compAs
Grupo de capacitación e investigación pedagógica

RED HEMABIGDAT
Senescyt-2018-040_ Reg. Red –18-0011
"Herramientas de Estadística Multivariante para el Análisis de Big Data".

APUNTES DE CLASE ESTADISTICA BASICA
Estadística y Probabilidad: Teoría y Ejercicios

APUNTES DE CLASE ESTADISTICA BASICA

Estadística y Probabilidad: Teoría y Ejercicios

Ing. Juan Calderón Cisneros. Mgs
UNIVERSIDAD ESTATAL DE MILAGRO

Ing. Carlos Triviño Ibarra, Mgs.
UNIVERSIDAD DE GUAYAQUIL

Msc. Laura del Rocio Cardenas Zhuma
UNIVERSIDAD DE GUAYAQUIL

Dra. June Lady Robles Amaya. MAE
UNIVERSIDAD ESTATAL DE MILAGRO

Msc. Sara Santillan Indacochea
UNIVERSIDAD DE GUAYAQUIL

Ing. Mauricio Guillen Godoy. MBA
UNIVERSIDAD ESTATAL DE MILAGRO

Ing. Javier Alcazar Espinoza. Mgs
UNIVERSIDAD ESTATAL DE MILAGRO

Ing. Paúl Freire Sierra, Mgs.
UNIVERSIDAD DE GUAYAQUIL

Lcda. Vanessa Vargas Párraga. Mgs
UNIVERSIDAD UNIR

Ing. Roberto Basurto Q. MAE
UNIVERSIDAD ESTATAL DE MILAGRO

Lcdo. Bolívar R. Duchi Ortega Mg.
UNIVERSIDAD ECOTEC

Econ. Ernesto Rangel Luzuriaga. M.Sc. Ph.D (c)
UNIVERSIDAD TECNOLÓGICA
EMPRESARIAL DE GUAYAQUIL UTEG

Ing. Salomón Arias Montero, Mgs.
UNIVERSIDAD TECNICA DE MACHALA

Ing. Glenda Blanc. MAE
UNIVERSIDAD ECOTEC

Ing. Virginia Zambrano Zambrano, MBA
UNIVERSIDAD DE GUAYAQUIL

Ing. Luis Palacios Torres. Mgs
UNIVERSIDAD ESTATAL DE MILAGRO

Mdc. Carlos Alcívar Trejo
UNIVERSIDAD ECOTEC

Msc. Roberto Israel Soria Vera
UNIVERSIDAD DE GUAYAQUIL

Lcda. Imelda Arias Montero, MSc.
UNIVERSIDAD ESTATAL DE MILAGRO

Ing. Juan Gonzalez Vaca, MSc.
ASESOR FINANCIERO

Ing. Marcos Guerrero Zambrano. Mgs
UNIVERSIDAD ESTATAL DE MILAGRO

Eco. Yecenia Escobar De la Cuadra. Mgs
UNIVERSIDAD ESTATAL DE MILAGRO

Ing. Zoila Noemi Merino Acosta, PhD
UNIVERSIDAD TÉCNICA DE BABAHOYO

Lcdo. Víctor Hugo Del Pozo R. Mgs.
UNIVERSIDAD DE GUAYAQUIL

APUNTES DE CLASE ESTADISTICA BASICA. Estadística y Probabilidad: Teoría y Ejercicios

compÁs
Grupo de capacitación e investigación pedagógica

RED HEMABIGDAT
Senescyt-2018-040_ Reg. Red -18-0011.
"Herramientas de Estadística Multivariante para el Análisis de Big Data".

Título original:
APUNTES DE CLASE ESTADISTICA BASICA
Estadística y Probabilidad: Teoría y
Ejercicios

Primera edición: septiembre 2020

Ing. Carlos Triviño Ibarra, Mgs.
Msc. Laura del Rocio Cardenas Zhuma
Dra. June Lady Robles Amaya. MAE
Msc. Sara Santillan Indacochea
Ing. Mauricio Guillen Godoy. MBA
Ing. Javier Alcazar Espinoza. Mgs
Ing. Paúl Freire Sierra, Mgs.
Lcda. Vanessa Vargas Párraga. Mgs
Ing. Roberto Basurto Q. MAE
Ing. Glenda Blanc. MAE
Ing. Virginia Zambrano Zambrano, MBA
Ing. Luis Palacios Torres. Mgs
Mdc. Carlos Alcívar Trejo
Msc. Roberto Israel Soria Vera
Lcda. Imelda Arias Montero, MSc.
Ing. Juan Gonzalez Vaca, MSc.
Ing. Marcos Guerrero Zambrano. Mgs
Eco. Yecenia Escobar De la Cuadra. Mgs
Lcdo. Bolívar R. Duchi Ortega Mg.
Econ. Ernesto Rangel Luzuriaga. M.Sc. Ph.D (c)
Ing. Salomón Arias Montero, Mgs.
Ing. Zoila Noemi Merino Acosta, PhD
Lcdo. Víctor Hugo Del Pozo R. Mgs.

Editado en Guayaquil - Ecuador

ISBN: 978-9942-33-286-8

compÁs
Grupo de capacitación e investigación pedagógica

ÍNDICE

CAPITULO 1

Autores:

Ing. Juan Calderón Cisneros. Mgs
UNIVERSIDAD ESTATAL DE MILAGRO

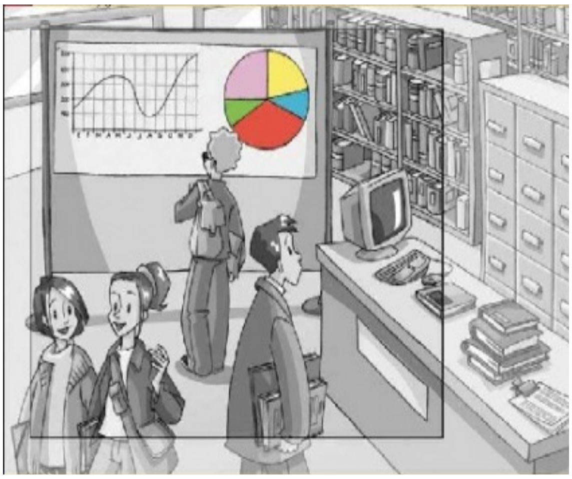
Ing. Carlos Triviño Ibarra, Mgs.
UNIVERSIDAD DE GUAYAQUIL

Msc. Laura del Rocio Cardenas Zhuma
UNIVERSIDAD DE GUAYAQUIL

"Democracia: es una superstición muy difundida, un abuso de la estadística."
JORGE LUIS BORGES

CAPÍTULO 1

La Estadística



OBJETIVOS

Saber qué significa estadística. Explicar qué es estadística descriptiva y estadística inferencial. Diferenciar entre una variable cualitativa y una variable cuantitativa. Distinguir entre una variable discreta y una variable continua. Diferenciar entre niveles de medición nominal, ordinal, por intervalo y de razón. Definir los términos mutuamente excluyentes y exhaustivos

Desarrollo:

Estadística es la ciencia que nos enseña a recolectar, organizar, presentar, analizar e interpretar datos para convertirlos en información útil que nos ayude a tomar decisiones más efectivas (Díaz-levicoy & Arteaga Cezón Dra Carmen Batanero Bernabeu, 2014).

Comentario:

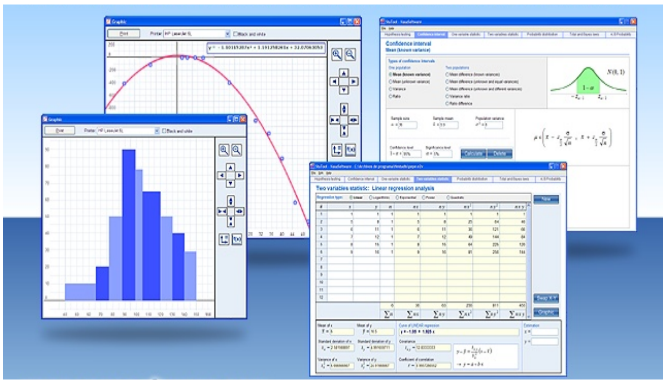
Como se puede observar, Estadística es una ciencia que cumple con cinco pasos muy importantes, la misma que nos permitirá tomar mejores decisiones.

- ***Ejemplos de Estadística***

Los analistas de investigación de una empresa evalúan muchas facetas de una acción en especial antes de elaborar una recomendación de "compra" o "venta". Recolectan los datos de ventas pasadas de la compañía y calculan las ganancias futuras. Asimismo, se consideran también otros factores antes de hacer una recomendación, como serían la demanda mundial proyectada de los productos *de la compañía, la fortaleza de la competencia, y el efecto del nuevo contrato entre gerencia y sindicato.*

Comentario:

En toda compañía en la cual se desea saber el nivel de aceptación que tienen los productos que se ofrecen a los consumidores, se utiliza la estadística la cual nos permitirá mediante cuestionarios saber que tan bien posesionado se encuentra un producto en el mercado y cuáles son las sugerencias que tienen los clientes o que esperan los clientes obtener de nuestro producto; la misma que nos permitirá hacer sugerencias acerca del producto. Así como saber las ventas futras en base a las ventas anteriores, o también cuantos insumos necesitarán comprar en el siguiente mes para poder satisfacer las necesidades del mercado.



EJEMPLO: El departamento de mercadotecnia de Unilever, compañía productora de jabones, tiene la responsabilidad de presentar recomendaciones acerca de la ganancia potencial de un nuevo grupo de jabones faciales recién desarrollados con olor a frutas, como uva, naranja y piña. Antes de tomar una decisión final, realizarán pruebas en diferentes mercados. Esto quiere decir que los puede presentar y vender en Guayaquil, Quito y Cuenca. Unilever tomará la decisión de introducir o no los jabones a nivel nacional basándose en las pruebas de mercadotecnia en ambas regiones.

