P(B/A_1) + P(A_2) * P(B/A_2)}$$

$$P(A_1/B) = \frac{(0.05)(0.90)}{(0.05)(0.90) + (0.95)(0.15)} = \frac{0.045}{0.1875} = 0.24$$

$$P(A_2/B) = \frac{P(A_2) * P(B/A_2)}{P(A_2) * P(B/A_2) + P(A_1) * P(B/A_1)}$$

$$P(A_2/B) = \frac{0.95 * 0.15}{0.95 * 0.15 + 0.05 * 0.90} = 0.76$$

En el ejemplo presentado se incluyó A_1 y A_2 , como probabilidades a priori; si hay más de dos probabilidades a priori, el denominador del teorema de Bayes, necesita de más términos. Si la distribución probabilística a priori consiste en n eventos mutuamente excluyentes, el Teorema de Bayes queda como:

$$P(A_i/B) = \frac{P(A_i)P(B/A_i)}{P(A_1)P(B/A_1) + P(A_2)P(B/A_2) + \dots + P(A_n)P(B/A_n)}$$

Donde A_i se refiere a cualquiera de los n posibles resultados.

Los cálculos para el país XXX se pueden resumir en la siguiente tabla:

Evento A_i	Probabilidad	Probabilidad	Probabilidad	Probabilidad
	a priori	Condicional	Conjunta	a posterior
	$P(A_i)$	$P(B/A_i)$	$P(A_i \text{ y } B)$	$P(A_i/B)$
Tiene la enfermedad A_1	0.05	0.90	0.045	0.045/0.1875=0.24
No tiene la enfermedad A_2	0.95	0.15	0.1425	0.1425/.1875=0.76
			P(B)=0.1875	1.00

Ejemplo

1. Un fabricante de celulares (xxx) compra un cierto microchip, llamado LS-24, de tres proveedores. Un 30% de los microcircuitos LS-24 se compran a Hall Electronics(HE), 20% a Schuller Sales(SS), y el 50% restante a Crawford Components(CC). El fabricante tiene un historial extenso acerca de los tres abastecedores y sabe que el 3% de los microchips de HE son defectuosos, el 5% de SS son no aceptables, y un 4% de los de CC también tienen defectos.

Cuando tales microcircuitos LS-24 llegan al fabricante, se colocan directamente en un depósito, no son inspeccionados o identificados de algún modo por el proveedor. Un trabajador selecciona uno para instalación en un celular, lo encuentra defectuoso. ¿Cuál es la probabilidad de que haya sido fabricado por SS?

Resumen de información:

- Existen 3 eventos, o sea 3 proveedores.

A_1 = LS-24 fue comprado a HE

A_2 = LS-24 fue comprado a SS

A_3 = LS-24 fue comprado a CC

- **Las probabilidades a priori son:**

$P(A_1) = 0.30$ La probabilidad de que un chip sea fabricado por HE

$P(A_2) = 0.20$ La probabilidad de que un chip sea fabricado por SS

$P(A_3) = 0.50$ La probabilidad de que un chip sea fabricado por CC

- **La información adicional a posteriori es que el chip es defectuoso.**

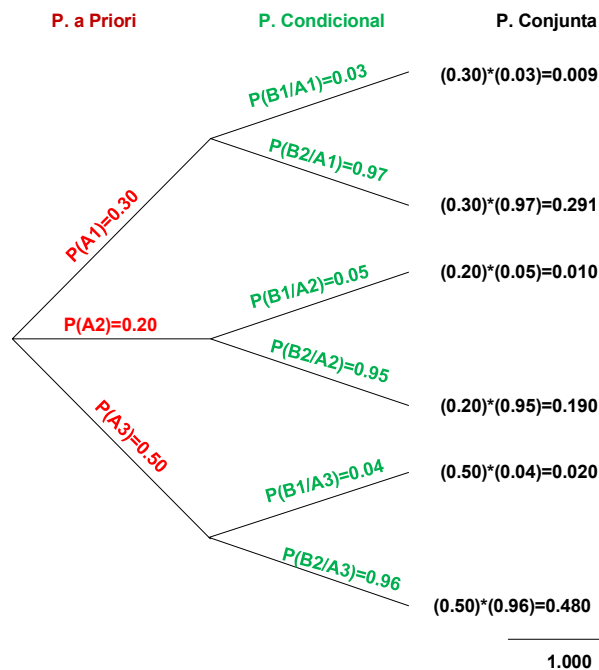
B_1 = es defectuoso

B_2 = no es defectuoso.

$P(B_1/A_1) = 0.03$

$P(B_1/A_2) = 0.05$

$P(B_1/A_3) = 0.04$



$$P(A_2/B_1) = \frac{0.01}{0.009 + 0.010 + 0.020} = 0.2564$$

- Cuando tales microcircuitos LS-24 llegan al fabricante, se colocan directamente en un depósito, no son inspeccionados o identificados de algún modo por el proveedor. Un trabajador selecciona uno para instalación en un celular, y no lo encuentra defectuoso. ¿Cuál es la probabilidad de que haya sido fabricado por HE?

$$P(A_1/B_2) = \frac{0.291}{0.291 + 0.19 + 0.48} = 0.3028$$

La información podemos resumirla en el cuadro siguiente:

Evento	Probabilidad	Probabilidad	Probabilidad	Probabilidad a
Ai	A Priori	Condicional	Conjunta	Posteriori
	P(Ai)	P(Bi/Ai)	P(Ai y Bi)	P(Ai/Bi)
HE A1	0.30	0.03	0.009	0.009/0.039=0.2308
SS A2	0.20	0.05	0.010	0.010/0.039=0.2564
CC A3	0.50	0.04	0.020	0.020/0.039=0.5128
			P(B)=0.039	10.000

Aplicando el Teorema de Bayes tenemos:

$$P(A_1/B_1) = \frac{P(A_1) * P(B_1/A_1)}{P(A_1) * P(B_1/A_1) + P(A_2) * P(B_1/A_2) + P(A_3) * P(B_1/A_3)}$$

$$2 \quad 1 \quad P(A_1) * P(B_1/A_1) + P(A_2) * P(B_1/A_2) + P(A_3) * P(B_1/A_3)$$

$$P(A_2/B_1) = \frac{(0.20)(0.05)}{(0.30)(0.03) + (0.20)(0.05) + (0.50)(0.04)} = \frac{0.010}{0.039} = 0.2564$$

c) ¿Cuál es la probabilidad de que haya sido fabricado por CC?

Ejercicio propuesto.

1. El departamento de crédito de la tienda departamental Lion's, California, informo que el 30% de sus ventas son en efectivo, 30% son pagadas con cheque en el momento de la adquisición y 40% son a crédito. Se tiene que 20% de las compras en efectivo, 90% de las pagadas con cheque, y 60% de las adquisiciones a crédito, son por más de \$50. La señora Tina Turner acaba de comprar un vestido nuevo que cuesta \$120. ¿Cuál es la probabilidad de que haya pagado en efectivo?

Evento Ai	P(Ai)	P(B/Ai)	P(Ai)*P(B/Ai)	P(Ai/B)
Efectivo A1	0.30	0.20	0.06	0.1052
Cheque A2	0.30	0.90	0.27	0.4737
Crédito A3	0.40	0.60	0.24	0.4211
			P(B)=0.57	1

$$P(A1/B)=0.1052$$

$$P(A1/B) = \frac{0.30 * 0.20}{0.30 * 0.20 + 0.30 * 0.90 + 0.40 * 0.60} = 0.1052$$

$$0.30 * 0.20 + 0.30 * 0.90 + 0.40 * 0.60$$

- Flashner Marketing Research, Inc., se especializa en la evaluación de las posibles tiendas de ropa para dama en centros comerciales. Al Flashner, el presidente, informa que evalúa las posibles tiendas como buenas, regulares y malas. Los registros de anteriores evaluaciones muestran que 60% fueron evaluados como buenos; 30% de las veces regulares y 10% de las ocasiones, malos. De los que fueron calificados como buenos, 80% hicieron mejoras el primer año; los que fueron calificados como regulares, 60% hicieron mejoras el primer año y de los que fueron mal evaluados, 20% hicieron mejoras el primer año. Connie's Apparel fue uno de los clientes de Flashner. Connie's Apparel hizo mejoras el año pasado. ¿Cuál es la probabilidad de que se le haya dado originalmente una mala calificación?

Probabilidad Ai	P(Ai)	P(B/Ai)	P(Ai)*P(B/Ai)	P(Ai/B)
Buenas A1	0.60	0.80	0.48	0.7059
Regulares A2	0.30	0.60	0.18	0.2647
Malas A3	0.10	0.20	0.02	0.0294
			P(B)=0.68	1

$$P(A3/B)=0.0294$$

3.7 Principios de conteo.

$$P(E) = \frac{n(E)}{n(N)} \quad 0 \leq P(E) \leq 1$$

Si el número de resultados posibles de un experimento es pequeño, resulta fácil enlistar y contar los eventos factibles, como por ejemplo el experimento de lanzar un dado, los resultados posibles son 6.

Pero cuando en número de resultados posibles son muchos, resultaría tedioso y a la vez complejo enlistar todos los resultados posibles; como por ejemplo enlistar todos los resultados de placas de circulación para la provincia de Pichincha. PTE049.

P	T	E	O	4	9
1	24	24	10	10	10 =

Para facilitar, el conteo se analizarán tres fórmulas a saber: Multiplicación, Permutación y Combinación.

3.7.1 Formula de la multiplicación

Si hay **m** modos de hacer una cosa y **n** formas de hacer otra, existen **mxn** formas de hacer ambas.

Número total de arreglos = (m)(n)

***lanzo dado---lanzo moneda---lanzo dado---lanzo moneda**

$$N=(6)(2)(6)(2)=144$$

Esta fórmula se puede extender a más de dos eventos. Sean los eventos m, n, o.

Número total de arreglos = (m)(n)(o)

Ejemplo 1: Un vendedor de automóviles desea anunciar que por \$20.000 usted puede comprar **un convertible**, **un sedán de 2 puertas** o **un modelo de 4 puertas**, con elección de **cubrerruedas deportivos** o **comunes**. ¿Cuántos arreglos diferentes de modelos y cubrerruedas pueden ofrecer el comerciante?

Evento 1 (m) = Tipo de auto

Evento 2 (n) = Cubrerruedas

Número total de arreglos = (m)(n)

$$\text{Número total de arreglos} = (3)*(2)=6$$

El vendedor podría encontrar la solución esquematizando la situación:

Convertible con cubreruedas comunes	Convertible con cubreruedas deportivas
Sedan de 2 puertas con cubreruedas comunes	Sedan de 2 puertas con cubreruedas deportivas
Sedan de 4 puertas con cubreruedas comunes	Sedan de 4 puertas con cubreruedas deportivas

En el ejemplo no ha sido difícil enlistar y contar las posibles combinaciones; pero si el vendedor decidiera ofrecer 9 modelos y 5 tipos de cubreruedas, esta posibilidad resulta ser demasiado tedioso. Para este caso el número de arreglos posibles es:

Número total de arreglos = $(9) \cdot (5) = 45$

Ejemplo 2: ¿De cuantas formas diferentes puede responderse un examen bajo cada una de las siguientes condiciones?

- El examen consiste en tres preguntas de opción múltiple con cuatro opciones para cada pregunta.

Número total de arreglos = $(4) \cdot (4) \cdot (4) = 64$

¿Cuál es la probabilidad de contestar correctamente su examen si contesta al azar?

$$P(E) = \frac{n(E)}{n(N)} \quad 0 \leq P(E) \leq 1$$

$$n(E) = (1) \cdot (1) \cdot (1) = 1$$

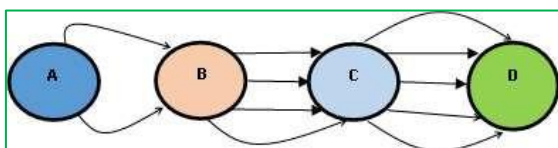
$$P(E) = \frac{1}{64} = 1.56\%$$

$$P(E) = \frac{3 \cdot 3 \cdot 3}{64} = 42.19\%$$

- El examen consiste en tres preguntas de opción múltiple con 4 opciones para cada una y cinco preguntas de falso-verdadero.

Número total de arreglos = $(4 \cdot 4 \cdot 4)(2 \cdot 2 \cdot 2 \cdot 2 \cdot 2) = 2.048$

Ejemplo 3: Dos caminos conectan las ciudades A y B, cuatro conectan a B y C, y cinco conectan a C y D. Para manejar de A a B, a C y luego a la ciudad D. ¿Cuántas rutas diferentes son posibles?



$$N = n_1 \cdot n_2 \cdot n_3$$

$$N = (2)(4)(5) = 40 \text{ rutas}$$

3.7.2 Permutaciones

La regla de la multiplicación se aplica para encontrar el número de arreglos posibles para dos o más grupos. La **fórmula de la permutación** sirve para determinar el número posible de arreglos cuando hay un solo grupo de objetos y donde el orden es importante.

PERMUTACION: Un arreglo o disposición de r objetos seleccionados a partir de un grupo único de n objetos posibles, donde el orden es importante y sin repetición.

$$\text{FORMULA DE PERMUTACION } nPr = \frac{n!}{(n-r)!}$$

$$6P2 = \frac{6!}{(6-2)!} = \frac{6!}{4!} = 30$$

$$9! = 9 * 8 * 7 * 6 * 5 * 4 * 3 * 2 * 1$$

Ejemplo 1: Supóngase que hay 8 máquinas disponibles, pero solo 3 espacios en el piso del taller donde se han de instalar tales maquinas. ¿De cuántos modos diferentes pueden colocarse las ocho en los tres espacios disponibles?

$$nPr = \frac{n!}{(n-r)!}$$

$$8P3 = \frac{8!}{(8-3)!} = 336$$

Ejemplo 2: Se van a utilizar los 10 números del 0 al 9, para crear cifrados de código de cuatro dígitos para identificar un artículo de ropa. El 1083 podría identificar una blusa azul talla mediana. No se permite repetición de los números. Es decir, no se puede tener el número 2334 por ejemplo. ¿Cuántos grupos distintos cifrados de código pueden establecerse?

$$nPr = \frac{n!}{(n-r)!}$$

$$10P4 = \frac{10!}{(10-4)!} = 5.040$$

Ejemplo 3: Un club tiene 20 miembros, deben ocuparse los puestos de presidente, vicepresidente, secretario y tesorero, y ningún miembro puede ocupar más de un puesto. ¿Cuántas planillas diferentes de candidatos son posibles?

$$nPr = \frac{n!}{(n-r)!}$$

$$20P4 = \frac{20!}{(20-4)!} = 116\,280$$

3.7.2 Combinaciones

Considere que en la permutación el orden de los arreglos es importante, en la combinación el orden no importa; por ejemplo, al ordenar los objetos a, b y c, sin importar el orden; una permutación es abc, otra bac, cba; de hecho, existen 6 opciones posibles para las permutaciones.

$$nPr = \frac{n!}{(n-r)!}$$

$$3P3 = \frac{3!}{(3-3)!} = 6$$

Si no importa el orden de los arreglos: abc, bac, cba; es el mismo arreglo, que se le llama combinación.