Comentario:

En este ejemplo podemos observar que la compañía Unilever, antes de poner un producto en el mercado realiza un análisis de investigación de mercados, mediante la entrega de jabones de muestra el mismo que nos permitirá saber si las personas que lo reciben les agrada el producto y si estarían dispuestos a comprarlo en el momento que salgan al mercado, así como cuanto estarían dispuesto a pagar, que recomendaciones le harían, y en qué lugar le gustaría que se expendieran estos productos.

¿Quién usa la Estadística?

Las técnicas estadísticas se usan ampliamente por personas en áreas de comercialización, contabilidad, control de calidad, consumidores, deportes, administración de hospitales, educación, política, medicina, etcétera...

Comentario:

Como se puede observar las estadísticas se pueden usar en cualquier rama profesional como por ejemplo la medicina, en la cual los médicos pueden saber con qué tipos de medicamentos se puede obtener mejores resultados en el tratamiento de ciertas enfermedades. En el área de la Política se pueden usar las estadísticas para saber que tan bien esta posesionado un candidato en alguna contienda electoral, y con los resultados poder tomar las mejores decisiones y correcciones si las hubiese.

¿Por qué hay que estudiar Estadística?

Existen tres razones muy importantes:

1. ***EN TODO LUGAR ENCONTRAMOS INFORMACIÓN NUMÉRICA.***

EJEMPLOS:

- *Los egresados de postgrado del programa de MBA en la UNEMI, contaron con un sueldo promedio inicial de $24,000.00 dólares y el 91% de ellos consiguieron trabajo a los tres meses de la graduación.*
- *En EE.UU. se consume mayor cantidad de café que en cualquier otro país, en promedio, 1.75 tazas por persona al día.*

2. ***LAS TÉCNICAS ESTADÍSTICAS SE UTILIZAN PARA TOMAR DECISIONES QUE AFECTAN A NUESTRA VIDA DIARIA.***

EJEMPLOS:

- *Las compañías de seguros utilizan análisis estadísticos para establecer las tarifas de los seguros de casa, automóvil, vida y salud. Existen tablas que resumen la probabilidad de que una mujer de 25 años de edad viva el año siguiente, los siguientes 5 años, etc. Las primas de seguro de vida se pueden establecer basándose en estas probabilidades.*
- *La Agencia de Protección al Medio Ambiente está interesada en la calidad del agua en el Estero Salado. Periódicamente toman muestras de agua para establecer el nivel de contaminación y mantener el nivel de calidad.*

3. ***SIN IMPORTAR SU LÍNEA DE TRABAJO FUTURA, UNO TOMARÁ DECISIONES QUE IMPLICAN DATOS.***

    - ***Para poder realizar una decisión basada en la información, necesitará:***

- *Determinar si la información existente es adecuada o si se requiere información adicional.*

- *Reunir información adicional, si es necesario, de tal forma que no haya resultados erróneos.*
- *Resumir la información de una forma útil e informativa.*
- *Analizar la información disponible.*
- *Sacar las conclusiones y realizar las deducciones necesarias, al tiempo que se evalúa el riesgo de llegar a una conclusión incorrecta.*

Tipos de Estadísticas

1. *Estadística Descriptiva*
2. *Estadística Inferencial*

Estadística Descriptiva:

Conjunto de métodos para organizar, resumir y presentar datos de manera informativa(GARG, 2018)**.**

- *Los conjuntos de datos no organizados son de poco valor.*
- *Sin embargo, hay técnicas estadísticas disponibles para organizar este tipo de información de manera significativa.*
- *Algunos datos pueden organizarse en una distribución de frecuencias.*
- *Los promedios específicos, como la media, pueden calcularse para describir el valor central de un grupo de datos numéricos.*

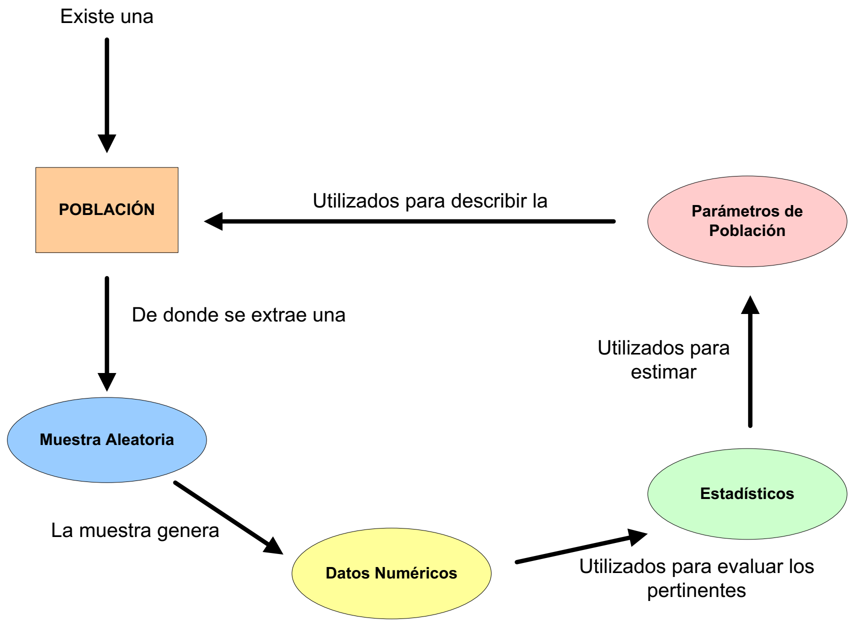
Ejemplos de Estadística Descriptiva

- ***EJEMPLO 1:*** *Un sondeo de opinión encontró que el 49% de las personas en una encuesta sabían el nombre del primer libro en la Biblia. La estadística "49" describe el número de cada 100 personas que saben la respuesta.*
- ***EJEMPLO 2:*** *Según el Consumer Reports, los dueños de lavadoras de ropa Whirlpool reportaron 9 problemas por cada 100 máquinas durante 1995. La estadística "9" describe el número de problemas por cada 100 máquinas.*

Estadística Inferencial:

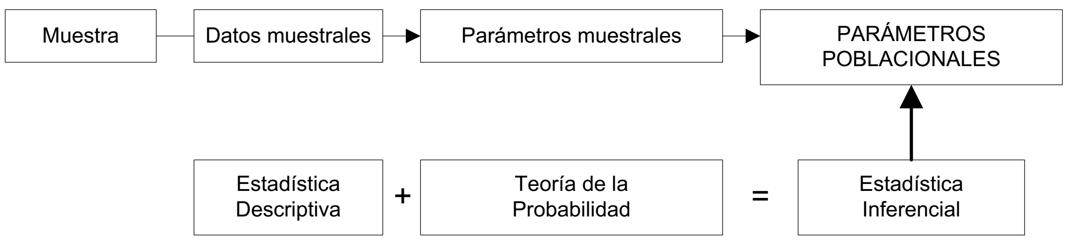
Conjunto de métodos utilizados para saber algo acerca de una población, con base en una muestra(Boslaugh & Watters, 2008)**.**

- *La principal preocupación referente a la Estadística Inferencial es encontrar algo sobre una población basada en una muestra tomada de esa población.*



Población: *Es un conjunto de todos los posibles individuos, objetos o medidas de interés.*

Muestra: *Es una porción o parte de la población de interés.*



Ejemplos de Estadística Inferencial

- ***EJEMPLO 1:*** *Las cadenas de TV monitorean la popularidad de sus programas contratando a alguna organización de consultoría para muestrear las preferencias de los televidentes.*

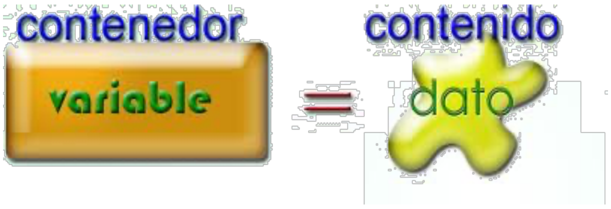
- ***EJEMPLO 2:*** *El departamento de contabilidad de una empresa elegirá una muestra de facturas para verificar la exactitud de todas las facturas de la compañía.*

- ***EJEMPLO 3:*** *Los catadores de vino prueban unas cuantas gotas para tomar la decisión de liberar todo el vino para la venta.*

Tipos de Variables:

1. *Los obtenidos a través de una población cualitativa.*

2. *Los obtenidos a través de una población cuantitativa.*



Variable cualitativa o de atributos:

Es cuando la característica o variable que se estudia no es numérica.

- *Cuando la información estudiada es cualitativa, generalmente nos interesa saber cuántas o que proporción entra en cada categoría.*

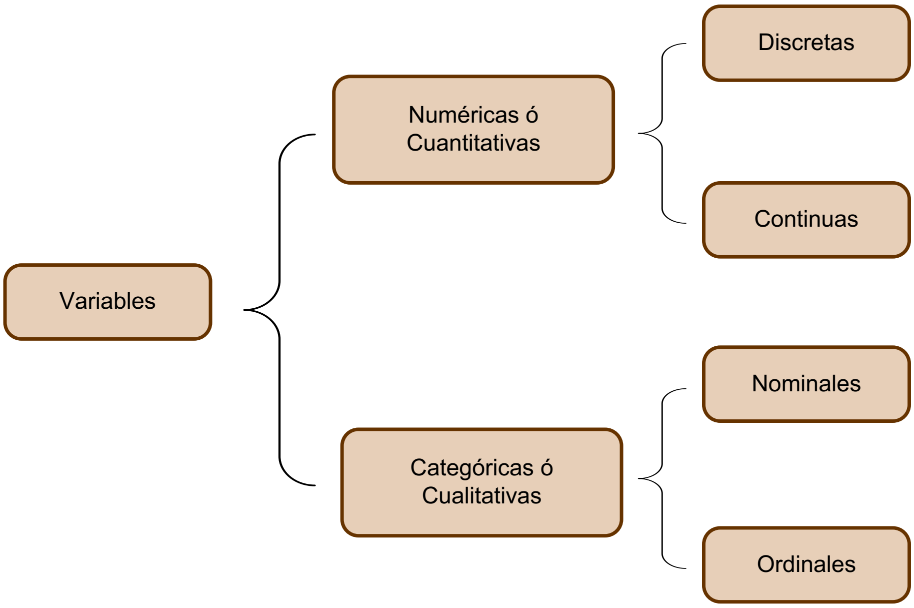
- ***EJEMPLOS:*** *Sexo, afiliación religiosa, tipo de automóvil que se posee, lugar de nacimiento, color de los ojos.*

Variable cuantitativa:

Es cuando la variable estudiada se puede expresar numéricamente.

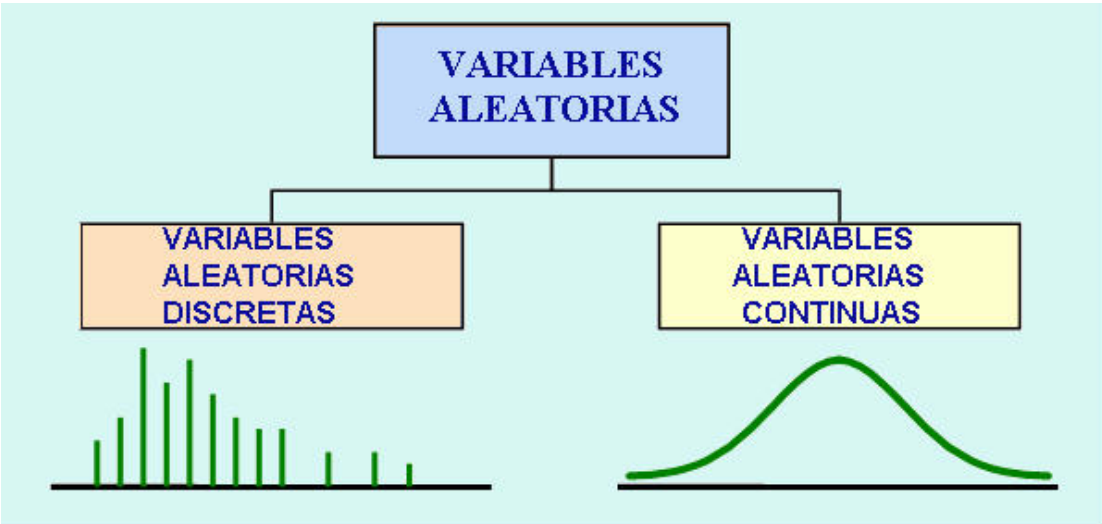
- *La población se conoce como población cuantitativa.*

- ***EJEMPLOS:*** *Saldo en una cuenta de cheques, minutos que faltan para que termine la clase, número de niños en una familia.*



Las variables cuantitativas se pueden clasificar como:

- *Variables Discretas*
- *Variables Continuas.*



Variables discretas:

Sólo pueden adquirir ciertos valores y casi siempre hay "brechas" entre esos valores. Por lo común las variables discretas casi siempre resultan del conteo(Laverty & Kelly, 1995)**.**

- ***EJEMPLO1:*** *el número de habitaciones en una casa (1, 2, 3,..., etc.).*
- ***EJEMPLO 2:*** *El número de vehículos que pasan por el peaje para ir a la costa.*

- ***EJEMPLO 3:*** *El número de vehículos que ingresan a la Universidad de Guayaquil desde las 6:00 hasta la 6:35 a.m.*

- ***EJEMPLO 4:*** *El número de estudiantes en el curso de Estadística del laboratorio B.*

Comentario:

Las variables discretas como bien lo dice resultan de un conteo, el mismo que tiene brechas o espacios entre cada valor. Como podemos observar en el Ejemplo1 cuando contamos el número de habitaciones en una casa podemos ver que hay 1, 2, 3,... habitaciones el hecho de que haya un cuarto pequeño esto no quiere decir que sea media habitación al contrario se debe de contar como una habitación. Otro ejemplo es el número de hermanos que una persona posee, por mucho que tenga un hermano que es un bebé, ha éste se lo debe de contar como una persona como tal.

Variables Continuas:

Pueden tomar cualquier valor dentro de un intervalo específico. Las variables continuas resultan generalmente de medir algo.

- ***EJEMPLO 1:*** *El tiempo que toma volar de Miami a New York.*

- ***EJEMPLO 2:*** *La presión del aire en una llanta.*

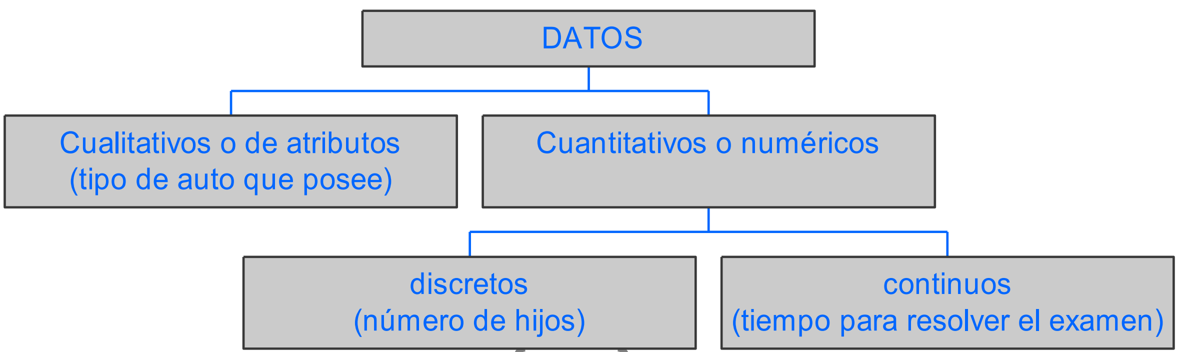
- ***EJEMPLO 3:*** *La hora en cada uno de los relojes de los alumnos en clase de estadística.*

- ***EJEMPLO 4:*** *La cantidad de dinero que tienen cada uno de ustedes en sus bolsillos en este momento.*

Comentario:

Como podemos observar las Variables Continuas resultan de una medición, como podemos observar en el Ejemplo1, el tiempo que toma volar de Miami a New York, puede tomar 2 horas, 2h15, 2h15 minutos 30 segundos y así sucesivamente. Y así mismo en los otros ejemplos.

RESUMEN



Fuentes de Datos Estadísticos

- *Los problemas de investigación suelen requerir datos publicados. Se pueden encontrar estadísticas relacionadas en artículos publicados, revistas y periódicos.*
- *No todos los temas disponen de datos publicados. En esos casos, la información deberá recolectarse y analizarse.*
- *Una manera de recolectar datos es mediante encuestas.*

Niveles de Medición

Generalmente el nivel de medición de un dato, marca los cálculos que pueden realizarse para resumir y presentar la información, y las pruebas estadísticas que pueden desarrollarse.

Tipos de Niveles de medición.

1. *Datos de nivel Nominal*
2. *Datos de nivel Ordinal*
3. *Datos de nivel de Intervalo*
4. *Datos de nivel de Razón*

Nivel nominal:

Los datos sólo se pueden clasificar en categorías ó contarse, y no se pueden ordenar. No existe un orden específico para las marcas.

Comentario:

Como su nombre lo dice el Nivel Nominal, nómina o etiqueta a una variable, la misma que no se puede medir, pero si se puede contar.

- ***Cuenta con las siguientes propiedades:***

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

- ***EJEMPLOS:*** *Color de los ojos, sexo, afiliación religiosa, etc...*

Mutuamente excluyente:

Cuando una persona, objeto o artículo, al ser incluido en una categoría, debe excluirse de las demás.

- ***EJEMPLO 1:*** *color de los ojos.*

- ***EJEMPLO 2:*** *El matrimonio.*

Comentario:
Esta propiedad indica que una persona, objeto o artículo, puede ser incluido o contado en tan solo una categoría es decir si ya fue incluido en una categoría ya no puede estar en la demás.

Nivel ordinal:

Involucra datos que se pueden ordenar, pero no es posible determinar las diferencias entre los valores de los datos o no tienen significado.

- ***Las propiedades del nivel ordinal son:***

  1. *Las categorías son mutuamente excluyentes y exhaustivas.*
  2. *Dichas categorías se pueden clasificar u ordenar de acuerdo con las características particulares que poseen.*

- ***EJEMPLO:*** *Como clasificaría la enseñanza del profesor de matemáticas: excelente, muy buena, buena, regular, y pésima.*

Comentario:

En el Nivel Ordinal a diferencia del Nivel Nominal el orden es importante el mismo que puede ser ordenado de Mayor a Menor o de Menor a Mayor. Los mismos que pueden ser Categorías o Numéricas.

Exhaustivo:

Cada persona, objeto o artículo debe clasificarse en al menos una categoría.

- ***ELEMPLO 1:*** *afiliación religiosa.*

- ***EJEMPLO 2:*** *Las materias en que se encuentran registradas en la universidad.*

Comentario:

Esta propiedad indica que una persona, objeto o artículo, puede ser incluido o contado en una o más categorías, es decir que puede ser incluido en como mínimo una categoría.

Nivel de Intervalo:

Similar al nivel ordinal, con la propiedad adicional de que se pueden determinar cantidades significativas de las diferencias entre los valores. No existe un punto cero naturales.