COMBINACION: Es el número de modos para elegir r objetos de un grupo de n, sin considerar el orden y sin repetición.

$$\text{FORMULA DE LA COMBINACION: } nCr = \frac{n!}{r!(n-r)!}$$

Ejemplo 1: Si un club tiene 20 miembros, ¿Cuántos comités diferentes de cuatro miembros son posibles?

$$nCr = \frac{n!}{r!(n-r)!}$$

$$20C4 = \frac{20!}{4!(20-4)!} = 4\,845$$

Ejemplo 2: Un entrevistador selecciono al azar 4 de 10 personas disponibles. ¿Cuántos grupos diferentes de 4 son posibles?

$$nCr = \frac{n!}{r!(n-r)!}$$

$$10C4 = \frac{10!}{4!(10-4)!} = 210$$

Ejemplo 3: Una organización de encuestas ha elaborado 15 preguntas destinadas a evaluar la actuación del presidente de Estados Unidos. El entrevistador seleccionara 10 de tales interrogantes. ¿Cuántos arreglos diferentes existen para el orden de las 10 preguntas seleccionadas?

$$nPr = \frac{n!}{(n-r)!}$$

$$15P10 = \frac{15!}{(15-10)!} = 10.897'286.400$$

Ejemplo 4: Un comité para la graduación en una escuela consta de 5 miembros. ¿En cuántas formas puede el comité lograr una decisión por mayoría a favor de un tipo de graduación?

$${}^5C_3 + {}^5C_4 + {}^5C_5 = \frac{5!}{3!(5-3)!} + \frac{5!}{4!(5-4)!} + \frac{5!}{5!(5-5)!} = 16$$

Ejemplo 5: Un artista ha creado 20 pinturas, y debe exhibir algunas de ellas en tres galerías. Se enviarán cuatro pinturas a la galería A, cuatro a la galería B y tres a la galería C. ¿De cuantas formas puede hacerse esto?

$${}^{20}C_4 * {}^{16}C_4 * {}^{12}C_3 = \frac{20!}{4!(20-4)!} * \frac{16!}{4!(16-4)!} * \frac{12!}{3!(12-3)!} = \frac{20!}{4! * 4! * 3! * 9!}$$

4. UNIDAD IV

4. Distribución de probabilidad discreta y continua



Distribuciones probabilistas

Una distribución de probabilidades da toda la gama de resultados posibles que pueden ocurrir con base en un experimento, resulta ser similar a una distribución de frecuencias estudiada en Estadística Descriptiva; sin embargo, en lugar de describir el pasado define que tan probable suceda algún evento futuro.

¿Qué es una distribución probabilística?

Una distribución de probabilidades muestra los posibles resultados de un experimento y la probabilidad de cada resultado.

4.1 Distribución probabilística

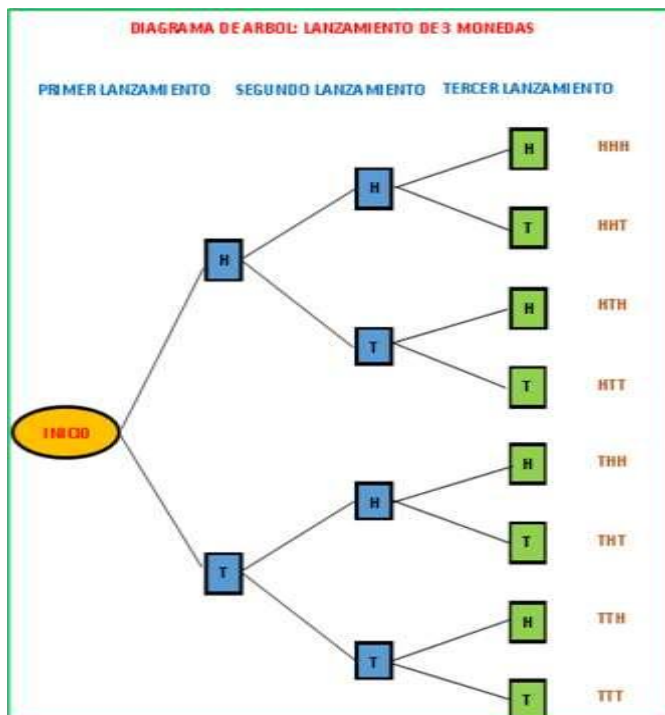
Enumeración de todos los resultados de un experimento junto con la probabilidad asociada a cada uno.

Ejemplo: Supóngase que se está interesado en el número de caras (H) que caen al lanzar una moneda tres veces. Este es el experimento. Los posibles resultados son: cero, uno, dos y tres caras. ¿Cuál es la distribución probabilística para el número de caras?

Numero de resultados posibles = $2 \times 2 \times 2 = 8$ Principio de conteo

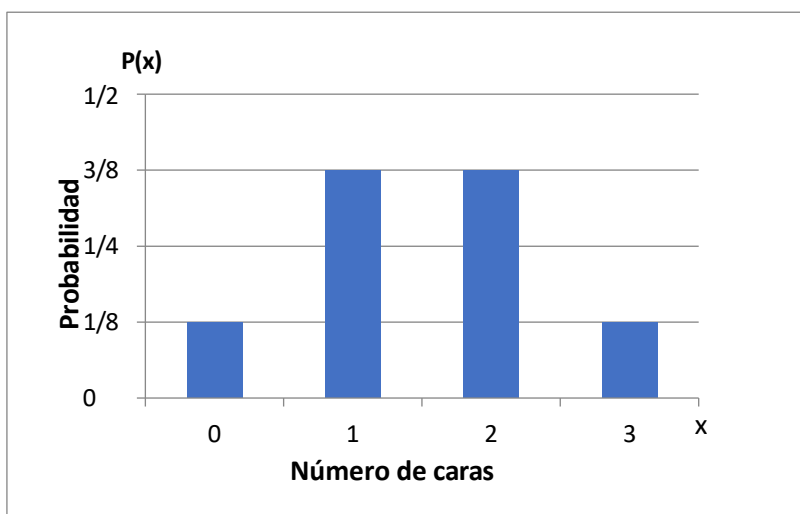
Posibles resultados: Cara (H); Cruz (T)

Posible resultado	Lanzamiento de la moneda			Número de Cara
	Primera	Segunda	Tercera	
1	T	T	T	0
2	T	T	H	1
3	T	H	T	1
4	T	H	H	2
5	H	T	T	1
6	H	T	H	2
7	H	H	T	2
8	H	H	H	3



Distribución de probabilidades para los eventos de cero, una, dos, tres caras; al lanzar 3 veces una moneda.

Número de caras x	Probabilidad del resultado $P(x)$	$P(x)$
0	$1/8$	$=0,125$
1	$3/8$	$=0,375$
2	$3/8$	$=0,375$
3	$1/8$	$=0,125$
Total	$8/8$	$=1,000$



Ejemplo: Para el experimento de lanzar un dado determine:

- Elabore una distribución de probabilidad para estos resultados.
- Represente en forma gráfica la distribución probabilística.
- ¿Cuál es el total de las probabilidades?

4.2 Variables aleatorias

En cualquier experimento aleatorio, los resultados ocurren al azar. Por ejemplo, si lanzamos un dado, puede ocurrir cualquiera de seis resultados posibles. Algunos experimentos dan resultados cuantitativos (como dólares, el peso, la estatura o el número de habitaciones), y otros cualitativos (como color racial, preferencia política o religiosa). En los ejemplos siguientes analicemos cual es la variable aleatoria:

- Si se considera el número de asistentes a la reunión de fin de semestre, el resultado podría ser 0, 1, 2, 3.....El número de asistentes es la variable aleatoria.
- Si se lanzan 3 monedas al aire y se considera el número de caras, el mismo podrá ser 0, 1, 2, y 3. El número de caras es la variable aleatoria.
- Si se pesa una persona, el resultado podría ser en libras 131,2, 131,7 o 131,1 dependiendo de la exactitud de la báscula. La variable aleatoria es el peso.

Variable aleatoria: Cantidad que es el resultado de un experimento aleatorio el cual, debido al azar, puede tomar valores diferentes. Y puede ser discreta y continua.

4.2.1 Variable aleatoria discreta

Variable que solo puede tener ciertos valores claramente separados, que resulta de contar algún elemento de interés. Por ejemplo:

- Si hay 35 estudiantes en el curso de Estadística Descriptiva, el número de estudiantes que no asisten puede ser 0, 1, 2, 3... 35. En general, una variable aleatoria discreta es el resultado de contar algo.
- Una variable discreta puede, en algunos casos tener valores fraccionarios o decimales. Dichos valores pueden estar separados, es decir, tener cierta distancia entre ellos, como, por ejemplo, las calificaciones de una prueba de admisión a una universidad como: 8.3, 8.7, 9.2. Estos valores son discretos porque existe una distancia entre ellos; por ejemplo 7,5 y 7,6. Una puntuación de 5,35 no puede ser.

4.2.2 Variable aleatoria continua

Si se mide algo, como el ancho de una habitación, la altura de una persona, se dice que es una variable aleatoria continua. Puede tomar uno de una cantidad infinitamente grande de valores, dentro de ciertas limitaciones, por ejemplo:

- La distancia aérea (en kilómetros) entre Quito y Ambato puede ser 122.50, 112.53 o 112.438 y así sucesivamente, dependiendo ello de la exactitud del dispositivo de medición.
- El peso de una persona podría ser en Kg: 45, 45.1, 45.13, 45.135...dependiendo de la exactitud de la báscula utilizada.

Si se organiza un conjunto de variables aleatorias discretas en una distribución de probabilidades, la distribución se denomina distribución de probabilística discreta. Los medios utilizados, así como las interpretaciones de probabilidad, son diferentes para las variables aleatorias discretas o las continuas.

4.2.3 Media, varianza y desviación estándar de una distribución de probabilidad.

En Estadística Descriptiva se estudió las medidas de localización y variación para una distribución de frecuencias, así como como la varianza y la dispersión de los datos alrededor de la media. De manera similar lo haremos para una distribución de probabilidades.

Media

Es el valor promedio a largo plazo de una variable aleatoria, la media de una distribución probabilística se denomina también valor esperado, $E(x)$ y viene a representar un promedio ponderado para el que los valores posibles que se consideran afectados por las probabilidades correspondientes de ocurrencia.

$$MEDIA DE UNA DISTRIBUCION PROBABILISTICA \quad \mu = E(x) = \sum[(x \cdot P(x))]$$

Varianza y desviación estándar

La media es un valor característico para representar una distribución probabilística discreta, pero no mide el grado de dispersión (variación) en una distribución. Para determinar el grado de dispersión utilizamos la varianza y la desviación estándar.

$$\text{VARIANCIA DE UNA DIST. PROBABILISTICA} \quad \sigma^2 = \sum[(X - \mu)^2 \cdot p(x)]$$

$$\text{DESVIACION ESTANDAR DE UNA DIST. PROBABILISTICA} \quad \sigma = \sqrt{\sigma^2}$$

EJEMPLO: Luis Macas vende automóviles nuevos de la Agencia Auto Regalado. Generalmente, negocia el mayor número de vehículos, los días sábado. Ha establecido la siguiente distribución probabilística para el número de autos que espera vender en un sábado en particular.

Numero de automoviles vendidos		Probabilidades
x		P(x)
0		0,10
1		0,20
2		0,30
3		0,30
4		0,10
	Total	1,00

1. ¿Qué tipo de distribución es ésta?
2. En un sábado común, ¿Cuántos autos debe esperar vender Luis?
3. ¿Cuál es la varianza de la distribución?

Solución:

1. ¿Qué tipo de distribución es ésta?

Una distribución probabilística discreta.

2. En un sábado común, ¿Cuántos autos debe esperar vender Juan?

Número de automóviles vendidos	Probabilidades	
x	P(x)	x.P(x)
0	0,1	0
1	0,2	0,2
2	0,3	0,6
3	0,3	0,9
4	0,1	0,4
	1	2,1

$$\mu = E(x) = \sum x * P(x) = 2.1$$

3. ¿Cuál es la varianza de la distribución?

$$\mu = E(x) = \sum x * P(x) = 2.1$$

Número de automóviles vendidos	Probabilidades	(x-μ)	(x-μ) ²	(x-μ) ² .P(x)
x	P(x)			
0	0,1	-2,1	4,41	0,441
1	0,2	-1,1	1,21	0,242
2	0,3	-0,1	0,01	0,003
3	0,3	0,9	0,81	0,243
4	0,1	1,9	3,61	0,361
			σ ²	1,29

$$\sigma^2 = \sum [(X - \mu)^2 \cdot p(x)] = 1.29$$

$$\sigma = \sqrt{1,29} = 1.136$$

Ejemplo 2: La siguiente tabla muestra la distribución probabilística para premios en efectivo de una rifa llevada a cabo en la tienda LQDSA.

Premio (\$)	Probabilidad
0	0,45
10	0,30
100	0,20
500	0,05

Si compra usted un solo boleto, ¿Cuál es la probabilidad de que gane:

- a) ¿Exactamente \$100? b) Por lo menos \$10? c) No más de \$100?
 d) Calcule la media, la varianza y la desviación estándar para esta distribución.

Solución:

- a) ¿Exactamente \$100?

$$P(X = 100) = 0.20$$

- b) Por lo menos \$10?

$$P(X \geq 10) = 0.30 + 0.20 + 0.05 = 0.55$$

c) ¿No más de \$100?

$$P(X \leq 100) = 0.20 + 0.30 + 0.45 = 0.95$$

d) Calcule la media, la varianza y la desviación estándar para esta distribución.

Premio (\$) x	Probabilidad P(x)	x.P(x)	(x-μ)	(x-μ) ²	(x-μ) ² .P(x)
0	0,45	0	-48	2304	1036,8
10	0,3	3	-38	1444	433,2
100	0,2	20	52	2704	540,8
500	0,05	25	452	204304	10215,2
	E(x) =	48		σ ²	12226

$$\mu = \sum x \cdot P(x)$$

$$\sigma^2 = \sum (x - \mu)^2 \cdot P(x)$$

$$\sigma = \sqrt{12226} = 110.57$$

Ejemplo 3: Se selecciona al azar a un grupo de 3 personas de una muestra de 10 que incluye a 4 menores de edad. Interesa conocer el número de estos últimos en el grupo.

a) ¿Cuáles son los valores posibles de esta variable aleatoria?

Variable aleatoria = número de menores de edad en el grupo.

Valores = 0, 1, 2, 3

b) Determine la distribución probabilística de esta variable.

Número de menores de edad	Probabilidad
x	P(x)
0	0,1667
1	0,5000
2	0,3000
3	0,0333
	1,0000

$$P(E) = \frac{N(E)}{N(T)}$$

$$N(T) = 10C3 = 120$$

$$N(E = 0) = 6C3 = 20$$

$$N(E = 1) = (6C2)(4C1) = 60$$

$$N(E = 2) = (6C1)(4C2) = 36 \quad N(E = 3) = 4C3 = 4$$

- c) ¿Cuál es la probabilidad de que haya: exactamente un menor de edad?, ¿por lo menos 1?, ¿dos o más?

$$P(X = 1) = 0.5000 \quad P(X \geq 1) = 0.5000 + 0.3000 + 0.0333 = 0.8333$$

$$P(X \geq 2) = 0.3000 + 0.0333 = 0.3333$$

- d) Calcule la media, la variancia y la desviación estándar de este número.

Número de menores de edad	Probabilidad	$x.P(x)$			$(x-\mu)^2.P(x)$
x	$P(x)$				
0	0,1667	0,0000			0,2400
1	0,5000	0,5000			0,0200
2	0,3000	0,6000			0,1920
3	0,0333	0,0999			0,1079
	1,0000	1,1999			0,5599

$$\sigma = \sqrt{0.5599} = 0.7483$$

FORMULA ALTERNATIVA PARA CALCULAR LA VARIANCIA.

$$\sigma^2 = \sum x^2.P(x) - \mu^2$$

Número de menores de edad	Probabilidad	x^2	$x^2.P(x)$
x	$P(x)$		
0	0,1667	0	0
1	0,5	1	0,5
2	0,3	4	1,2
3	0,0333	9	0,2997
	1		1,9997

$$\sigma^2 = 1.9997 - (1.1999)^2 = 0.5599$$

4.3 Distribución probabilística binomial

Es un ejemplo de una distribución probabilística discreta. La característica de una distribución binomial es que solo hay dos resultados posibles en una realización específica de un experimento. Por ejemplo, la respuesta a una pregunta de verdadero/falso es precisamente verdadero o falso. Los resultados son mutuamente excluyentes, lo cual significa que la respuesta a una pregunta de verdadero/falso no puede ser verdadero o falso al mismo tiempo. Frecuentemente clasificamos los dos

resultados posibles como “éxito” y “fracaso”, sin embargo, esta clasificación no implica que un resultado sea bueno y el otro malo.