- ***Propiedades del nivel de intervalo:***

  1. *Las categorías son mutuamente excluyentes y exhaustivas.*
  2. *Las citadas características se clasifican de acuerdo con la cantidad de características que poseen.*
  3. *Las diferencias iguales en la característica son representadas por iguales diferencias en los números asignados a las categorías.*

- ***EJEMPLO:*** *Los grados de temperatura en la ciudad de Guayaquil van de 38, 37, 40 y 35 grados centígrados estas temperaturas se pueden medir fácilmente pero también podemos determinar la diferencia entre las mismas.*

  Comentario:
  En este Nivel de Intervalo, sus intervalos son fijos o constantes es decir no varían, y este nivel no tiene un punto cero naturales es decir que este punto no indica la ausencia de algo por ejemplo cero grados en temperatura no significa que no haga ni calor ni frío al contrario significa que está haciendo mucho frío.

Nivel de razón:

El nivel de intervalo con un punto cero iniciales inherentes. Las diferencias y razones son significativas para este nivel de medición.

- ***Propiedades del nivel de razón:***

  1. *Las categorías son mutuamente excluyentes y exhaustivas.*
  2. *Dichas categorías se pueden clasificarse de acuerdo con la cantidad de la característica que poseen.*
  3. *Las diferencias iguales en las características son representadas por diferencias iguales en los números asignados a las categorías.*
  4. *El punto (o valor) 0 representa una ausencia de la categoría.*

  EJEMPLOS: *Dinero, altura de los jugadores de básquetbol de la NBA.*

EJEMPLO 2: *El peso es otro ejemplo, ya que si la escala de la báscula está en cero hay una ausencia de peso.*

Comentario:

- *En este Nivel de Razón, al igual que el nivel de Intervalo, los intervalos son constantes es decir no varían, y este nivel tiene un punto cero naturales es decir que este punto indica la ausencia de algo por ejemplo si una báscula marca cero indica la ausencia de algo.*

AUTOEVALUACIÓN # 1

CONTESTE:

1. Para poder realizar una decisión basada en la información, se necesitará:

____

____

2. Indique los Tipos de Estadísticas:

____

____

3. ***Escriba 2 ejemplos de Estadística Descriptiva.***

EJEMPLO 1: ____

____

EJEMPLO 2: ____

____

4. ***Estadística Inferencial es:*** ____________________________________

____

5. ***Unidades experimentales Identifique las unidades experimentales en los que se miden las variables siguientes:***
a. Género de un estudiante
b. Número de errores en un examen de medio semestre
c. Edad de un paciente con cáncer
d. Número de flores en una planta de azalea

e. Color de un auto que entra a un estacionamiento

6. ***Población y Muestra cuáles son sus diferencias:***

7. ***Indique 3 ejemplos de Estadística Inferencial***

EJEMPLO 1:

EJEMPLO 2:

EJEMPLO 3:

8. ***Escriba los Tipos de Variables:***

9. ***Indique que variables son cualitativas y cuales cuantitativas:***

   a. Comida Favorita.

   b. Profesión que te gusta.

   c. Número de goles marcados por tu equipo favorito en la última temporada.

   d. Número de alumnos de tu Instituto.

   e. El color de los ojos de tus compañeros de clase.

   f. Coeficiente intelectual de tus compañeros de clase.

10. ***De las siguientes variables indique ¿cuáles son discretas y cuáles son continuas?***

    a. Número de acciones vendidas cada día en la Bolsa.

    b. Temperaturas registradas cada hora en un observatorio.

    c. Período de duración de un automóvil.

d. El diámetro de las ruedas de varios coches.

e. Número de hijos de 50 familias.

f. Censo anual de los españoles.

11. Clasifique las variables en cualitativas y cuantitativas, en discretas o continuas.

a. La nacionalidad de una persona.

b. Número de litros de agua contenidos en un depósito.

c. Número de libros en un estante de librería.

d. Suma de puntos tenidos en el lanzamiento de un par de dados.

e. La profesión de una persona.

f. El área de las distintas baldosas de un edificio

12. ¿Cualitativa o cuantitativa? Identifique cada una de las variables como cuantitativa o cualitativa:

a. Tiempo para ensamblar un rompecabezas sencillo
b. Número de estudiantes en un salón de clases de primer año
c. Califi cación de un político recién electo (excelente, bueno, regular, malo)
d. Estado en que vive una persona

13. ¿Discreta o continua? Identifique cada una de las variables cuantitativas como discretas o continuas.

a. Número de accidentes en botes en un tramo de 50 millas del río Colorado
b. Tiempo para completar un cuestionario
c. Costo de una lechuga
d. Número de hermanos y hermanas que tenga el lector
e. Rendimiento en kilogramos de trigo para un terreno de 1 hectárea de un trigal

14. Tiempos de supervivencia al cáncer Un investigador médico desea estimar el tiempo de supervivencia de un paciente, después del inicio de un tipo particular de cáncer y después de un régimen particular de radioterapia.

a. ¿Cuál es la variable de interés para el investigador médico?
b. ¿La variable del inciso a) es cualitativa, cuantitativa, discreta o cuantitativa continua?
c. Identifica que la población de interés para el investigador médico.
d. Describa la forma en que el investigador podría seleccionar una muestra de entre la población.
e. ¿Qué problemas podrían surgir al muestrear desde esta población?

15. ***Presidentes de Estados Unidos Un conjunto de datos contiene las edades al fallecimiento de cada uno de los anteriores 24 presidentes de Ecuador ahora desaparecidos.***
    a. ¿Este conjunto de mediciones es una población o una muestra?
    b. ¿Cuál es la variable que se mide?
    c. ¿La variable del inciso
    b) es cuantitativa o cualitativa?

16. ***Actitudes del electorado Usted es candidato a la legislatura de su estado y desea hacer una encuesta de las actitudes del electorado, respecto a las probabilidades que tenga usted para ganar. Identifique la población que es de interés para usted y de la que le gustaría seleccionar una muestra. ¿En qué forma esta población depende del tiempo?***

CAPITULO 2

Autores:

Dra. June Lady Robles Amaya. MAE
UNIVERSIDAD ESTATAL DE MILAGRO

Msc. Sara Santillan Indacochea
UNIVERSIDAD DE GUAYAQUIL

Ing. Mauricio Guillen Godoy. MBA
UNIVERSIDAD ESTATAL DE MILAGRO

"Algo muy importante para el principiante es comprender que, si los resultados van a ser tratados estadísticamente, es necesario tener en cuenta la estadística al planear el experimento."
WILLIAM IAN BEARDMORE BEVERIDGE

CAPITULO 2

Descripción de los datos: Distribuciones de Frecuencias y Representaciones Gráficas

OBJETIVOS:

Organizar los datos en una distribución de frecuencias. Presentar una distribución de frecuencias en un histograma, un polígono de frecuencias y un polígono de frecuencias acumuladas. Desarrollar una representación de tallo y hoja. Presentar datos mediante técnicas de graficación; como gráficas de líneas, de barras y circulares.

Desarrollo:

En años recientes, las computadoras se han hecho fácilmente accesibles para muchos estudiantes y son una valiosa herramienta. En el estudio de estadísticas, incluso un principiante puede usar paquetes de programas para realizar análisis estadísticos con un alto grado de rapidez y precisión. Algunos de los paquetes estadísticos más comunes que se pueden adquirir en centros de cómputo son el MINITAB, INFOSTAT, SAS (Statistical Analysis System) y el SPSS (Statistical Package for the Social Sciences); las computadoras personales tienen capacidad para paquetes como el R++, MS Excel y otros. Hay incluso programas de estadística en línea y "applets" interactivos en la internet.

En este libro usamos principalmente el MS Excel como herramienta estadística; entender la salida básica de este paquete ayudará al estudiante a interpretar la salida de otros sistemas de cómputo.

Distribución de frecuencias:

Agrupamiento de datos en categorías que muestran el número de observaciones en cada categoría mutuamente excluyente(Fischetti, 2015)**.**

Pasos para organizar los datos en una distribución de frecuencias

1. ***Establezca grupos conocidos como clases.***

- *La práctica común es permitir que el límite inferior de la primera clase sea un poco menor que el valor más pequeño en el grupo de datos, y que el límite superior de la clase mayor sea algo mayor que el valor más grande.*

- *Esto quiere decir que no es necesario que el límite inferior de la primera clase sea igual al valor más pequeño.*

- *Las clases son mutuamente excluyentes, lo cual significa que una observación especifica solamente puede entrar en una sola categoría.*

- ***Fórmula para hallar el Número de Clases:***

$$2^K \geq n$$

Donde:

2 es un valor constante (es decir no cambia)
n es el número de observaciones o datos
K es el número de clases

Ejemplo:

n = 30

$$2^K \geq n$$
$$2^K \geq 30$$
$$2^5 \geq 30$$
$$32 \geq 30$$
$$K = 5$$

Por lo tanto, el número de clases será 5.

2. ***Halle el tamaño del Intervalo de Clases.***

Su Fórmula es la siguiente:

$$i = \frac{\text{Valor Mayor - Valor Menor}}{\text{K}}$$

Ejemplo:

$$i = \frac{90-45}{5}$$

$$i = \frac{45}{5}$$

$$i = 9$$

Entonces el tamaño del intervalo será de 9.

3. ***Distribuya en clases los datos Observados.***

- *Cada clase incluye el límite inferior pero no el límite superior de la clase.*
- *En consecuencia, no se puede sobrepasar, es decir un dato específico solamente aparecerá en una clase.*

4. ***Cuente el número de artículos en cada clase.***

- *El número de observaciones en cada una se llama frecuencias de clase.*
- *Suele ser útil expresar los datos en millares, o en algunas otras unidades convenientes, más que con las cifras reales.*
- *Admitimos que el colocar información sobre precios de venta en una distribución de frecuencia da como resultado la pérdida de alguna información detallada.*
- *Las ventajas de considerar los datos en una forma más comprensible contrarrestan con mucho esta desventaja.*

5. ***Su juicio profesional puede determinar el número de clases.***

- *Demasiadas clases o muy pocas no podrían dar a conocer la forma básica del conjunto de datos*
- *Por regla general, es mejor no utilizar menos de 5 ni más de 15 clases en la elaboración de una distribución de frecuencias.*

6. ***Evite la superposición de límites de clase establecidos.***

- *Las clases establecidas de esta manera no son mutuamente excluyentes, lo que infringe la definición de distribución de frecuencias.*

- *Con la superposición de clases no estaría claro dónde clasificar un valor.*

7. Trate de evitar tener clases abiertas.

- *Este tipo de clases ocasiona problemas al hacer gráficas, y al determinar ciertas medidas de tendencia central y de dispersión.*

EJEMPLO 1:

Suponiendo que los siguientes datos pertenecen a la edad de las personas que se encontraban almorzando en un restaurante.

65	**46**	**49**	**84**	**79**	**56**	**28**	**43**	**67**	**36**
43	**78**	**37**	**40**	**68**	**72**	**55**	**62**	**22**	**82**

Se pide que se elabore una Tabla de Distribución de Frecuencias

Solución:

Como se puede observar existen 20 observaciones es decir n=20. Entonces procedemos a calcular el número de Clases, y el tamaño del intervalo con las fórmulas antes vista.

$$2^K \geq n$$
$$2^K \geq 20$$
$$2^5 \geq 20$$
$$32 \geq 20$$
$$K = 5$$

$$i = \frac{\text{Valor Mayor - Valor Menor}}{\text{K}}$$

$$i = \frac{84-22}{5}$$
$$i = \frac{62}{5}$$
$$i = 12,4$$
$$i \cong 12$$

Clases	**Frecuencias**
[20; 32]	**2**
[32; 44]	**5**
[44; 56]	**3**
[56; 68]	**4**
[68; 80]	**4**

[80; 92]	**2**
	20

Como podemos observar en las clase de [20;32] hay 2 personas
Como podemos observar en las clase de [32;44] hay 5 personas
Como podemos observar en las clase de [44;56] hay 3 personas
Como podemos observar en las clase de [56;68] hay 4 personas
Como podemos observar en las clase de [68;80] hay 4 personas
Como podemos observar en las clase de [80;92] hay 4 personas

Noten que en la columna de las frecuencias la suma total de los datos debe ser igual al número de observaciones, en este caso 20.

Marca de Clase (punto medio):

Es el punto que divide a la clase en dos partes iguales. Es el promedio entre los límites superior e inferior de la clase.

Ejemplo:

Si la clase fuese [20;32], su Marca de clase o punto medio seria:

$$= \frac{20+32}{2}$$
$$= \frac{52}{2}$$
$$= 26$$

Para el Ejemplo 1 podemos obtener lo siguiente:

Clases	Frecuencias	Marca de Clase
[20; 32]	2	26
[32; 44]	5	38
[44; 56]	3	50
[56; 68]	4	62
[68; 80]	4	74
[80; 92]	2	86
	20	

Intervalo de clase:

Para una distribución de frecuencias que tiene clases del mismo tamaño, el intervalo de clase se obtiene restando el límite inferior de una clase del límite inferior de la siguiente.

Distribución de Frecuencia Relativas

Puede resultar conveniente convertir las frecuencias de clase a frecuencias de clase relativas para mostrar el porcentaje del número total de observaciones en cada clase.

- *Para convertir una distribución de frecuencias a una distribución de frecuencias relativas, cada frecuencia de clase se divide entre el número total de observaciones.*

Ejemplo 1:

Clases	Frecuencias	Marca de Clase	Frec. Relativa
[20; 32]	2	26	2/20 = 0,1
[32; 44]	5	38	5/20 = 0,25
[44; 56]	3	50	3/20 = 0,15
[56; 68]	4	62	4/20 = 0,2
[68; 80]	4	74	4/20 = 0,2
[80; 92]	2	86	2/20 = 0,1
	20		1

Distribución de Frecuencias Acumulada

Se usa para determinar cuántos o qué proporción de los valores de los datos es menor o mayor que cierto valor.

Ejemplo 1:

Clases	Frecuencias	Marca de Clase	Frec. Relativa	Frec. Acumulada
[20; 32]	2	26	2/20 = 0,1	2
[32; 44]	5	38	5/20 = 0,25	7
[44; 56]	3	50	3/20 = 0,15	10
[56; 68]	4	62	4/20 = 0,2	14
[68; 80]	4	74	4/20 = 0,2	18
[80; 92]	2	86	2/20 = 0,1	20
	20		1	

Y, aquí tenemos la tabla de Distribución de frecuencias Terminada.

Presentación Gráfica de una Distribución de Frecuencias

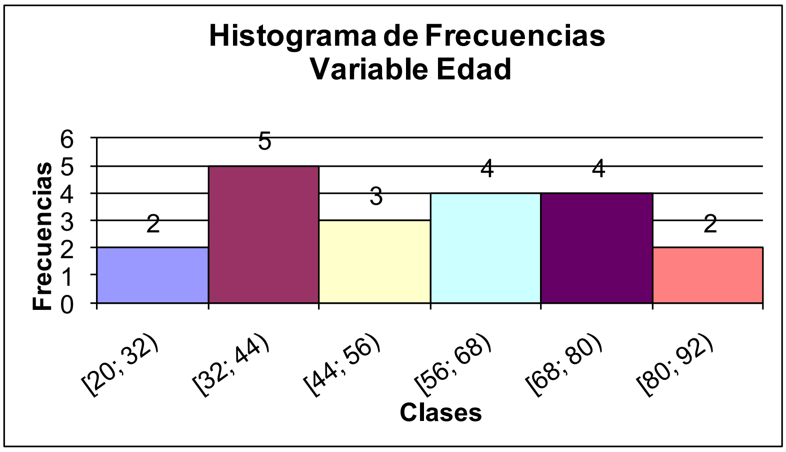
Las tres formas de gráficas más usadas son histogramas, polígonos de frecuencia y distribuciones de frecuencias acumuladas (ojiva).

Histograma de Frecuencias:

Gráfica donde las clases se marcan en el eje horizontal (Eje de las "X"), y las frecuencias de clase en el eje vertical (Eje de las "Y"). Las frecuencias de clase se representan por las alturas de las barras y éstas se trazan adyacentes entre sí.

De manera que un histograma describe una distribución de frecuencias utilizando una serie de rectángulos adyacentes, en donde la altura de cada rectángulo es proporcional a la frecuencia que la clase representa.

Ejemplo1:



Comentario:

Como podemos observar en el intervalo de [20;32] hay 2 personas.
Como podemos observar en el intervalo de [32;44] hay 5 personas, y es aquí donde hubo más personas.
Como podemos observar en el intervalo de [44;56] hay 3 personas.
Como podemos observar en el intervalo de [56;68] hay 4 personas.
Como podemos observar en el intervalo de [68;80] hay 4 personas.
Como podemos observar en el intervalo de [80;92] hay 4 personas.

Polígono de Frecuencias:

Consiste en segmentos de línea que conectan los puntos formados por el punto medio de la clase y la frecuencia de clase.

El punto medio de cada clase está marcado en el eje X, y las frecuencias de cada clase en el eje Y.

Recuerde que el punto medio de clase es el valor que se encuentra al centro de una clase y representa los valores en esa.

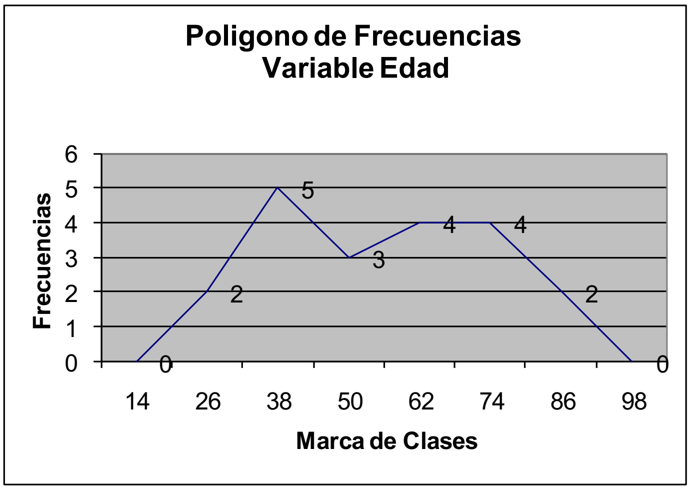
La frecuencia de clase es el número de observaciones en una clase específica.

Los puntos se unen en orden.

Para completar el polígono de frecuencias, se agregan puntos medios al eje X para "anclar" el polígono de frecuencias.

Estos dos valores, se obtienen al restar el intervalo de clase al punto medio más bajo, y al sumar el intervalo de clase al punto medio más alto en la distribución de frecuencias.