Otra característica de la distribución binomial es que la variable aleatoria es el resultado del conteo. Esto es se cuenta el número de éxitos en la totalidad de ensayos, por ejemplo, se lanza una moneda 5 veces y se cuenta el número de caras que resultan; se seleccionan 6 trabajadores y se evalúa el número de ellos que tienen menos de 40 años.

Otra particularidad de esta distribución es que la probabilidad de un éxito permanece igual de un ensayo a otro. Ejemplo: La probabilidad de que se adivine la primera pregunta de verdadero/falso en forma correcta (éxito) es igual a $\frac{1}{2}$, al igual que en la segunda pregunta sigue siendo $\frac{1}{2}$; así como en la tercera y sucesivamente en las otras preguntas.

Otra característica en una distribución probabilística binomial es que un ensayo es independiente de cualquier otro, esto significa que no existe una configuración rítmica con respecto a los resultados; es decir los resultados podría ser VVFFVFFVFF.....

Características de una distribución binomial:

1. El resultado de cada ensayo o realización de un experimento se clasifica en una de dos categorías mutuamente excluyentes: éxito o fracaso.
2. La variable aleatoria es el resultado de contar el número de éxitos en una cantidad fija de ensayos.
3. La probabilidad de un éxito permanece igual para cada ensayo. Lo mismo sucede con la probabilidad de un fracaso.
4. Los ensayos son independientes, lo cual significa que el resultado de un ensayo no afecta el resultado de algún otro.

¿Cómo se elabora una distribución probabilística binomial?

Se debe saber:

1. El número de ensayos.
2. La probabilidad de éxito en cada ensayo.

Ejemplo: Si un examen de 10 preguntas de opción múltiple y cada pregunta tiene 4 opciones de respuesta; el número de ensayos es 10 y la probabilidad de “éxito” es 0.25. Por lo tanto, la probabilidad de que una persona sin conocimiento del tema adivine la respuesta a una pregunta en forma correcta, tiene un valor de 0.25.

FORMULA DE DISTRIBUCION PROBABILISTICA BINOMIAL:

$$P(x) = nCx \cdot \pi^x (1 - \pi)^{n-x}$$

n número de ensayos.

x número de éxitos

π Probabilidad de éxito en cada ensayo; $\pi \neq 3.1416$

Ejemplo: Un examen consiste en 4 preguntas de verdadero/falso y un estudiante no sabe nada acerca de la materia. ¿Cuál es la probabilidad de:

1. ¿No obtener exactamente ninguna de las cuatro en forma correcta?
2. Obtener exactamente una de las cuatro.

Solución:

Hacer análisis si cumple las condiciones de una distribución probabilística binomial:

- a) Se clasifica el experimento con éxito o fracaso.
- b) La variable aleatoria es el resultado del conteo
- c) La probabilidad de éxito permanece igual para cada ensayo 0.50
- d) Los ensayos son independientes; igual probabilidad de ocurrencia en cada lanzamiento, no dependiendo del anterior.

1. ¿No obtener exactamente ninguna de las cuatro en forma correcta?

$$n = 4 \quad \pi = 0.50 \quad x = 0$$

$$P(x) = nC_x \cdot \pi^x (1 - \pi)^{n-x}$$

$$P(X = 0) = 4C_0 * (0.50)^0 (1 - 0.50)^{4-0} = 0.0625$$

2. Obtener exactamente una de las cuatro.

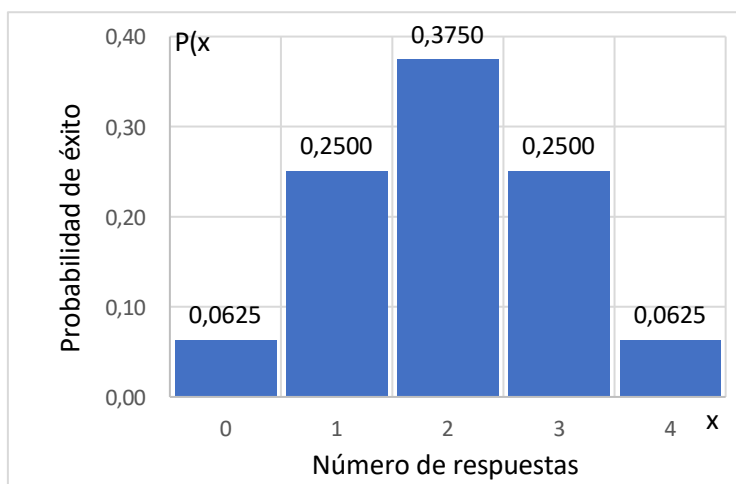
$$n = 4 \quad \pi = 0.50 \quad x = 1$$

$$P(X = 1) = 4C_1 * (0.50)^1 (1 - 0.50)^{4-1} = 0.2500$$

Planteemos por medio de una tabla todas las probabilidades de éxito, esto es 0, 1, 2, 3, 4 y su respectivo diagrama de distribución de probabilidad para $n = 4$ $\pi = 0.50$

Numero de	Probabilidad
-----------	--------------

éxitos	Fracción	Decimal
0	1/16	0,0625
1	4/16	0,2500
2	6/16	0,3750
3	4/16	0,2500
4	1/16	0,0625
Total	16/16	1,0000



Uso de tablas de probabilidad binomial

Para encontrar los valores descriptos de éxito podemos utilizar la tabla del apéndice A, del texto ADMINISTRACION PARA ADMINISTRACION Y ECONOMIA. Mason/Lind/Marchal.

$$n = 4 \quad \pi = 0.50$$

x	P(x)
0	0,063
1	0,250
2	0,375
3	0,025
4	0,063

Media y variancia de una distribución binomial

Media de una distribución binomial:

$$\mu = n \cdot \pi$$

Variancia de una distribución binomial:

$$\sigma^2 = n\pi(1 - \pi)$$

Ejemplo: Con base a la experiencia, 5% de los engranajes producidos por una maquina automática de alta velocidad Carter-Bell, resultan defectuosos. ¿Cuál es la probabilidad de que, si entran seis engranajes seleccionados al azar, exactamente cero sean defectuosos? ¿Exactamente uno? ¿Dos? ¿Tres? ¿Cuatro? ¿Cinco? ¿O exactamente 6?

Solución: Condiciones binomiales: 1) Existe la probabilidad constante de éxito (0.05). 2) Hay un número fijo de ensayos (6). 3) Los ensayos son independientes. 4) Existen solo dos resultados posibles (aceptable o defectuoso).

Tabla apéndice A: $n = 6$ $\pi = 0.05$

Número de engranajes defectuosos	Probabilidad de ocurrencia
x	$P(x)$
0	0,735
1	0,232
2	0,031
3	0,002
4	0,000
5	0,000
6	0,000

- Nota para el numero de engranajes defectuoso 4, 5 y 6; la tabla muestra como probabilidad cero, pero si utilizamos la formula podemos apreciar que es un valor muy pequeño, pero existe esa probabilidad.

$$P(x) = nCx \cdot \pi^x (1 - \pi)^{n-x}$$

$$P(5) = \frac{6!}{5!(6-5)!} (0.05)^5 (1 - \pi)^{6-5} = 0.00000178$$

$$P(5) = 6C5 * (0.05)^5 (1 - 0.05)^{6-5} = 0.00000178$$

- Para el ejemplo calculemos la media y la variancia.

$$\mu = n \cdot \pi \quad \mu = 6 * 0.05 = 0.30$$

$$\sigma^2 = n\pi(1 - \pi) \quad \sigma^2 = 6 * 0.05(1 - 0.05) = 0.285$$

- Estos valores los podemos calcular siguiendo el procedimiento antes descrito.

Número de menores de edad	Probabilidad	$x \cdot P(x)$	$(x - \mu)$	$(x - \mu)^2$	$(x - \mu)^2 \cdot P(x)$
x	$P(x)$				
0	0,735	0			0,06615
1	0,232	0,232			0,11368
2	0,031	0,062			0,08959
3	0,002	0,006			0,01458
4	0	0			0
5	0	0			0
6	0	0			0
	μ	0,3		σ^2	0,284

Ejemplo: Una encuesta de corretaje reporta que 30% de los inversionistas individuales, ha utilizado a un corredor de descuento; esto es, uno que no cobra las comisiones completas. En una muestra seleccionada al azar de nueve inversionistas, ¿Cuál es la probabilidad de que:

- ¿Exactamente dos de los individuos de la muestra de la muestra hayan empleado a un corredor de descuento?
- ¿Exactamente cuatro de ellos hayan utilizado a un corredor de ese tipo?
- ¿Ninguno haya recurrido a un corredor de descuento?

Solución: Condiciones binomiales: 1) Existe la probabilidad constante de éxito (0.30). 2) Hay un número fijo de ensayos (9). 3) Los ensayos son independientes. 4) Existen solo dos resultados posibles (haya utilizado o no lo haya utilizado).

Tabla apéndice A: $n = 9$ $\pi = 0.30$

Número de empleados de descuento	Probabilidad de ocurrencia
x	$P(x)$
0	0,040
1	0,156
2	0,267
3	0,267
4	0,172
5	0,074
6	0,021
7	0,004
8	
9	0,000

- a) ¿Exactamente dos de los individuos de la muestra de la muestra hayan empleado a un corredor de descuento?

$$n=9 \quad \pi = 0.30 \quad x = 2$$

$$P(x) = nCx \cdot \pi^x (1 - \pi)^{n-x}$$

$$P(x = 2) = 0.267$$

$$P(X = 2) = 9C2 \cdot (0.30)^2 (1 - 0.30)^{9-2} = 0.267$$

- b) Exactamente cuatro de ellos hayan utilizado a un corredor de ese tipo?

$$P(x = 4) = 0.172$$

$$P(4) = 9C4 * (0.30)^4 (1 - 0.30)^{9-4} = 0.172$$

- c) Ninguno haya recurrido a un corredor de descuento?

$$P(x = 0) = 0.004$$

$$P(0) = \frac{9!}{4! (9 - 0)!} (0.30)^0 (1 - 0.30)^{9-0} = 0.004$$

Ejemplo: Un 10% de los empleados de producción de la empresa Nome Steel están ausentes del trabajo en un determinado día de verano. Supóngase que se seleccionan al azar 10 trabajadores de producción para un estudio riguroso de ausentismo.

- a) ¿Cuál es la variable aleatoria?

El ausentismo

- b) ¿Tal variable es discreta o continua?
- c) ¿Cuál es la probabilidad de que ninguno de los 10 empleados seleccionados esté ausente?

$$\pi = 0.10 \quad n = 10$$

$$P(x) = nCx \cdot \pi^x (1 - \pi)^{n-x}$$

$$P(X = 0) = 10C0 \cdot (0.10)^0 (1 - 0.10)^{10} =$$

- d) Desarrolle una distribución probabilística binomial para el experimento.
- e) Calcule la media, la variancia y la desviación estándar de la distribución.

$$\mu = n\pi = (10)(0.10) = 1$$

$$\sigma^2 = 10 * 0.10(1 - 0.10) = 0.90$$

Ejemplo: La rapidez con la que las compañías de servicios resuelven problemas es de suma importancia. Georgetown Telephone Company afirma que es capaz de resolver 70% de los problemas de los clientes el mismo día en que se reportan. Suponga que los 15 casos que se reportaron el día de hoy son representativos de todas las quejas.

- Cuántos problemas esperaría que se resolvieran el día de hoy? .Cual es la desviación estándar?
- Cuál es la probabilidad de que 10 problemas se resuelvan el día de hoy?
- De que 10 u 11 problemas se resuelvan el día de hoy?
- Y de que más de 10 problemas se resuelvan el día de hoy?

4.3.2 Distribuciones probabilísticas acumulativas.

Podría ser necesario conocer la probabilidad de adivinar correctamente las respuestas a 5 o más preguntas de verdadero/falso.

Ejemplo: Un estudio reciente por la Asociacion de Vigilantes de Carreteras revelo que el 60% de los conductores en Estados Unidos se coloca el cinturón de seguridad al manejar. Se seleccionó una muestra de 10 automovilistas en una carretera de Florida.

- ¿Cuál es la probabilidad de que exactamente 7 llevaran fijo el cinturón?
- ¿Cuál es la probabilidad de que 7 o menos lleven puesto el cinturón?

Solución: Condiciones binomiales:1) Existe la probabilidad constante de éxito (0.60). 2) Hay un número fijo de ensayos (10). 3) Los ensayos son independientes. 4) Existen solo dos resultados posibles (lleve puesto o no lleve puesto).

Tabla apéndice A: $n = 10$ $\pi = 0.60$

Número de personas	Probabilidad de
lleven cinturón	ocurrencia
x	$P(x)$
0	0,000
1	0,002
2	0,011
3	0,042
4	0,111
5	0,201
6	0,251
7	0,215
8	0,121
9	0,040
10	0,006

- ¿Cuál es la probabilidad de que exactamente 7 llevaran fijo el cinturón?

$$P(x = 7) = 0.215$$

b) ¿Cuál es la probabilidad de que 7 o menos lleven puesto el cinturón?

$$P(x \leq 7) = P(0) + P(1) + P(2) + P(3) + P(4) + P(5) + P(6) + P(7)$$

$$P(x \leq 7) = 0.000 + 0.002 + 0.011 + 0.042 + 0.111 + 0.201 + 0.251 + 0.215$$

$$P(x \leq 7) = 0.833$$

$$P(x \leq 7) = 1 - [P(8) + P(9) + P(10)]$$

$$P(x \leq 7) = 1 - [0.121 + 0.040 + 0.006] = 0.833$$

Ejemplo: En un estudio reciente se halló que el 90% de las casas de EEUU tienen televisión a colores. En una muestra de 9 viviendas, ¿Cuál es la probabilidad de que:

- a) Las nueve tengan televisión?
- b) Menos de 5 posean dichos aparatos?
- c) Más de 5 tengan televisión a color?
- d) Por lo menos siete de las casas cuenten con ella?

4.3.3 Distribución probabilística hipergeométrica.

Para aplicar la distribución binomial, la probabilidad de éxito debe permanecer igual para cada ensayo sucesivo. La probabilidad de adivinar la respuesta correcta en un ensayo de verdadero/falso es 0.50 que permanece igual para cada pregunta de un examen.

Supóngase que el 60% de los votantes registrados en Quito son de AP. Si se van a seleccionar al azar 30 votantes registrados, la probabilidad de elegir uno de AP es 0.60; la de obtener uno de ese partido en la siguiente selección, también es 0.60, con la consideración que el muestreo se realiza con reposición, lo cual significa que la persona seleccionada se pone nuevamente en la población antes de elegir la siguiente.

La mayor parte del muestreo se realiza **sin reposición**. Es decir, si la población es pequeña, la probabilidad para cada observación cambiará. Por ejemplo si la población está conformada por 15 elementos, la probabilidad de seleccionar uno en particular es $1/15$. Si el muestreo se realiza sin reposición, la probabilidad de escoger un elemento ahora es de $1/14$ y así sucesivamente. Esto considerando que la población es finita, es decir, se conoce el número de elementos en la población y es relativamente pequeño.

Población finita: Una población formada por un número pequeño de individuos, objetos o medidas.

Ejemplos de población finita, tenemos: 285 estudiantes de la Carrera de Contabilidad y Auditoría de la UPS en tercer semestre, los 2570 hinchas registrados del Nacional de Quito-Ecuador.

Cuando el muestreo se realiza sin reposición y la muestra se obtiene de una población relativamente pequeña, la probabilidad de éxito cambia de un ensayo a otro y no debe aplicarse la distribución binomial sino la hipergeométrica, la misma que para su aplicación se debe tomar en cuenta las siguientes características:

1. Si se selecciona una muestra de una población finita sin reposición.
2. Si el tamaño de la muestra n es mayor que el 5% del tamaño de la población N ($n > 0.05N$)
3. Si $n < 0.05N$ es suficiente la distribución binomial.

Con las dos características anotadas se utiliza la distribución hipergeométrica para determinar la probabilidad de un número específico de éxitos o fracasos.

FORMULA DE LA DISTRIBUCION HIPERGEOMETRICA:

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

- N Tamaño de la población
- S Cantidad de éxitos de la misma
- x número de éxitos que interesan
- n Tamaño de la muestra o el número de ensayos
- C Símbolo para una combinación.

Ejemplo: Supóngase que durante la semana se fabricaron 50 juegos de Play Station. Operaron 40 sin problemas, y 10 tuvieron al menos un defecto. Se selecciona una muestra al azar de 5. **Utilizando la formula hipergeométrica**, ¿Cuál es la probabilidad de que 4 de los 5 funcionan perfectamente?

¿Cuál es la probabilidad de que 4 de los 5 funcionan perfectamente?

Solución:

Análisis de condiciones: (1) Muestreo se realiza sin reposición. (2) Tamaño de la muestra es $5/50=10\%$ de la población que es mayor al 5%.

Datos:

$N=50$

$n=5$

$S=40$

$$x=4$$

$$P(x) = \frac{(sC_x)(N-sC_{n-x})}{(NC_n)}$$

$$P(4) = \frac{(40C4)(10C1)}{50C5} = 0.431$$

La probabilidad de seleccionar 5 juegos al azar y descubrir que 4 operan bien es de 0.431.

- Encontrar la probabilidad de que tres de los 5 operen de forma correcta.

$$P(5) = \frac{(40C3)(10C2)}{50C5} = 0.2098$$

- Encontrar la probabilidad de que tres de los 5 no operen de forma correcta.

$$N=50$$

$$n=5$$

$$S=10$$

$$x=4$$

$$P(4) = \frac{(10C4)(40C1)}{50C5} = 0.00396$$

Ejemplo: En KOPA se acaba de recibir un embarque de 10 aparatos de televisión. Poco después de haberse efectuado la entrega, el fabricante llamo para informar que por descuido se habían enviado 3 televisores defectuosos. La Sra. Pérez, propietaria de la empresa, decidió probar 2 de los 10 aparatos recibidos. ¿Cuál es la probabilidad de que ninguno de los dos tenga defectos?

Solución: Análisis de condiciones: (1) Muestreo se realiza sin reposición. (2) Tamaño de la muestra es $3/10=30\%$ de la población que es mayor al 5%.

$$N = 10$$

$$n = 2$$

$$S = 7$$

$$x = 0$$

$$P(x = 0) = \frac{(7C2)(3C0)}{10C2} = 0.4667$$

Ejemplo: Keith's Florists tiene 15 camiones de entrega, que emplea sobre todo para entregar flores y arreglos florales en la zona de Greenville, Carolina del Sur. De estos 15 camiones, 6 presentan problemas con los frenos. En forma aleatoria se seleccionó una muestra de 5 camiones. ¿Cuáles la probabilidad de que 2 de los camiones probados presenten frenos defectuosos?

4.3.4 Distribución probabilística de Poisson

Las distribuciones probabilísticas binomiales para probabilidades de éxito (π) menores que el 5% (0.05) podrían calcularse, pero tomaría demasiado tiempo (en especial para una n grande, por ejemplo, 100 o más).

La forma límite de la distribución binomial cuando la probabilidad de éxito es muy pequeña y n es grande se denomina **distribución probabilística de Poisson**. Generalmente se le conoce como la ley de eventos improbables. Lo cual significa que la probabilidad π de que suceda un evento específico es muy pequeña.

Tiene muchas aplicaciones como: Se utiliza como modelo para describir la distribución de errores en la captura de datos, el número de clientes en espera de servicio en un restaurante.

La distribución de Poisson puede utilizarse con la formula siguiente:

$$P(x) = \frac{\mu^x e^{-\mu}}{x!} \quad \mu = n \cdot \pi$$

μ Es la media (aritmética) del número de ocurrencias (éxito) en un intervalo de tiempo específico

e constante 2.71828

n es el número de ocurrencias (éxito)

$P(x)$ es la probabilidad que se va a calcular para un valor dado de x .

Ejemplo: Considérese que el Aerolínea TAME rara vez se pierde el equipaje. En la mayoría de los vuelos no se observa un mal manejo de las maletas; algunos reportan una valija perdida; unos cuantos tienen dos maletas extraviadas; rara vez un vuelo tiene tres; y así sucesivamente. Supóngase que en una muestra aleatoria de 1000 viajes aéreos revela que un total de 300 maletas perdidas.

La media aritmética del número de equipajes extraviados por vuelo es: $\mu = 300/1000 = 0.30$. Si la cantidad de maletas perdidas sigue una distribución de Poisson con $\mu = 0.30$. Podemos calcular las diferentes probabilidades como:

- Probabilidad de no perder ninguna maleta.

$$P(x) = \frac{\mu^x e^{-\mu}}{x!} \quad P(0) = \frac{0.30^0 e^{-0.30}}{0!} = 0.7404$$

- Probabilidad de que exactamente dos maletas extraviadas es:

$$P(2) = \frac{0.30^2 e^{-0.30}}{2!} = 0.0333$$

Todas estas probabilidades las podemos encontrar utilizando el apéndice C para $\mu = 0.30$

x	P(x)
0,0000	0,7408
1,0000	0,2222
2,0000	0,0333
3,0000	0,0033
4,0000	0,0003
5,0000	0,0000
6,0000	0,0000
7,0000	0,0000

Ejemplo: En una distribución de Poisson $\mu = 0.40$

- a) ¿Cuál es la probabilidad de que $x = 0$?

$$P(x) = \frac{\mu^x e^{-\mu}}{x!} \quad P(0) = \frac{0.40^0 e^{-0.40}}{0!} = 0.6703$$

- b) ¿Y cuál es la que $x > 0$?

$$P(x > 0) = 1 - P(0) = 1 - 0.6703 = 0.3297$$

EJEMPLO: La Sra. Zurita está encargada de los préstamos en el Banco del Austro. Con base en sus años de experiencia, estima que la probabilidad de que un solicitante no sea capaz de pagar un préstamo, es 0.025. El mes pasado realizó 40 préstamos.

- a) ¿Cuál es la probabilidad de que 3 préstamos no sean pagados oportunamente?

$$P(x) = \frac{\mu^x e^{-\mu}}{x!}$$

$$\mu = n \cdot \pi \quad \mu = 40 \times 0.025 = 1$$

$$P(x) = \frac{\mu^x e^{-\mu}}{x!} \quad P(X = 3) = \frac{1^3 e^{-1}}{3!} = 0.0613$$

- b) ¿y cuál es la probabilidad de que por lo menos 3 préstamos no se liquiden a tiempo?

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

APENDICE C: $\mu = 1$

x	P(x)
0	0,3679
1	0,3679
2	0,1839
3	0,0613
4	0,0153
5	0,0031
6	0,0005
7	0,0001
8	0,0000

$$P(x) = \frac{1^x e^{-1}}{x!}$$

$$P(\geq 3) = 1 - (0.3679 + 0.3679 + 0.1839) = 0.0803$$

Ejemplo ¿35?: Se calcula que 0.5% de quienes se comunican al departamento de servicio al cliente de Dell, Inc., escuchará un tono de línea ocupada. ¿Cuál es la probabilidad de que de las 1 200 personas que se comunicaron hoy (a) exactamente 5 no hayan escuchado la llamada? (b) Por lo menos 5 hayan escuchado un tono de línea ocupada?

Clasificación: La probabilidad de éxito es pequeña y n es grande, entonces utilizamos la distribución de Poisson.

$$P(x) = \frac{\mu^x e^{-\mu}}{x!}$$

$$\mu = n\pi = 1200 * 0.005 = 6$$

$$a) P(x = 5) = \frac{6^5 e^{-6}}{5!} = 0.1606$$

$$b) P(X \geq 5) = 1 - (P(0) + P(1) + P(2) + P(3) + P(4))$$

Ejemplo ¿49? : Un estudio federal informó que 7.5% de la fuerza laboral de Estados Unidos tiene problemas con las drogas. Una oficial antidroga del estado de Indiana decidió investigar esta afirmación. En una muestra de 15 trabajadores:

a) ¿Cuántos trabajadores cree que presenten problemas de adicción a las drogas? ¿Cuál es la desviación estándar?

b) ¿Cuál es la probabilidad de que ninguno de los trabajadores de la muestra manifieste problemas de adicción?

c) ¿Cuál es la probabilidad de que por lo menos uno de los trabajadores de la muestra presente problemas de adicción?

Ejemplo ¿59? El cargo de jefe de la policía en la ciudad de Corry, Pennsylvania, se encuentra vacante. Un comité de búsqueda, integrado por los residentes de Corry, tiene

la responsabilidad de recomendar al alcalde de la ciudad al nuevo jefe de la policía. Hay 12 candidatos, 4 de los cuales son mujeres o miembros de una minoría. El comité decide entrevistar a los 12 candidatos.

Primero seleccionaron al azar a cuatro candidatos para entrevistarlos el primer día, ninguno de los cuales resultó ser mujer ni miembro de una minoría. El periódico local, CorryPress, en una de sus columnas editoriales, sugiere que hay discriminación. ¿Cuál es la probabilidad de que así sea?

Ejemplo ¿65?: Los informes recientes relacionados con el crimen indican que cada minuto ocurren 3.1 robos a vehículos motorizados en Estados Unidos. Suponga que la distribución de los robos por minuto se puede aproximar por medio de una distribución de probabilidad de Poisson.

a) Calcule la probabilidad de que ocurran exactamente cuatro robos en un minuto.

b) ¿Cuál es la probabilidad de que no haya robos en un minuto?

c) ¿Cuál es la probabilidad de que por lo menos haya un robo en un minuto?

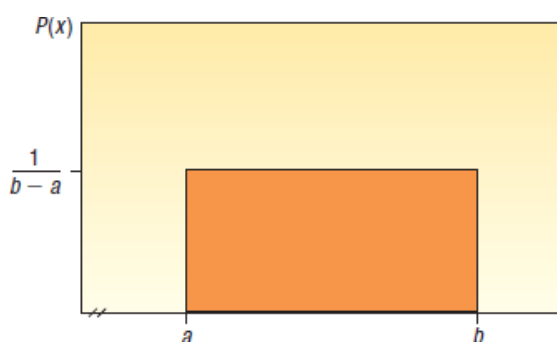
4.4 Distribuciones probabilísticas continuas

En el capítulo anterior estudiamos las distribuciones probabilísticas discretas, continuamos el estudio con las distribuciones probabilísticas continuas, donde una variable aleatoria continua es una variable que generalmente es el resultado de medir algo, como la estatura de una persona, el peso de un producto que está en venta etc.

4.4.1 Distribuciones de probabilidad uniforme

La distribución de probabilidad uniforme es la más simple de una variable aleatoria continua. La distribución tiene forma rectangular y queda definida por valores mínimos y máximos.

- Para ejemplificar tenemos el caso la resistencia de un bloque de cemento de construcción está entre 70 a 90 kg/cm². La variable aleatoria es la resistencia del bloque de cemento, que es continua en el intervalo de 70 a 90 kg/cm².



Grafica tomada de texto base

La media y la desviación estándar de una distribución uniforme está dada por:

$$\mu = \frac{a+b}{2} \quad \text{Media}$$

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

La probabilidad de distribución uniforme está dada.

$$P(x) = \frac{1}{b-a} \quad \text{si } a \leq x \leq b \text{ y } 0 \text{ en cualquier otro lugar} \quad \text{Probabilidad uniforme}$$

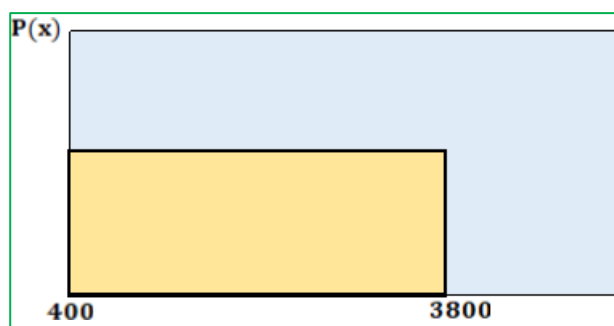
En el caso de distribuciones que describen una variable aleatoria continua, las áreas dentro de la distribución representan probabilidades.

$$\text{Área} = (\text{largo})(\text{ancho}) = \frac{1}{(b-a)} \cdot (b-a) = 1$$

Ejemplo: Estudios de familias establece que una familia de cuatro miembros gasta entre \$400 y \$3 800 anuales en toda clase de seguros. Suponga que el dinero que se gasta tiene una distribución uniforme entre estas cantidades.

- ¿Cuál es la media de la suma que se gasta en seguros?
- ¿Cuál es la desviación estándar de la suma que se gasta?
- Si elige una familia al azar, ¿cuál es la probabilidad de que gaste menos de \$2 000 anuales en seguros?
- ¿Cuál es la probabilidad de que una familia gaste más de \$3 000 anuales?

Solución



- ¿Cuál es la media de la suma que se gasta en seguros?

$$\mu = \frac{a+b}{2}$$

$$\mu = \frac{400 + 3800}{2} = \$2\,100$$

- ¿Cuál es la desviación estándar de la suma que se gasta?

$$\sigma = \sqrt{\frac{(b-a)^2}{12}}$$

$$\sigma = \sqrt{\frac{(3\,800 - 400)^2}{12}} = \$981.50$$

c) Si elige una familia al azar, ¿cuál es la probabilidad de que gaste menos de \$2 000 anuales en seguros?

$$P(X \leq 2000) = \frac{1}{(3\,800 - 400)} \cdot (3800 - 2000) = 0.5294$$

d) ¿Cuál es la probabilidad de que una familia gaste más de \$3 000 anuales?

$$P(X \geq 3000) = \frac{1}{(3\,800 - 400)} \cdot (3800 - 3000) = 0.2353$$

4.4.2 Distribuciones de probabilidad normal

Enseguida se estudia la distribución de probabilidad normal, la distribución de probabilidad normal tiene una formula algo compleja.

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

X: Valor de la variable aleatoria

σ : Desviación estándar

μ : Media

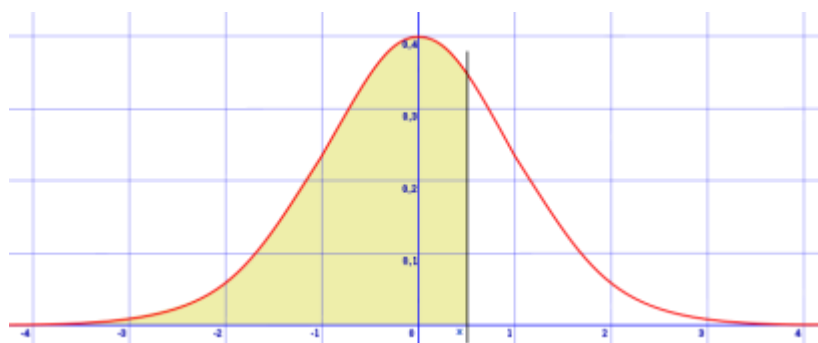
π : 3.1416

e: 2.7183

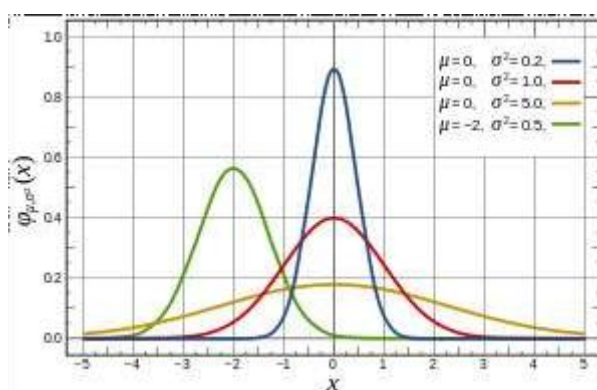
Para calcular la probabilidad no recurriremos a la formula, sino utilizaremos una tabla a la que nos referiremos posteriormente.