Ejemplo 1:

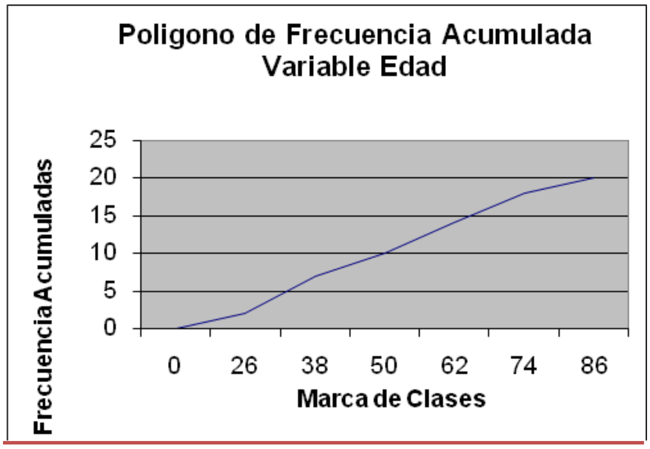


Polígono de Frecuencia Acumulado (OGIVA)

Para graficar una distribución de frecuencias acumuladas, localice el límite superior de la clase en el eje X y las frecuencias acumuladas a lo largo del eje Y.

Para proporcionar información adicional, puede graduarse el eje vertical a la izquierda en unidades, y el eje vertical a la derecha en porcentajes.

Ejemplo 1:

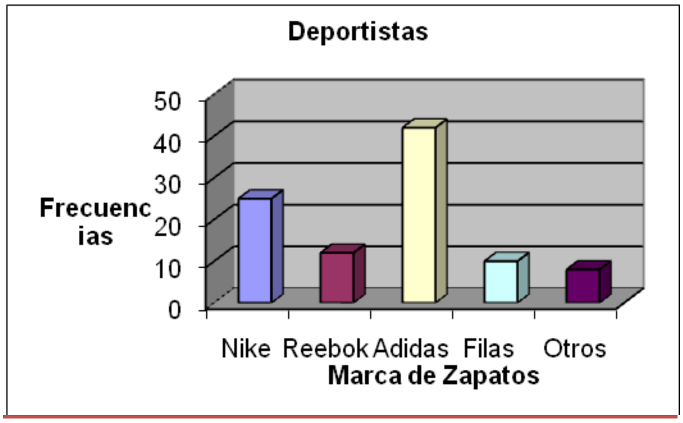


Gráfica de Barras:

Se puede usar para describir cualquier nivel de medición (nominal, ordinal, de intervalo o de razón).

EJEMPLO 1:

Construya una gráfica de barras para el número de deportistas que usan ciertas marcas de zapatos para las competencias de atletismo.



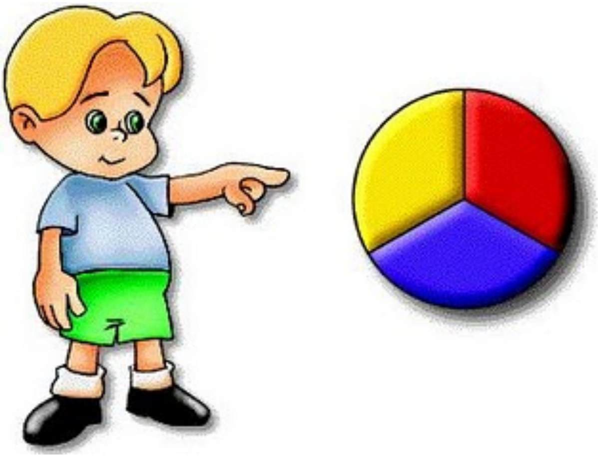
- ***Observe que hay un espacio entre las barras que representan los diversos grupos. Esta es una de las formas en las que difieren un histograma de frecuencias y una gráfica de barras.***

- ***No hay espacio entre las barras de un histograma de frecuencia, porque los son de intervalo o de nivel de razón, pero este no es el caso de los datos en una gráfica de barras.***

Gráfica circular o pastel:

Es en especial útil para desplegar una distribución de frecuencias relativas. Se divide un círculo de manera proporcional a la frecuencia relativa y las rebanadas representan los diferentes grupos.

El primer paso es registrar los porcentajes.



EJEMPLO 1:

Basándonos en el mismo ejemplo anterior el de los deportistas, elaboremos un gráfico de Pastel.

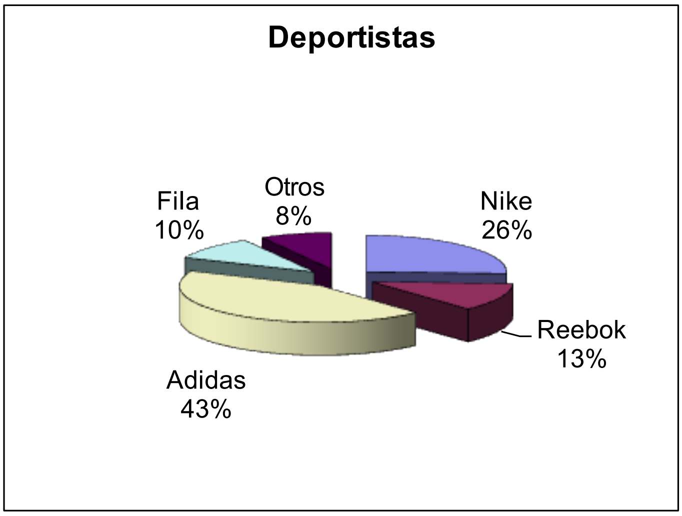


Diagrama de Tallo y Hojas:

Es una técnica estadística para representar un conjunto de datos. Cada valor numérico se divide en dos partes: los dígitos principales son el tallo y el dígito siguiente es la hoja.

EJEMPLO 1:

26

Donde: **2** *Es el Digito principal y por lo tanto será el TALLO*
6 *Es el Digito secundario y por lo tanto será la HOJA*

EJEMPLO 2:

115

Donde: **11** *Es el Digito principal y por lo tanto será el TALLO*

5 *Es el Digito secundario y por lo tanto será la HOJA*

- ***NOTA:*** *una ventaja de la representación de tallo y hoja, comparado con la distribución de frecuencias es que no se pierde la identidad de cada observación.*

  La principal ventaja al organizar los datos en una distribución de frecuencias es que se consigue un cuadro visual rápido de la forma de la distribución sin hacer cálculos adicionales.

  Una representación de tallo y hoja realmente es un histograma con más información; es decir, valores de datos en vez de grupos.

  EJEMPLO 3:

  Peter logró las siguientes calificaciones en el doceavo examen de contabilidad del semestre: 86, 79, 92, 84, 69, 88, 91, 83, 96, 78, 82, 85. Construya una representación de tallo y hoja para los datos.

  Solución:

  1. Tenemos que ordenar los Datos

  69, 78, 79, 82, 83, 84, 85, 86, 88, 91, 92, 96

  2. Separe los Tallos de las hojas.

TALLO	HOJAS
6	9
7	8 , 9
8	2, 3, 4, 5, 6, 8
9	1, 2, 6

Note, que se puede observar un forma de histograma en el Diagrama de tallo y hoja, como podemos observar en este caso no se pierde la identidad de los datos es decir sabemos con exactitud qué valores forman parte de este Diagrama de tallo y hoja.

Podemos ver también que en el intervalo de [82; 88], es donde se encuentran la mayor cantidad de datos o información. Y que en el tallo sea solo tenemos una hoja, que juntos conforman la calificación 69. etc.

AUTOEVALUACIÓN # 2

REALICE LOS SIGUIENTES EJERCICIOS:

1. ***Para determinar la cuantía que la beca para libros otorgará el Ministerio de Educación el curso que viene, se desea tener una idea de lo que realmente invierten los alumnos en libros, por cuatrimestre. Para ello, se han conseguido los datos de 50 alumnos recogidos en la siguiente tabla, x100 pts. Aproximadas al precio "redondo" más próximo:***

125	**157**	**145**
165	**191**	**148**
148	**197**	**225**
136	**205**	**119**
179	**217**	**221**
205	**148**	**222**
209	**152**	**165**
235	**113**	**148**
265	**119**	**127**
245	**117**	**148**

Elabore una Tabla de Distribución de Frecuencias Por conveniencia agrupar los datos utilizando: Fórmula para hallar el Número de Clases, Halle el tamaño del Intervalo de Clases, Marca de clase (punto medio), Intervalo de clase, Distribución de frecuencia Relativas, Distribución de Frecuencias Acumulada,

2. ***En un experimento que medía el porcentaje de encogimiento al secar 50 muestras de arcilla plástica, produjeron los siguientes resultados:***

19	20	17	17	17	15	16	17	19	22
20	18	22	19	17	18	18	18	17	17
14	12	19	16	19	17	19	17	16	18
21	23	18	19	19	16	18	17	18	17
18	18	16	17	17	20	19	17	18	22

a) Construir el gráfico las frecuencias,
b) Hallar las frecuencias relativas para una tabla de frecuencias de 5 intervalos.
c) Construye el gráfico de la frecuencia acumulada.
d) Calcular las medidas estadísticas para dicha tabla. Comentar.