Una distribución probabilística continua muy importante es la distribución probabilística normal y su respectiva curva normal que tiene las características siguientes:

1. La curva normal es acampanada y presenta un solo pico en el centro de la distribución. La media (aritmética), la mediana y moda son iguales y están localizadas en el pico. De esta forma, la mitad del área bajo la curva normal se encuentra hacia la derecha y la otra hacia la izquierda.
2. La distribución probabilística normal es simétrica con respecto a su media.
3. La curva normal decrece en ambas direcciones a partir del valor central. Es asintótica, lo que significa que se acerca al eje X, pero sin tocarlo.



No existe una sola distribución probabilística normal, sino que hay una “familia de ellas, dependiendo del valor de μ y σ como se muestra en el grafico siguiente:



4.4.2.1 Distribución probabilística estándar.

Para la distribución normal resulta imposible proporcionar una tabla de probabilidades (como la binomial o de Poisson) para cada combinación de μ y σ . Pero puede utilizarse un elemento de la familia de distribuciones normales para todos los problemas donde la distribución resulte aplicable. Tiene una media igual a cero y una desviación estándar igual a 1, y se denomina distribución normal estándar. Cualquier distribución normal puede convertirse en una distribución normal estándar restando la media a cada observación y dividiendo luego entre la desviación estándar.

Primero se convierte, o se estandariza, la distribución a una distribución normal estándar utilizando el valor de z , (también denominado desvío normal estandarizado o simplemente desvío normal).

Valor z : Diferencia (desviación) entre un valor seleccionado, denotado por X , y la media μ , dividida tal diferencia entre la desviación estándar σ .

$$\text{VALOR NORMAL } Z \qquad Z = \frac{X - \mu}{\sigma}$$

X es el valor de cualquier medida u observación específica.

μ es la media de la distribución

σ es la desviación estándar de la distribución.

El valor z mide la distancia entre el valor específico X y la media aritmética, en unidades de la desviación estándar. Conocido Z y utilizando la tabla, se puede obtener el área o la probabilidad bajo la curva normal.

Ejemplo: La media de un grupo de ingresos semanales para un conjunto de gerentes de nivel medio, es de \$1200, y la desviación estándar es de \$150 ¿Cuál es la probabilidad de que los ingresos estén entre \$1200 y \$1500? ¿Cuál es la probabilidad que esté entre \$950 y \$1200? ¿La probabilidad que sean más de \$1500?

$$\mu = \$1\,200$$

$$\sigma = \$150$$

$$Z = \frac{X - \mu}{\sigma}$$

$$X = 1500 \quad Z = \frac{1500 - 1200}{150} = 2.00$$

$$\mu = 1.200$$

$$P(1500) = 0.4772$$

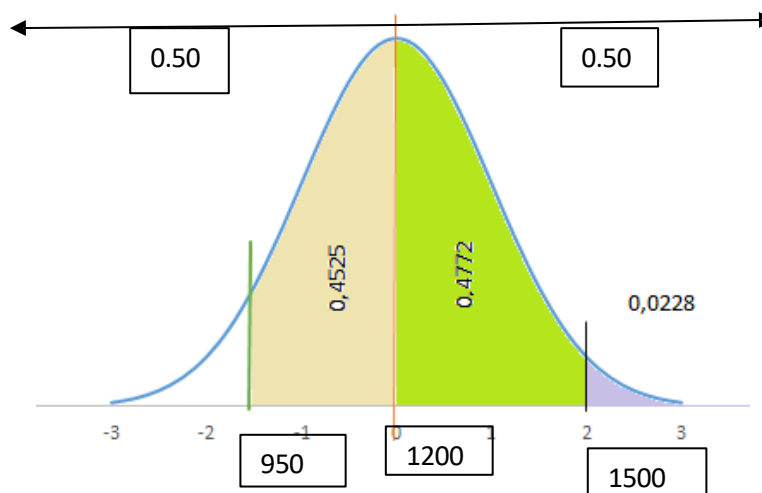
$$\sigma = 150$$

$$X = 950 \quad Z = \frac{950 - 1200}{150} = -1.67$$

$$\mu = 1.200$$

$$P(950) = 0.4525$$

$$\sigma = 150$$



- $P(X > 1500) = 0.50 - 0.4772 = 0.0228$

- Probabilidad que este \$950 y \$1500

$$P(950 \leq X \leq 1500) = 0.4525 + 0.4772 = 0.9297$$

- Probabilidad que este entre \$1300 y \$1400

$$Z = \frac{X - \mu}{\sigma}$$

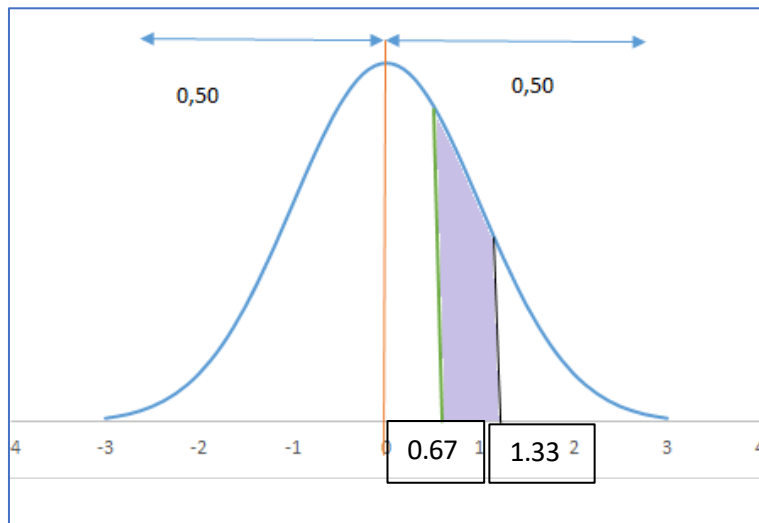
$$Z(1300) = 0.67$$

$$P(1300) = 0.2486$$

$$Z(1400) = 1.33$$

$$P(1400) = 0.4082$$

$$P(\text{entre } 1300 \text{ y } 1400) = 0.4082 - 0.2486 = 0.1594$$

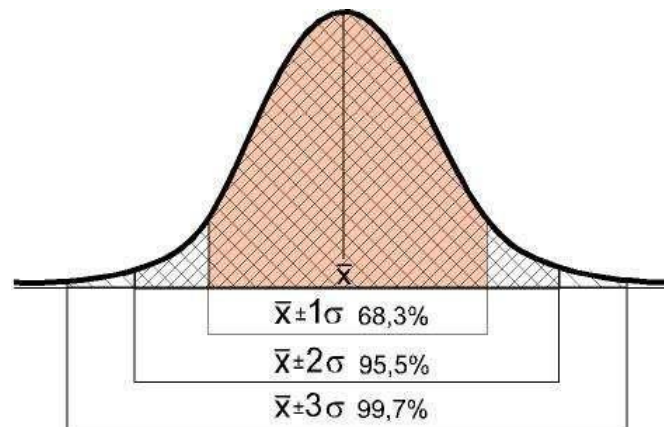


RESUMEN:

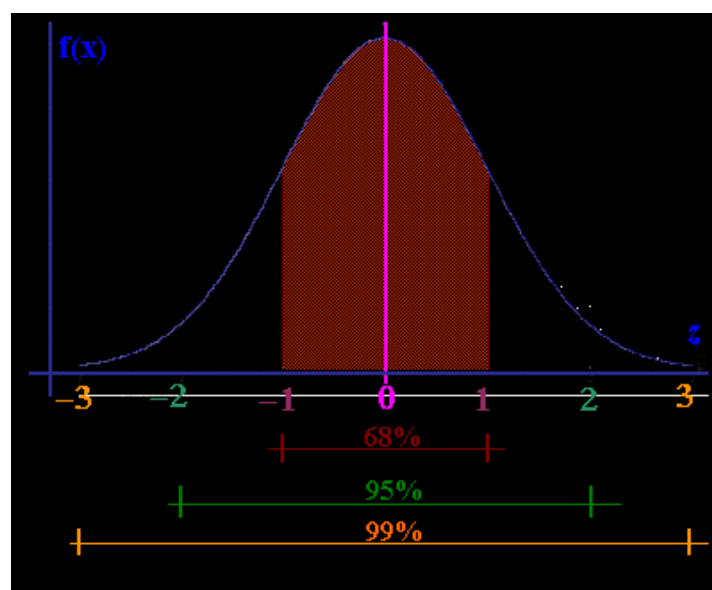
- Si se desea encontrar el área entre 0 y Z, ó entre 0 y $-Z$, se busca el valor directamente de la tabla.
- Si se desea encontrar una área más allá de Z o $-Z$, se localiza la probabilidad de Z y se resta del valor 0.5000
- Para el área entre 2 puntos en diferentes lados de la media, determine los valores Z y sume las áreas.
- Para el área entre dos puntos en el mismo lado de la media, determine los valores Z y reste las áreas.

4.4.3 Regla empírica

Para una distribución de frecuencias simétrica de campana; aproximadamente 68% de las observaciones estará a $\pm$ una desviación estándar desde la media; aproximadamente el 95% entre ± 2 desviaciones estándar y prácticamente todas las observaciones el 99.7% estarán a ± 3 desviaciones estándar con respecto a la media.



- Al transformar las mediciones a valores Z (o desvíos normales estándar) cambia la escala a los valores indicados en el gráfico.



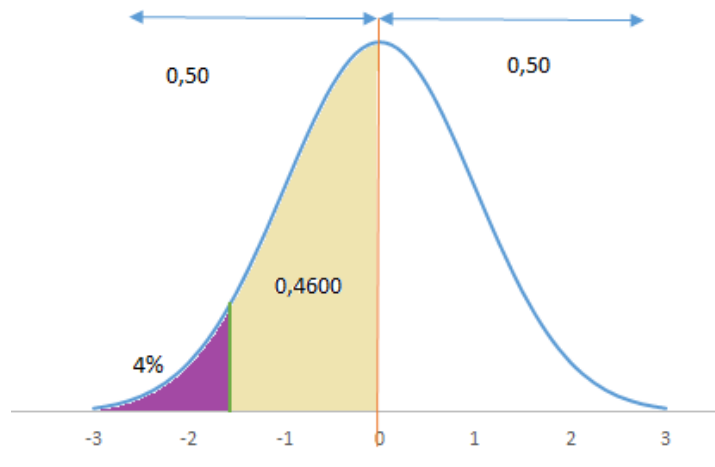
- Texto: 68% equivale a 68.26%; 95% a 95,44% ; 99% a 99.74%

EJEMPLO: Una prueba de vida útil para un gran número de pilas alcalinas tipo D, reveló que la duración media para un uso específico antes de la falla, es 19.0 horas (h). La distribución de las duraciones se aproxima a una distribución normal. La desviación estándar fue 1.2h

- ¿Entre que par de valores fallo aproximadamente 68% de las pilas?
- ¿Entre cuales dos valores ocurrieron la falla de alrededor de 95% de las pilas?
- ¿Entre que par de valores fallaron prácticamente todas las pilas?

EJEMPLO: Un fabricante de neumáticos desea fijar una garantía mínima de millas recorridas para su nueva llanta MX100. Las pruebas de duración revelaron que la media de las millas recorridas es 47900 y con una desviación estándar de 2050 millas, y con

una distribución normal. El fabricante desea fijar las millas recorridas de garantía de manera que no sea necesario más del 4% de los neumáticos. ¿Cuántas millas de recorrido de garantía debe anunciar el fabricante?



$$Z = \frac{X - \mu}{\sigma}$$

$$Z = \frac{X - 47900}{2050}$$

de Tabla valor aproximado a 0.4600 que corresponde a 0.4599

donde $Z = -1.75$

$$-1.754 = \frac{X - 47900}{2050}$$

$$X = 44.312 \text{ millas}$$

El fabricante deberá anunciar que reemplazará gratis cualquier neumático que se gaste antes de llegar a las 44.312 millas, y la compañía sabe que solo el 4% será sustituido bajo ese plan.

47. Un estudio de llamadas telefónicas de larga distancia que se realizó en las oficinas centrales de Pepsi Botting Group, Inc., en Somers, Nueva York, demostró que las llamadas, en minutos, se rigen por una distribución de probabilidad normal. El lapso medio de tiempo por llamada fue de 4.2 minutos, con una desviación estándar de 0.60 minutos.

- ¿Qué porcentaje de llamadas duró entre 4,2 y 5 minutos?
- ¿Qué porcentaje de llamadas duró más de 5 minutos?
- ¿Qué porcentaje de llamadas duró entre 5 y 6 minutos?
- ¿Qué porcentaje de llamadas duró entre 4 y 6 minutos?

e) Como parte de su informe al presidente, el director de comunicaciones desea informar la duración de 4% de las llamadas más largas. ¿Cuál es este tiempo?

SOLUCION.

DATOS:

$$\mu = 4.2 \text{ min}$$

$$\sigma = 0.60 \text{ min}$$

$$Z = \frac{X - \mu}{\sigma}$$

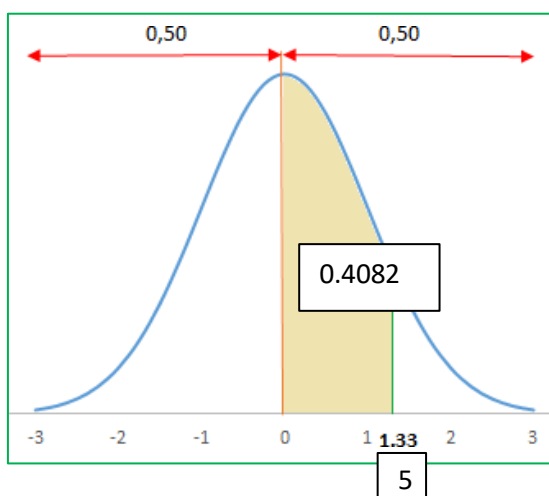
$$z_5 = 1.33 \qquad P(5) = 0.4082$$

$$z_6 = 3 \qquad P(6) = 0.4987$$

$$z_4 = -0.33 \qquad P(4) = 0.1293$$

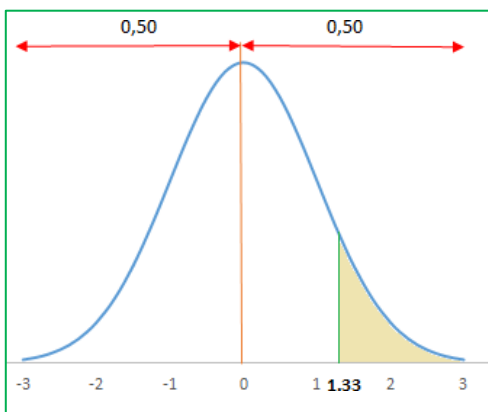
a) ¿Qué porcentaje de llamadas duró entre 4,2 y 5 minutos?

$$P(4.2 \leq X \leq 5) = 0.4082$$



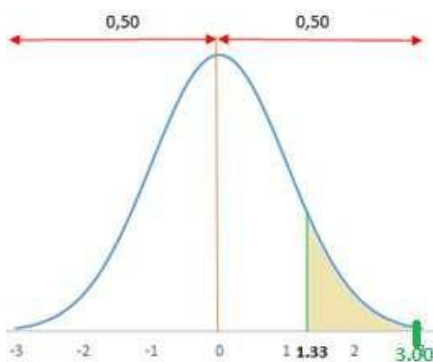
b) ¿Qué porcentaje de llamadas duró más de 5 minutos?