3. ***Una empresa grande de equipos deportivos está probando el efecto de dos planes publicitarios sobre las ventas de los últimos 4 meses. Dadas las ventas que se ven aquí, ¿cuál programa de publicidad parece producir el crecimiento promedio más alto en ventas mensuales?***

Mes	**Plan 1**	**Plan 2**
Enero	1657,0	4735,0
Febrero	1998,0	5012,0
Marzo	2267,0	5479,0
Abril	3432,0	5589,0

4. ***Los estadísticos del programa de Meals on Wheels (comida sobre ruedas), el cual lleva comidas calientes a enfermos confinados en casa, desean evaluar sus servicios. El número de comidas diarias que suministran aparece en la siguiente tabla de frecuencia. Calcular la media, mediana y la moda e intérprete.***

Número de comidas por día	Número de días
0 - 5	3
5 - 10	6
10 - 15	5
15 - 20	8
20 - 25	2
25 - 30	3

5. ***MOUTS CRADS compró 20 acciones a $ 15 cada una, 50 acciones a $20 cada una, 100 acciones a $30 cada una y 75 acciones a $35 cada una. ¿Cuál es el precio promedio por acción?***

6. ***Las edades de 50 de los directores ejecutivos de las mejores corporaciones de la nación reportadas aparecen en la siguiente tabla de frecuencias. Calcule e interprete la media, la mediana y la moda. Además, calcule e interprete: Q1, Q2, D10, D60, P15, P90.***

Edades	Frecuencias
50 y menos de 55	8
55 y menos de 60	13
60 y menos de 65	15
65 y menos de 70	10
70 y menos de 75	3
75 y menos de 80	1

CAPITULO 3

Autores:

Ing. Javier Alcazar Espinoza. Mgs
UNIVERSIDAD ESTATAL DE MILAGRO

Ing. Paúl Freire Sierra, Mgs.
UNIVERSIDAD DE GUAYAQUIL

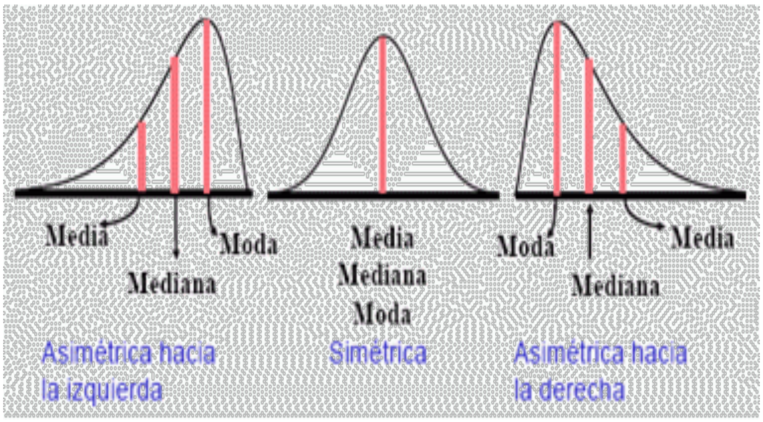
Lcda. Vanessa Vargas Párraga. Mgs
UNIVERSIDAD UNIR

"La falacia del cuadro estadístico estriba en que es unilateral, en la medida en que representa sólo el aspecto promedio de la realidad y excluye el cuadro total. La concepción estadística del mundo es una mera abstracción, y es incluso falaz, en particular cuando atañe a la psicología del hombre."

CARL GUSTAV JUNG

CAPITULO 3

Medidas de Ubicación o de Tendencia Central



OBJETIVOS:

Calcular la media aritmética, mediana, moda, media ponderada y media geométrica. Explicar las características, utilización, ventajas y desventajas de cada medida de ubicación. Identificar la posición de la media aritmética, la mediana, y la moda, tanto para distribuciones simétricas como asimétricas o sesgadas.

Desarrollo:

Medidas de Tendencia Central para Datos Desagrupados

Media aritmética de la población

Para datos no agrupados, la media de la población es la suma de todos los valores en ella dividida entre el total de valores en la población(Bruce & Bruce, 2017)*:*

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

Donde:

- ***μ*** *Representa la media de la población.*
- ***N*** *Es el número total de elementos en la población.*
- ***X*** *Representa cualquier valor en particular.*
- ***Σ*** *Indica la operación de sumar.*

Nota: *Esta fórmula se la utiliza cuando los datos que se utilizan son todos aquellos que forman una población.*

EJEMPLO1:

Se tiene 15 estudiantes en un curso de CÁLCULO y se desea saber el promedio de Edad de los estudiantes que están en dicha clase. Las edades son las siguientes:

18, 29, 22, 24, 20, 23, 25, 19, 18, 17, 22,21, 20, 18, 19

$$\mu = \frac{\sum_{i=1}^{15} X_i}{15} = \frac{18+29+22+24+20+23+25+19+18+17+22+21+20+18+19}{15}$$

$$\mu = \frac{315}{15} = 21$$

Entonces podemos decir que el promedio de edad de los estudiantes en dicho curso es de 21 años.

Parámetro:

Es una característica de una población.

Media Aritmética de la Muestra

Para datos no agrupados, la media de la Muestra es la suma de todos los valores en ella dividida entre el total de valores en la Muestra:

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

Donde:

- $\overline{X}$ *Representa la media de la población.*
- ***n*** *Es el número total de elementos en la población.*
- ***X*** *Representa cualquier valor en particular.*
- ***Σ*** *Indica la operación de sumar.*

Nota: *Esta Fórmula se la utiliza cuando los datos que se utilizan son todos aquellos que forman una Muestra.*

EJEMPLO 1:

Si del ejemplo anterior el de los estudiantes de un curso de inglés, tomamos una muestra de 5 estudiantes para tener una idea de la edad promedio de los estudiantes de dicho curso. Cuya muestra aleatoria está compuesta de los siguientes datos:

18, 25, 20, 19, 21

$$\overline{X} = \frac{\sum_{i=1}^{5} X_i}{5} = \frac{18+25+20+19+21}{5} = \frac{103}{5} = 20,6$$

$$\overline{X} \cong 21$$

Entonces podemos decir que el promedio de edad de los estudiantes en dicho curso, basado en una muestra de dicho curso es de 21 años.

Dato estadístico:
Es una característica de una muestra*.*

Propiedades de la Media Aritmética

- *Todo conjunto de datos de nivel de intervalo y de nivel de razón tiene un valor medio.*
- *Al evaluar la media se incluyen todos los valores.*
- *Un conjunto de valores sólo tiene una media.*
- *La cantidad de datos a evaluar rara vez afecta la media.*
- *La media es la única medida de ubicación donde la suma de las desviaciones de cada valor con respecto a la media, siempre es cero.*

$$\sum_{i=1}^{n}(X_i - \overline{X}) = 0$$

EJEMPLO 1:

Basándonos en la muestra anterior de los estudiantes del curso de inglés. Probar si esta propiedad es cierta.

$$\sum_{i=1}^{n}(X_i - \overline{X}) = 0$$

$$(18-20,6)+(25-20,6)+(20-20,6)+(19-20,6)+(21-20,6)=0$$

$$-2,6+4,4-0,6-1,6+0,4=0$$

Por lo tanto queda comprobada dicha propiedad.

Mediana:

Es el punto medio de los valores después de ordenarlos de menor a mayor, o de mayor a menor. La misma cantidad de valores se encuentra por arriba de la mediana que por debajo de ella (Brown, 2018)**.**

Cálculo de la Mediana para una Población o Muestra PAR:

Para un conjunto con un número par de números, la mediana se calculará de la siguiente forma:

$$\tilde{X} = \frac{X'_{n/2} + X'_{(n/2)+1}}{2}$$

EJEMPLO 1:

Siguiendo con los del Ejemplo1, para poder mantener los mismos datos:

18, 29, 22, 24, 20, 23, 25, 19, 18, 17, 22,21, 20, 18, 19, 30

Solución:

1) Lo primero que debemos hacer es ordenar los datos:

17, 18, 18, 18, 19, 19, 20, 20, 21, 22, 22, 23, 24, 25, 29, 30

2) Ahora debemos ver si la población o la muestra es Par o Impar, como podemos observar la población es PAR. Entonces utilizamos la Fórmula cuando es PAR.

$$\tilde{X} = \frac{X'_{16/2} + X'_{(16/2)+1}}{2} = \frac{X'_8 + X'_{8+1}}{2} = \frac{X'_8 + X'_9}{2}$$

Ahora debemos buscar en el conjunto de datos ordenados, la posición 8 y 9 de los datos.

$$\tilde{X} = \frac{20+21}{2} = \frac{41}{2} = 20,5$$

Como podemos ver la mediana es 20,5.

Cálculo de la Mediana para una Población o Muestra IMPAR:

Para un conjunto con un número impar de números, la mediana se calculara de la siguiente forma:

$$\tilde{X} = X'_{\frac{n+1}{2}}$$

EJEMPLO 1:

18, 29, 22, 24, 20, 23, 25, 19, 18, 17, 22,21, 20, 18, 19

Solución:

1) Lo primero que debemos hacer es ordenar los datos:

17, 18, 18, 18, 19, 19, 20, 20, 21, 22, 22, 23, 24, 25, 29

2) Ahora debemos ver si la población o la muestra es Par o Impar, como podemos observar la población es PAR. Entonces utilizamos la Fórmula cuando es IMPAR.

$$\tilde{X} = X'_{\frac{n+1}{2}} = X'_{\frac{15+1}{2}} = X'_{16/2} = X'_{8}$$

Ahora debemos buscar en el conjunto de datos ordenados, la posición 8 de los datos y ese valor es la Mediana.

$$\tilde{X} = 20$$

Como podemos ver la Mediana es 20.