$$P(X \geq 5) = 0.5000 - 0.4082 = 0.0918$$



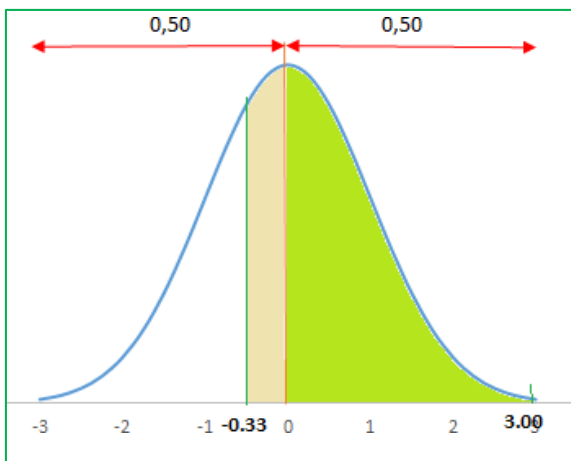
c) ¿Qué porcentaje de llamadas duró entre 5 y 6 minutos?

$$P(5 \leq X \leq 6) = 0.4987 - 0.4082 = 0.0895$$

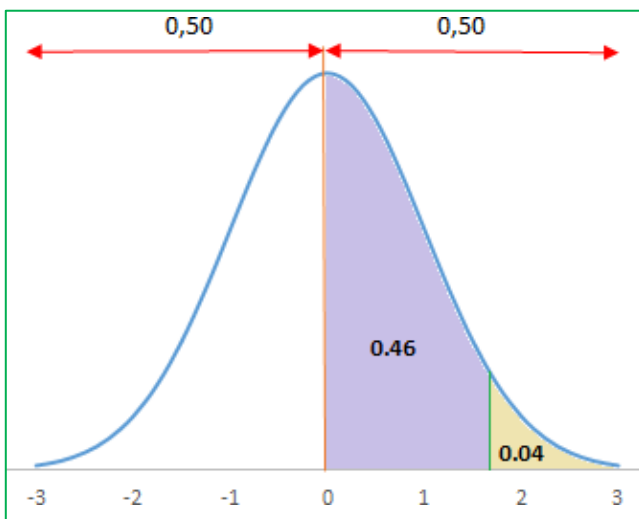


d) ¿Qué porcentaje de llamadas duró entre 4 y 6 minutos?

$$P(4 \leq X \leq 6) = 0.1293 + 0.4987 = 0.628$$



e) Como parte de su informe al presidente, el director de comunicaciones desea informar la duración de 4% de las llamadas más largas. ¿Cuál es este tiempo?



$$P(X) = 0.46$$

$$z = 1.75$$

$$z = \frac{X - \mu}{\sigma}$$

$$1.75 = \frac{X - 4.2}{0.6}$$

$$X = 5.25 \text{ min}$$

e) Como parte de su informe al presidente, el director de comunicaciones desea informar la duración de 4% de las llamadas menos largas. ¿Cuál es este tiempo?

$$P(X) = 0.46$$

$$z = -1.75$$

$$z = \frac{X - \mu}{\sigma}$$

$$-1.75 = \frac{X - 4.2}{0.6}$$

$$X = 3.15 \text{ min}$$

48. Un informe reciente publicado en USA Today indicaba que una familia común de cuatro miembros gasta \$490 al mes en alimentos. Suponga que la distribución de gastos de alimento de una familia de cuatro miembros sigue una distribución normal, con una media de \$490 y una desviación estándar de \$90.

a) ¿Qué porcentaje de familias gasta más de \$430 y menos de \$490 en alimentos al mes?

b) ¿Qué porcentaje de familias gasta menos de \$430 al mes en alimentos?

c) ¿Qué porcentaje de familias gasta entre \$430 y \$600 mensuales en alimentos?

d) ¿Qué porcentaje de familias gasta entre \$500 y \$600 mensuales en alimentos?

SOLUCION:

DATOS:

$$\mu = \$490$$

$$z = \frac{X - \mu}{\sigma}$$

$$\sigma = \$90$$

$$z_{30} = -5.11 \quad P(30) = 0.5000$$

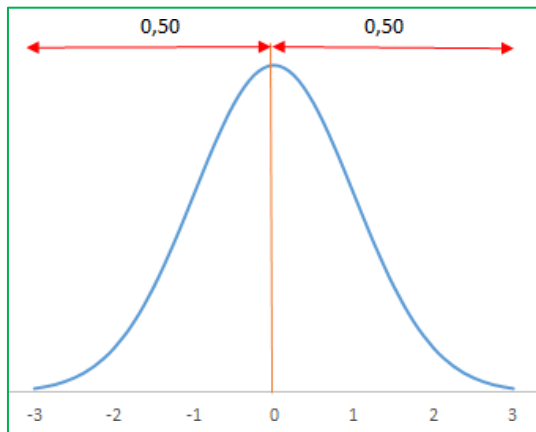
$$z_{430} = -0.67 \quad P(430) = 0.2486$$

$$z_{600} = 1.22 \quad P(600) = 0.3888$$

$$z_{500} = 0.11 \quad P(500) = 0.0438$$

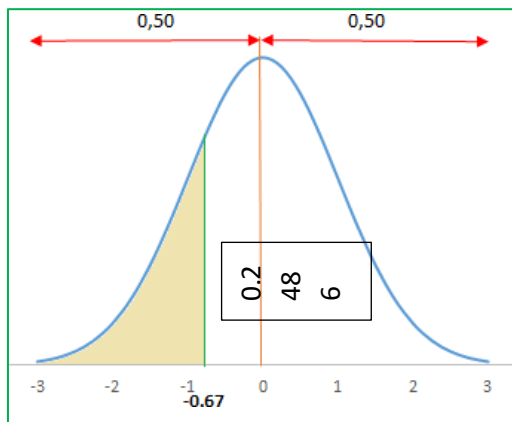
a) ¿Qué porcentaje de familias gasta más de \$30 y menos de \$490 en alimentos al mes?

$$P(30 \leq X \leq 490) = 0.5000$$



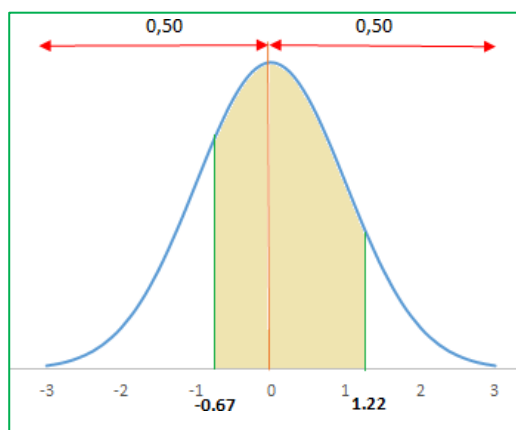
b) ¿Qué porcentaje de familias gasta menos de \$430 al mes en alimentos?

$$P(X \leq 430) = 0.5000 - 0.2486 = 0.2514$$



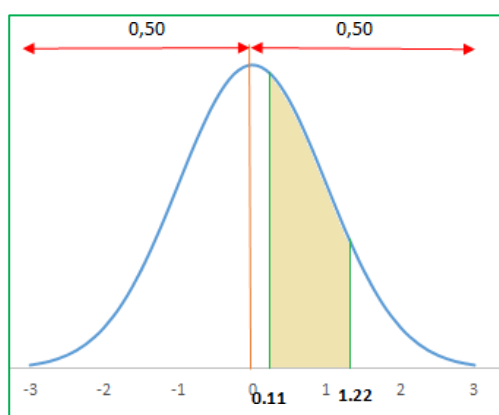
c) ¿Qué porcentaje de familias gasta entre \$430 y \$600 mensuales en alimentos?

$$P(430 \leq X \leq 600) = 0.2486 + 0.3888 = 0.6374$$



d) ¿Qué porcentaje de familias gasta entre \$500 y \$600 mensuales en alimentos?

$$P(500 \leq X \leq 600) = 0.3888 - 0.0438 = 0.345$$



65. La mayoría de las rentas de automóviles por cuatro años abarcan hasta 60 000 millas. Si el arrendador rebasa esa cantidad, se aplica una sanción de 20 centavos la milla de renta. Suponga que la distribución de millas recorridas en rentas por cuatro años tiene una distribución normal. La media es de 52 000 millas, y la desviación estándar, de 5 000 millas.

- ¿Qué porcentaje de rentas generará una sanción como consecuencia del exceso en millas?
- Si la compañía automotriz quisiera modificar los términos de arrendamiento de manera que 25% rentas rebasaran el límite de millas, ¿en qué punto debe establecerse el nuevo límite superior?
- Por definición, un automóvil de bajo millaje es uno con 4 años de uso y que ha recorrido menos de 45 000 millas. ¿Qué porcentaje de automóviles devueltos se considera de bajo millaje?

$$\mu = 52$$

$$\sigma = 5$$

$$z_{60} = 1.6$$

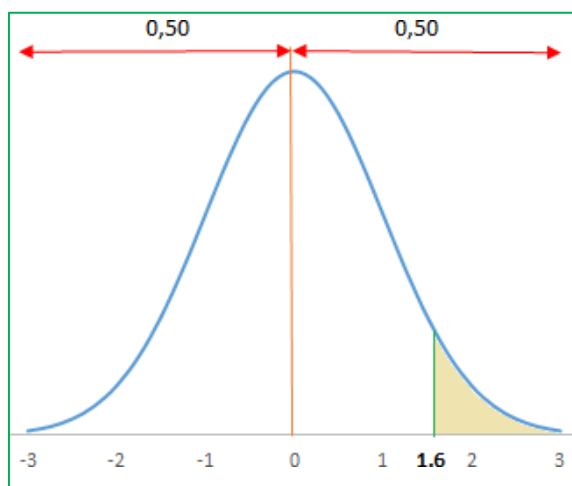
$$P(60) = 0.4452$$

$$z_{45} = -1.4$$

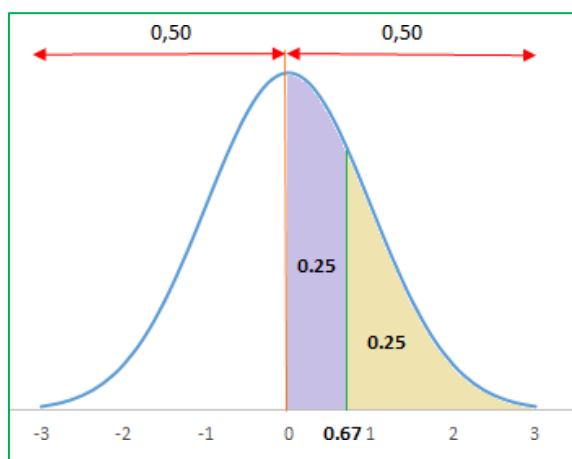
$$P(45) = 0.4192$$

a) ¿Qué porcentaje de rentas generará una sanción como consecuencia del exceso en millas?

$$P(X \geq 60) = 0.5000 - 0.4452 = 0.0548$$



b) Si la compañía automotriz quisiera modificar los términos de arrendamiento de manera que 25% rentas rebasaran el límite de millas, ¿en qué punto debe establecerse el nuevo límite superior?



$$z = 0.67 \quad z = \frac{X - \mu}{\sigma}$$

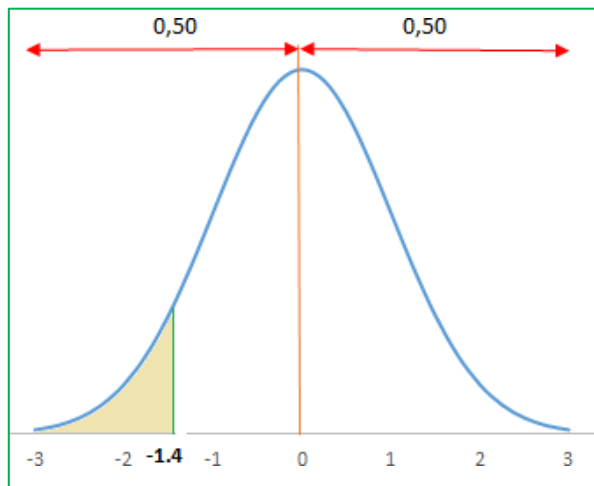
$$X - 52$$

$$0.67 = \frac{\quad}{5}$$

$$X = 55.35$$

c) Por definición, un automóvil de bajo millaje es uno con 4 años de uso y que ha recorrido menos de 45 000 millas. ¿Qué porcentaje de automóviles devueltos se considera de bajo millaje?

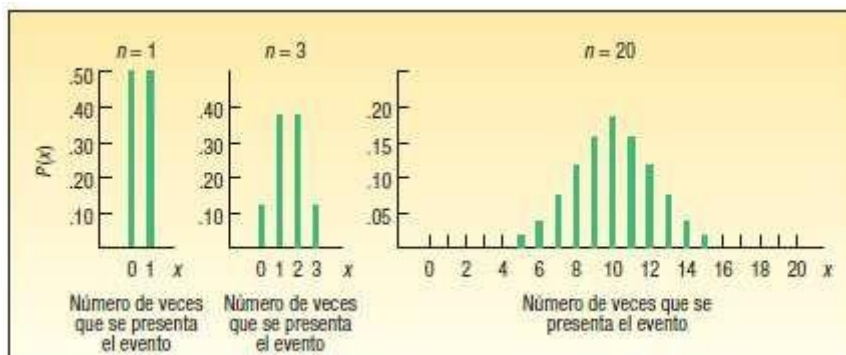
$$P(X \leq 45) = 0.500 - 0.4192 = 0.0808$$



4.4.4 Aproximación de normal a binomial

Cuando se trató la distribución binomial, que es una distribución discreta donde n es pequeña, cuando n es grande es conveniente utilizar la aproximación de la distribución normal a binomial.

Parece ser razonable utilizar la distribución normal en sustitución de la binomial, cuando n es grande la distribución binomial se aproxima cada vez más a la normal. En la gráfica se observa una distribución binomial con $\pi = 0.50$ y $n=1, 3, 20$



Tomado de texto base

La distribución de probabilidad normal constituye una buena aproximación de la distribución de probabilidad binomial cuando $n\pi$ y $n(1 - \pi)$ son 5 por lo menos. Sin embargo, antes de aplicar la aproximación normal, debe estar seguro de que la distribución de interés es en verdad una distribución binomial.

1. Sólo existen dos resultados mutuamente excluyentes en un experimento: éxito o fracaso.
2. La distribución resulta del conteo del número de éxitos en una cantidad fija de pruebas.
3. La probabilidad de un éxito, es la misma de una prueba a otra.
4. Cada prueba es independiente.

4.4.4.1 Factor de corrección de continuidad

Para interpretar la aplicación de la aproximación de la distribución normal a binomial, así como la necesidad de un factor de corrección, realicemos un ejemplo.

Ejemplo: Suponga que el 70% de los nuevos clientes de un restaurante regresan a comer. ¿Cuál es la probabilidad de que el 60 o más clientes regresen a comer durante una semana en la que 80 nuevos (primera vez) clientes comieron en el restaurante?

Paso 1: Verificar si está en condiciones binomiales.

- (a) Solo existe 2 posibles resultados, el cliente regresa o no.
- (b) Se puede contar el número de éxitos, por ejemplo 55 de los 80 regresan.
- (c) La probabilidad de que una persona regrese se mantiene constante 0.70
- (d) La prueba es independiente, significa que, si una persona regresa, no influye en que otra persona regrese.

Paso 2: Utilizar la formula binomial.

$$P(x) = nCx \cdot \pi^x(1 - \pi)^{n-x}$$

$$P(60) = {}_{80}C_{60} \cdot (0.70)^{60}(1 - 0.70)^{20} = 0.0626$$

$$P(61) = {}_{80}C_{61} \cdot (0.70)^{61}(1 - 0.70)^{19} = 0.0479$$

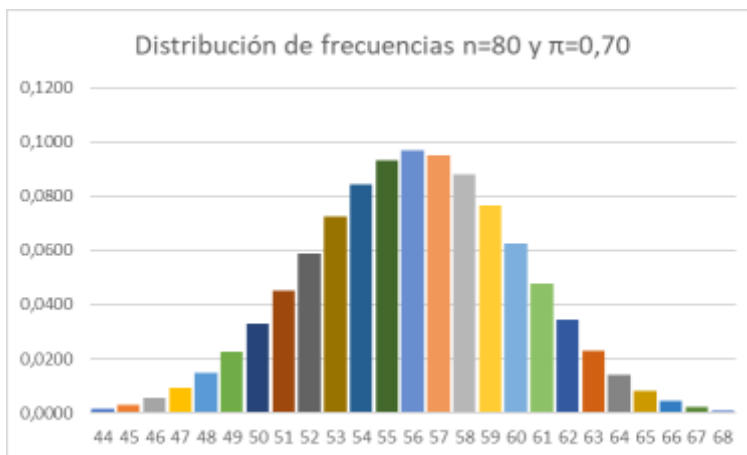
Se puede calcular todas las probabilidades hasta 80, utilizando la función en Excel, para valores menores a 40 y superiores a 70 son 0.0000 con aproximación a 4 decimales.

x	P(x)	x	P(x)
40	0,0001	55	0,0931
41	0,0002	56	0,0970
42	0,0004	57	0,0953
43	0,0008	58	0,0881
44	0,0017	59	0,0767
45	0,0031	60	0,0626
46	0,0055	61	0,0479
47	0,0093	62	0,0343
48	0,0149	63	0,0228
49	0,0227	64	0,0142
50	0,0328	65	0,0081
51	0,0451	66	0,0043
52	0,0587	67	0,0021
53	0,0723	68	0,0009
54	0,0844	69	0,0004
		70	0,0001

Podemos determinar con estos valores la probabilidad de que 60 o más regresen a comer.

$$P(X \geq 60) = 0.0626 + \dots + 0.0001 = 0.1978$$

Si se representa la distribución de probabilidades de la variable discreta, se puede observar la similitud con una distribución normal. Lo que se necesita es transformar las probabilidades discretas para obtener una distribución continua, el procedimiento consiste en permitir que la probabilidad discreta de 53 clientes quede representada por un área bajo la curva continua entre 52.5 y 53.5; después permitir que la probabilidad de 54 clientes quede representada por un área de 53.5 y 54.5, etc. Esto permite que exista continuidad diferente a un número entero.



En el ejercicio para determinar la probabilidad binomial de 60 o más éxitos, se debe restar, en este caso, 0.5 de 60. El valor 0.5 recibe el nombre de factor de corrección de continuidad.

Factor de corrección de continuidad: Valor de 0.5 restado o sumado, según se requiera, aun valor seleccionado cuando una distribución de probabilidad discreta se aproxima por medio de una distribución de probabilidad continua.

Y se aplica en los siguientes cuatro casos:

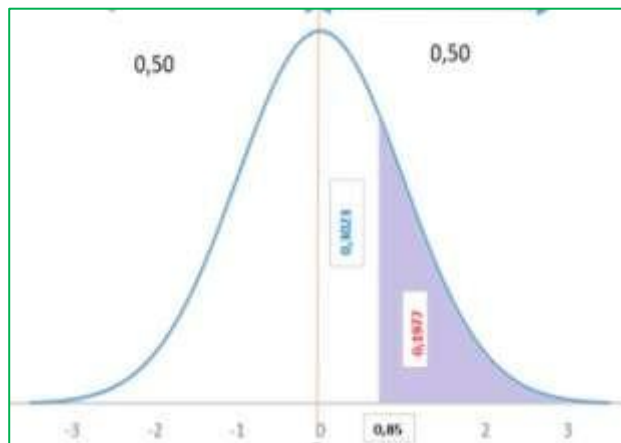
1. Para la probabilidad de que por lo menos ocurra X, se utiliza el área por encima de $(X - 0.5)$.
2. Para la probabilidad de que ocurra más que X, se utiliza el área por encima de $(X + 0.5)$.
3. Para la probabilidad de que ocurra X o menos, se utiliza el área debajo de $(X + 0.5)$.
4. Para la probabilidad de que ocurra menos que X, se utiliza el área debajo de $(X - 0.5)$.

Para utilizar en el ejemplo la probabilidad de que regresen 60 o más de los clientes que de los 80 que van al restaurante por primera vez, seguimos el procedimiento conocido.

$$\mu = n \cdot \pi = 80(0.70) = 56$$

$$\sigma^2 = n\pi(1 - \pi) = 80(0.70)(1 - 0.70) = 16.8 \quad \sigma = 4.10$$

$$z = \frac{X - \mu}{\sigma} = \frac{59.5 - 56}{4.1} = 0.85 \quad P(59.5) = 0.3023$$



$$P(x \geq 60) = 0.50 - 0.3023 = 0.1977$$

Ejemplo 1: Dottie's Tax Service se especializa en declaraciones del impuesto sobre la renta de clientes profesionales, como médicos, dentistas, contadores y abogados. Una auditoría reciente de las declaraciones que elaboraba la empresa, que llevó a cabo el Internal Revenue Service, IRS, indicó que 7% de las declaraciones que había elaborado durante el año pasado contenía errores. Si esta tasa de error continúa este año y Dottie's elabora 80 declaraciones, ¿cuál es la probabilidad de que cometa errores en:

- a) más de seis declaraciones?
- b) por lo menos seis declaraciones?
- c) seis declaraciones exactamente?

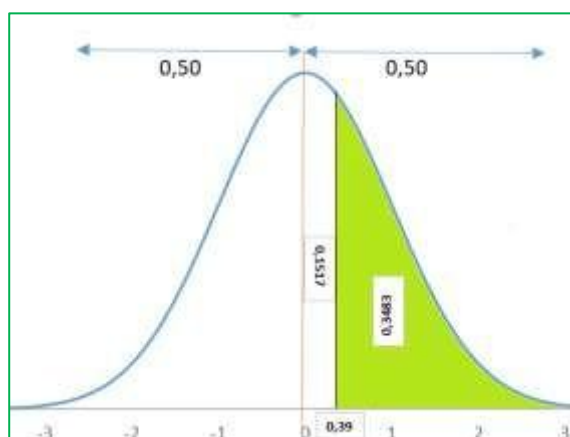
Solución.

- a) más de seis declaraciones?

$$\mu = n \cdot \pi = 80(0.07) = 5.6$$

$$\sigma^2 = n\pi(1 - \pi) = 80(0.07)(1 - 0.07) = 5.21 \quad \sigma = 2.28$$

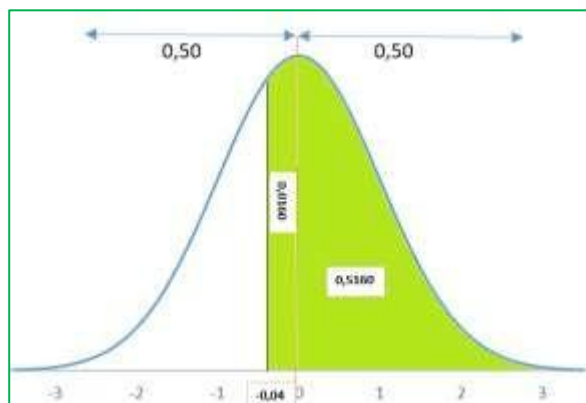
$$z = \frac{X - \mu}{\sigma} = \frac{6.5 - 5.6}{2.2821} = 0.39 \quad P(6.5) = 0.1517$$



$$P(X > 6) = 0.5 - 0.1517 = 0.3483$$

b) por lo menos seis declaraciones?

$$z = \frac{X - \mu}{\sigma} = \frac{5.5 - 5.6}{2.2821} = -0.04 \quad P(5.5) = 0.0160$$



$$P(X \geq 6) = 0.5 + 0.016 = 0.5160$$

c) seis declaraciones exactamente?

$$P(X = 6) = 0.5160 - 0.3483 = 0.1677$$

Ejemplo. Un estudio que realizó Taurus Health Club, famoso en Estados Unidos, reveló que 30% de sus nuevos miembros tiene un significativo exceso de peso. Una campaña de promoción de membresías en un área metropolitana dio como resultado la captación de 500 nuevos miembros.

a) Se sugirió utilizar la aproximación normal de la distribución binomial para determinar la probabilidad de que 175 o más de los nuevos miembros se encuentren muy excedidos de peso. ¿Es este problema de naturaleza binomial? Explique.

b) ¿Cuál es la probabilidad de que 175 o más de los nuevos miembros se encuentren muy pasados de peso?

c) ¿Cuál es la probabilidad de que 140 o más de los nuevos miembros se encuentren muy pasados de peso?

Solución.

a) Se sugirió utilizar la aproximación normal de la distribución binomial para determinar la probabilidad de que 175 o más de los nuevos miembros se encuentren muy excedidos de peso. ¿Es este problema de naturaleza binomial? Explique.

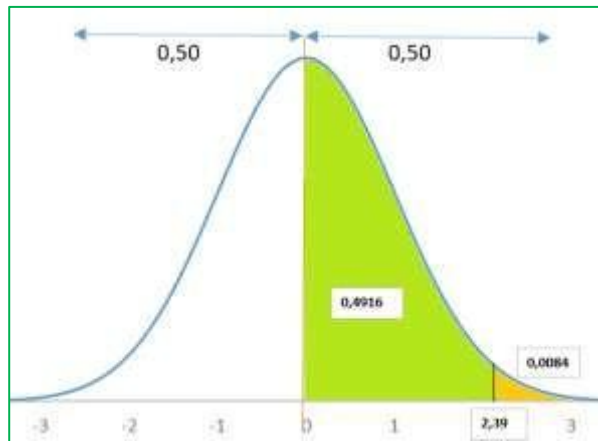
- Solo existe 2 posibles resultados, los clientes tienen exceso de peso o no.
- Se puede contar el número de éxitos, por ejemplo 130 de los 500 tienen exceso de peso.
- La probabilidad de que un cliente tiene exceso de peso se mantiene constante 0.30
- La prueba es independiente, significa que, si un cliente tiene exceso de peso, no influye en que otra cliente lo tenga.

b) ¿Cuál es la probabilidad de que 175 o más de los nuevos miembros se encuentren muy pasados de peso?

$$\mu = n \cdot \pi = 500(0.30) = 150$$

$$\sigma^2 = n\pi(1 - \pi) = 500(0.30)(1 - 0.30) = 105 \quad \sigma = 10.2470$$

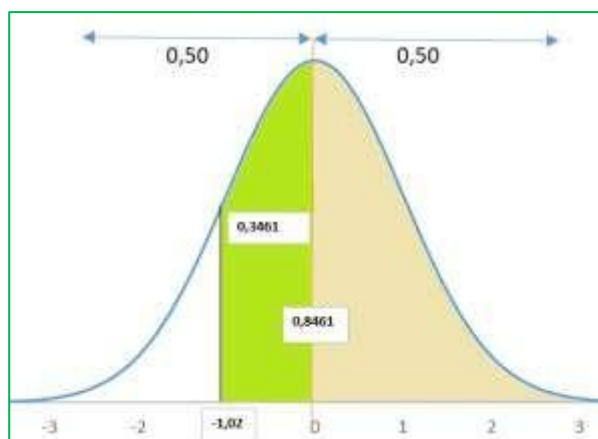
$$z = \frac{X - \mu}{\sigma} = \frac{174.5 - 150}{10.2470} = 2.39 \quad P(174.5) = 0.4916$$



$$P(X \geq 175) = 0.5 - 0.4916 = 0.0084$$

c) ¿Cuál es la probabilidad de que 140 o más de los nuevos miembros se encuentren muy pasados de peso?

$$z = \frac{X - \mu}{\sigma} = \frac{139.5 - 150}{10.2470} = -1.02 \quad P(139.5) = 0.3461$$



$$P(X \geq 140) = 0.3461 + 0.5 = 0.8461$$

4.5 Distribución probabilística exponencial

Una distribución exponencial, es una distribución continua que describe los tiempos entre eventos que ocurren en secuencia. Las acciones suceden independientemente a un ritmo constante por unidad o duración de tiempo, la variable exponencial es siempre positiva. La distribución exponencial suele describir situaciones como:

- Los tiempos de servicio en un sistema.
- El tiempo de entre entradas en un sitio web.
- El tiempo de vida de un componente eléctrico.
- El tiempo que transcurre hasta que la siguiente llamada telefónica llega a un centro de servicio al tiempo.

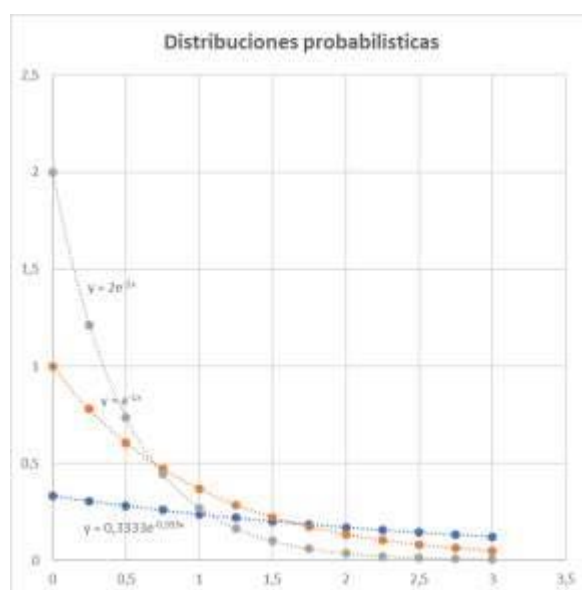
A diferencia de la distribución uniforme y normal que son simétricas la exponencial tiene un sesgo positivo, la misma que es descrita por un solo parámetro, identificado por λ , la misma que nos referiremos como un parámetro de ritmo y está determinada por: $\lambda = \frac{1}{\mu}$.

Existe una estrecha relación entre la distribución de Poisson y Exponencial, la primera por ejemplo describe la probabilidad del número de clientes lleguen a un restaurante, la segunda nos permitirá determinar el tiempo entre cada llegada de los clientes. Para determinar el tiempo de llegada utilizamos las fórmulas siguientes:

$$P(x) = \lambda e^{-\lambda x} \quad \text{Distribución exponencial}$$

$$P(\text{Tiempo de llegada} < x) = 1 - e^{-\lambda x}$$

Conforme el valor de λ , disminuye la distribución exponencial se vuelve menos sesgada, como se aprecia en la gráfica siguiente:



Ejemplo 1: Los tiempos de espera para recibir la comida después de hacer el pedido en una tienda siguen una distribución exponencial con una media de 60 segundos. Calcule la probabilidad de que un cliente espere:

- Menos de 30 segundos.
- Más de 30 segundos.
- Entre 40 y 80 segundos.

d) ¿Cincuenta por ciento de los clientes espera menos de cuántos segundos? ¿Cuál es la mediana?

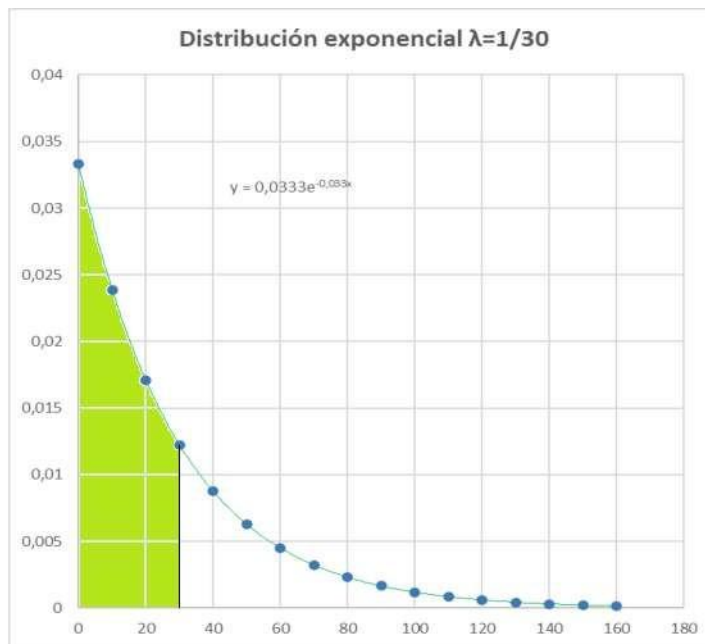
Solución.

a) Menos de 30 segundos.

$$\lambda = \frac{1}{\mu} = \frac{1}{30}$$

$$P(\text{Tiempo de llegada} < x) = 1 - e^{-\lambda x}$$

$$P(\text{Tiempo de llegada} < 30) = 1 - e^{-\frac{1}{30}(30)} = 0.6321$$



b) Más de 30 segundos.



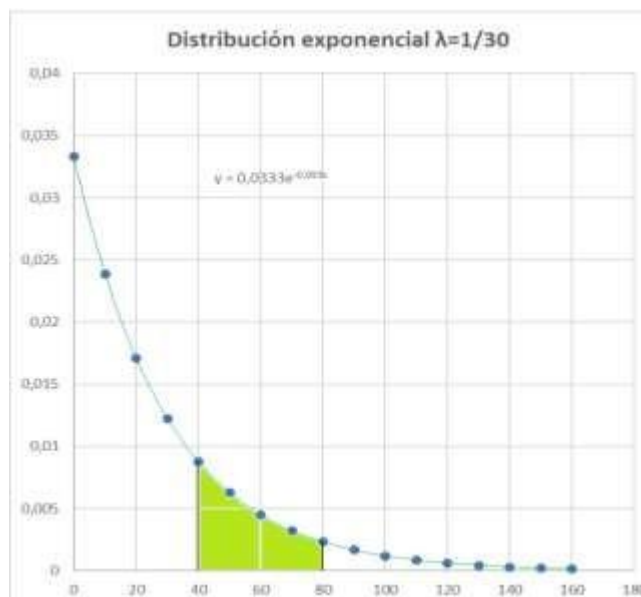
$$P(\text{Tiempo de llegada} > x) = 1 - P(\text{Tiempo de llegada} < x)$$

$$P(\text{Tiempo de llegada} > 30) = 1 - 0.6321 = 0.3679$$

$$P(\text{Tiempo de llegada} > 30) = e^{-kx}$$

$$P(\text{Tiempo de llegada} > 30) = e^{-(1/30)(30)} = 0.3679$$

c) Entre 40 y 80 segundos.



$$P(40 \leq x \leq 80) = P(\text{Llegada} \leq 80) - P(\text{Llegada} \leq 40)$$

$$P(40 \leq x \leq 80) = (1 - e^{-\frac{1}{30}(80)}) - (1 - e^{-\frac{1}{30}(40)}) = 0.1941$$

d) ¿Cincuenta por ciento de los clientes espera menos de cuántos segundos? ¿Cuál es la mediana?

$$P(\text{Tiempo de llegada} < x) = 1 - e^{-\lambda x}$$

$$0.50 = 1 - e^{-(1/30)x}$$

$$0.50 = e^{-(1/30)x} \quad \ln(0.50) = -\frac{1}{30}x$$

$$-0.6931 = -\frac{1}{30}x \quad x = 20.79 \text{ segundos (mediana)}$$

5. Anexo: Correlación y Regresión Lineal.

La correlación y regresión lineal son métodos estadísticos que estudian la relación entre dos variables de intervalo o razón.

Como ejemplos de relación entre dos variables, tenemos:

- ¿Existe alguna relación en las que una empresa gasta por mes y sus ventas totales anuales?
- ¿Existe relación entre el número de delitos en una ciudad con el número de policías destinadas al control?
- ¿El número de metros cuadrados en una casa está relacionado con el costo de la factura de energía eléctrica?

¿Qué es el análisis de correlación?

Cuando se estudia la relación entre dos variables en escala de intervalo (o de razón), es usual comenzar con un diagrama de dispersión. Este procedimiento proporciona una representación visual de la relación entre las variables. El siguiente paso suele ser calcular el coeficiente de correlación, que brinda una medida cuantitativa de la fuerza de la relación entre dos variables.

Suponga que el gerente de ventas de una empresa que vende aspiradoras, desea determinar si hay alguna relación entre el número de llamadas de ventas en un mes y el número de aspiradoras que se vendieron en él. El gerente selecciona una muestra aleatoria de 10 representantes de ventas y determina el número de llamadas de ventas que cada uno hizo el mes pasado y el número de aspiradoras que vendió. La información muestral aparece en la tabla.

Representante de ventas	Número de llamadas de ventas	Número de aspiradoras vendidas
Represente 1	20	40
Represente 2	40	70
Represente 3	20	50
Represente 4	30	70
Represente 5	10	40
Represente 6	10	50
Represente 7	20	50
Represente 8	20	60
Represente 9	20	40
Represente 10	30	80

Diagrama de dispersión



Al visualizar la tabla y el análisis de dispersión se observa que existe una relación entre el número de llamadas de ventas y el número de aspiradoras vendidas. Aunque la relación no es perfecta, se puede realizar análisis estadísticos numérico para determinar la fuerza de la relación entre las variables en estudio a través de un análisis de correlación.

ANÁLISIS DE CORRELACIÓN: Grupo de técnicas para medir la asociación entre dos variables.

La idea básica del análisis de correlación es reportar la asociación entre dos variables. La implicación es que el número de aspiradoras vendidas se relaciona con el número de llamadas de ventas. Conforme aumenta el número de llamadas de venta, parece que el número de aspiradoras vendidas también lo hace. De este modo, el número de llamadas de ventas se considera **variable independiente**, y el de copiadoras vendidas, **variable dependiente**.

Para el análisis es importante definir a la variable dependiente e independiente, en el ejemplo:

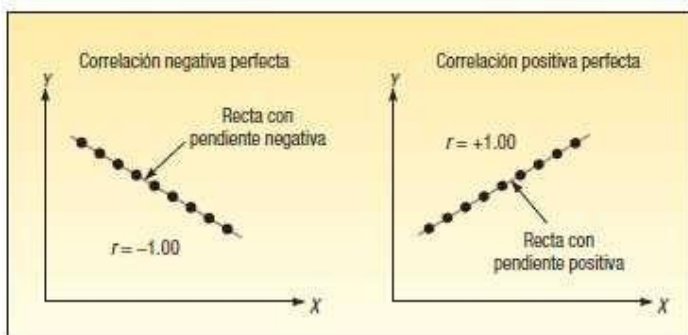
$y = f(x)$ = variable dependiente(número de aspiradoras vendidas)

x = variable independiente(número de llamadas de venta)

La variable independiente proporciona la base para la estimación, es la variable predictora, en base a la misma se puede conocer la variable dependiente.

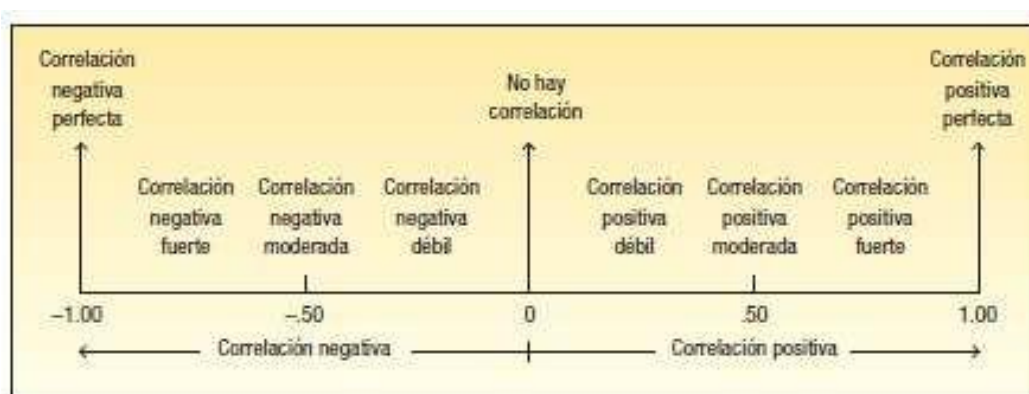
Coeficiente de correlación

El coeficiente de correlación, creado por Karl Pearson alrededor de 1900, describe la fuerza de la relación entre dos conjuntos de variables en escala de intervalo o de razón. Se designa con la letra r , y con frecuencia se le conoce como r de Pearson y coeficiente de correlación producto-momento. Puede adoptar cualquier valor de -1.00 a +1.00, inclusive. Un coeficiente de correlación de -1.00 o bien de +1.00 indica una correlación perfecta. En la gráfica se muestra cómo aparecería el diagrama de dispersión si la relación entre los dos conjuntos de datos fuera lineal y perfecta.



Si no hay ninguna relación entre los dos conjuntos de variables, la r de Pearson es cero.

En el siguiente diagrama se resumen la fuerza y la dirección del coeficiente de correlación.



COEFICIENTE DE CORRELACIÓN: Medida de la fuerza de la relación lineal entre dos variables.

Las características del coeficiente de correlación se resumen a continuación.

CARACTERÍSTICAS DEL COEFICIENTE DE CORRELACIÓN

1. El coeficiente de correlación de la muestra se identifica con la letra minúscula r .
2. Muestra la dirección y fuerza de la relación lineal (recta) entre dos variables en escala de intervalo o en escala de razón.
3. Varía de -1 hasta +1, inclusive.
4. Un valor cercano a 0 indica que hay poca asociación entre las variables.
5. Un valor cercano a 1 indica una asociación directa o positiva entre las variables.
6. Un valor cercano a -1 indica una asociación inversa o negativa entre las variables.

Cálculo del coeficiente de correlación: Continuamos con el ejemplo que estamos desarrollando.

Representante de ventas	Llamadas X	Ventas Y	$X - \bar{X}$	$Y - \bar{Y}$	$(X - \bar{X})(Y - \bar{Y})$
Represente 1	20	40	-2	-15	30
Represente 2	40	70	18	15	270
Represente 3	20	50	-2	-5	10
Represente 4	30	70	8	15	120
Represente 5	10	40	-12	-15	180
Represente 6	10	50	-12	-5	60
Represente 7	20	50	-2	-5	10
Represente 8	20	60	-2	5	-10
Represente 9	20	40	-2	-15	30
Represente 10	30	80	8	25	200
Media	22	55	0	0	0
Desv. Estándar	9,1894	14,3372			900

Coefficiente de correlación:

$$r = \frac{\sum(X - \bar{X})(Y - \bar{Y})}{(n - 1)s_x s_y}$$

$$r = \frac{\sum(X - \bar{X})(Y - \bar{Y})}{(n - 1)s_x s_y} = \frac{900}{(10 - 1)(9.1894)(14.3372)} = 0.7590$$

Como **r** es positiva, se tiene una relación directa entre el número de llamadas de ventas y el número de aspiradoras vendidas. El valor de 0.759 está muy cercano a 1.00, y por ende se concluye que la asociación es fuerte.

Análisis de regresión

Expresada la fuerza y la dirección de la relación lineal entre dos variables a través del coeficiente de correlación. En esta sección se elabora una ecuación para expresar la relación lineal entre dos variables. Además, se desea estimar el valor de la variable dependiente Y con base en un valor seleccionado de la variable independiente X. La técnica para desarrollar la ecuación y proporcionar las estimaciones se denomina análisis de regresión.

ECUACIÓN DE REGRESIÓN: Ecuación que expresa la relación lineal entre dos variables.

Forma General de la ecuación de regresión: **$Y = a + bX$**

donde:

Y : que se lee Y prima, es el valor de la estimación de la variable Y para un valor X seleccionado.

a : es la intersección Y.

b : es la pendiente de la recta, o el cambio promedio en por cada cambio de una unidad (ya sea aumento o reducción) de la variable independiente X.

X es cualquier valor de la variable independiente que se seleccione.

Pendiente de la recta de regresión: **$b = r \frac{s_y}{s_x}$**

donde:

r : es el coeficiente de correlación.

s_y : es la desviación estándar de Y (la variable dependiente).

s_x : es la desviación estándar de X (la variable independiente).

Intersección con el eje y: **$a = Y - bX$**

donde:

Y : es la media de Y (la variable dependiente).

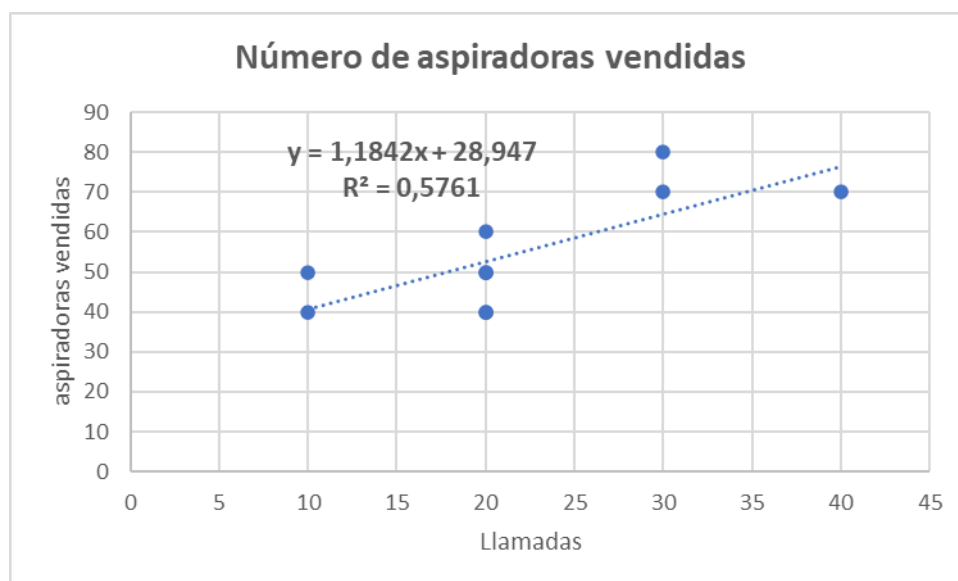
X : es la media de X (la variable independiente).

Ejemplo: Utilizando los datos del ejemplo anterior, tenemos:

$$b = r \frac{s_y}{s_x} \quad b = 0.7590 \frac{14.3372}{9.1894} = 1.1842$$

$$a = Y - bX \quad a = 55 - (1.1842)(22) = 28.9476$$

$$Y = a + bX \quad Y = 28.9476 + 1.1842X$$



Con la ecuación de regresión se puede predecir el número de aspiradoras vendidas en relación al número de llamadas, por ejemplo:

- $X = 25$ llamadas

$$Y = 28.9476 + 1.1842X$$

$$Y = 28.9476 + 1.1842(25)$$

$$Y = 58.55 \cong 59 \text{ aspiradoras vendidas}$$

El coeficiente de determinación

El error estándar de estimación proporciona una medida relativa de la capacidad de predicción de una ecuación de regresión, a través del coeficiente de determinación.

COEFICIENTE DE DETERMINACIÓN: Proporción de la variación total de la variable dependiente Y que se explica, o contabiliza, por la variación de la variable dependiente X.

En el ejemplo que hemos estado analizando $r = 0.759$; el coeficiente de determinación $r^2 = 0.576$. Así, se dice que 57.6% de la variación del número de aspiradoras vendidas se explica, o está representado por la variación del número de llamadas de ventas.

Nuestro análisis muestra que sólo 57.6% de la variación del número de copadoras vendidas se explica por la variación del número de llamadas de ventas que se realizó. Es claro que estos datos no forman una línea perfecta.

El coeficiente de determinación proporciona una medida cuantitativa de la calidad del modelo para predecir o explicar la variable de interés.

Bibliografía

- Douglas A. Lind, William G. Marchal y Samuel A. Wathen, “ESTADÍSTICA APLICADA A LOS NEGOCIOS Y LA ECONOMÍA”, decimoquinta edición. 2012
- Llinás Solano, Humberto, “Estadística descriptiva y distribuciones de probabilidad”. Universidad del Norte 2017.
- Leonard J. Kazmier, Alfredo Díaz Mata, and Guillermina Eslava Gómez, “Estadística aplicada a la administración y la economía”. 2013.