Propiedades de la Mediana

- *La mediana es única para cada conjunto de datos.*
- *No se ve afectada por valores muy grandes o muy pequeños, y por lo tanto es una medida valiosa de tendencia central cuando ocurren.*
- *Puede obtenerse para datos de nivel de razón, de intervalo y ordinal.*
- *Puede calcularse para una distribución de frecuencias con una clase de extremo abierto, si la mediana no se encuentra en una de estas clases.*

La Moda $\hat{X}$ ::

- ***Es el valor que más veces se repite en un conjunto de Datos.***

EJEMPLO 1:

Siguiendo con los datos del ejemplo anterior:

17, 18, 18, 18, 19, 19, 20, 20, 21, 22, 22, 23, 24, 25, 29

$$\hat{X} = 18$$

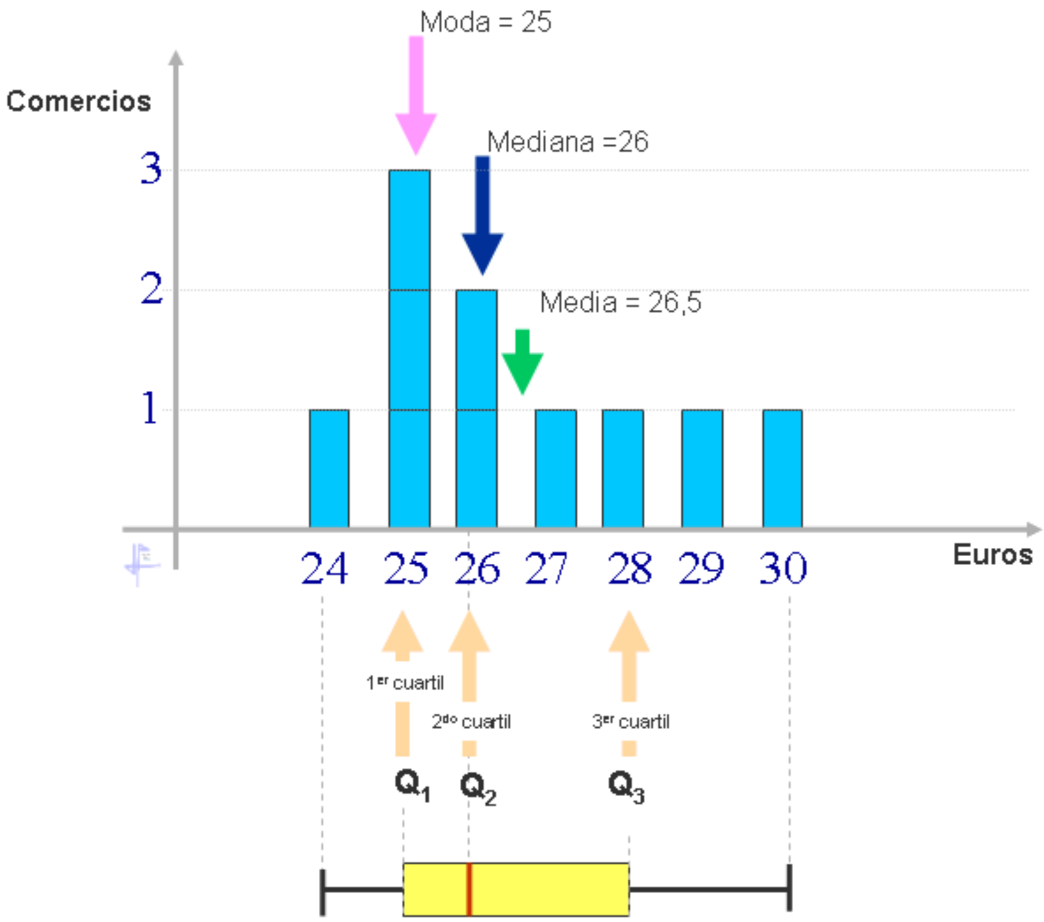
Como el valor que más veces se repite es el 18 está es la moda.

Medidas de Tendencia Central Para datos Agrupados

Media de Datos Agrupados

La media de una muestra de datos organizados en una distribución de frecuencias se calcula mediante la siguiente fórmula:

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



EJEMPLO 1:

Los siguientes datos representan las ventas de un departamento en miles de dólares.

Clases	***Frecuencias(f)***
[20,30]	***6***
[30,40]	***7***
[40,50]	***4***
[50,60]	***8***
[60,70]	***5***
	30

Solución:

Tenemos que calcular la Marca de Clases (X), y luego multiplicamos la Columna de la frecuencia por la Columna de la Marca de clases. A continuación se muestra como quedaría el cuadro:

A continuación calcularemos la Media Aritmética.

$$\bar{X} = \frac{\sum_{i=1}^{30} X_i * f_i}{30} = \frac{1340}{30} = 44,67$$

Como podemos ver la Media es de 44,67 (en miles de dólares), en Ventas de dicho departamento.

Mediana de Datos Agrupados

La mediana de una muestra de datos organizados en una distribución de frecuencias se calcula mediante la siguiente fórmula:

$$X = L + \left(\frac{n/2 - FA}{f} \right) * i$$

Donde:

L *Es el límite inferior de la clase que contiene a la Mediana*
n *Es el tamaño de la Muestra o Población*
FA *Es la Frecuencia Acumulada de la clase que contiene a la Mediana*
f *Es la Frecuencia de la clase que contiene a la Mediana*
i *Es el tamaño del intervalo.*

EJEMPLO:

Siguiendo con los datos de la tabla anterior tenemos que:

Solución:

Clases	*Frecuencias(f)*	*Marca de Clases (X)*	*f * X*	*Frecuencia Acumulada (FA)*
[20,30]	*6*	*25*	*150*	*6*
[30,40]	*7*	*35*	*245*	*13*
[40,50]	*4*	*45*	*180*	*17*
[50,60]	*8*	*55*	*440*	*25*
[60,70]	*5*	*65*	*325*	*30*
	30		*1340*	

Para determinar la clase de la mediana de datos agrupados:

- ***Elabore una distribución de frecuencias acumulada.***
- ***Divida el número total de datos entre 2. Determine qué clase contiene este valor. Por ejemplo, si n=30, 30/2 = 15, después determine qué clase contiene el 15° valor (la clase de la mediana). El mismo que se encuentra en la frecuencia Acumulada donde la suma llega a 17, ya que aquí esta contenido la posición 15 de los datos ordenados.***

Clases	*Frecuencias(f)*	*Marca de Clases (X)*	*f * X*	*Frecuencia Acumulada (FA)*
[20,30)	*6*	*25*	*150*	*6*
[30,40)	*7*	*35*	*245*	*13*
[40,50)	*4*	*45*	*180*	*17*
[50,60)	*8*	*55*	*440*	*25*
[60,70)	*5*	*65*	*325*	*30*
	30		*1340*	

$$X = L + \left(\frac{n/2 - FA}{f} \right) * i = 40 + \left(\frac{30/2 - 17}{4} \right) * 10 = 40 + \left(\frac{15 - 17}{4} \right) * 10$$

$$X = 40 + \left(\frac{-2}{4} \right) * 10 = 40 + (-0.5) * 10 = 40 - 5 = 35$$

$$X = 35$$

Por lo tanto la Mediana es de 35.

Moda de Datos Agrupados $\hat{X}$

La moda de los datos agrupados se aproxima a la Marca de clase que contiene la frecuencia de mayor.

EJEMPLO:

Siguiendo con la tabla anterior.

Clases	Frecuencias (f)	Marca de Clases (X)	$f * X$	Frecuencia Acumulada (FA)
[20,30)	6	25	150	6
[30,40)	7	35	245	13
[40,50)	4	45	180	17
[50,60)	8	55	440	25
[60,70)	5	65	325	30
	30		1340	

Como podemos ver la frecuencia Mayor es 8, y por lo tanto la moda $\hat{X} = 55$ ***ya que la Marca de clase es 55.***

La Media Ponderada:

La media ponderada de un conjunto de números X1, X2, ..., Xn, con las ponderaciones correspondientes w1, w2, ...,wn, se calcula con la fórmula:

$$\bar{X}w = \frac{\sum_{i=1}^{n} w_i * X_i}{\sum_{i=1}^{n} w_i} = \frac{w_1 X_1 + w_2 X_2 + \ldots\ldots + w_n X_n}{W_1 + w_2 + \ldots\ldots + w_n}$$

EJEMPLO:

Durante un periodo de una hora en una tarde calurosa de un sábado, el cantinero José sirvió cincuenta bebidas. Calcule la media ponderada de los

precios de las bebidas. (Precio ($), cantidad vendida): (.50,5), (.75,15), (.90,15), (1.10,15).

Solución:

$$\bar{X}w = \frac{\sum_{i=1}^{4} w_i * X_i}{\sum_{i=1}^{4} w_i} = \frac{5*(0.50)+15*(0.75)+15*(0.90)+15*(1.10)}{50} = 0.875$$

Entonces podemos decir que la media ponderada de los precios de las bebidas fue de 0.875.

Media Geométrica:

La media geométrica (MG) de un conjunto de n números positivos se define como la raíz n-ésima del producto de los n valores. Su fórmula es:

$$MG = \sqrt[n]{(X_1)*(X_2)*\ldots\ldots*(X_n)}$$

La media geométrica se usa para encontrar el promedio de porcentajes, razones, índices o tasas de crecimiento.

EJEMPLO:

Las tasas de interés de tres bonos son 5%, 7% y 4%.

$$MG = \sqrt[3]{(5)*(7)*(4)} = 5.192$$

Entonces la Media Geométrica de los Bonos es de 5.192%.

Otra aplicación de la media geométrica es determinar el porcentaje promedio del incremento en ventas, producción u otros negocios o series económicas de un periodo a otro. La fórmula para este tipo de problema es:

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

EJEMPLO:

El número total de mujeres inscritas en colegios americanos aumentó de 755000 en 1986 a 835 000 en 1995.

Solución:

Aquí n = 10.

$$MG = \sqrt[10]{\frac{(835000)}{735000}} = 0.0658$$

Es decir, la media geométrica de la tasa de crecimiento es 6.58%.

AUTOEVALUACIÓN # 3

REALICE LOS SIGUIENTES EJERCICIOS:

1. ***Las calificaciones finales en la asignatura de Matemáticas de 1º de la ESO de 50 alumnos son las siguientes:***

   5 – 6 – 5 – 4 – 3 – 2 – 7 – 9 – 5 – 8 – 6 – 5 – 4 – 7 – 9 – 1 – 7 – 9 – 3 – 5 – 6 – 6 – 8 4 – 3 – 2 – 5 – 3 – 7 – 3 – 2 – 5 – 4 – 9 – 9 – 7 – 5 – 4 – 6 – 3 – 3 – 7 – 3 – 8 – 4 – 6 5 – 3 – 9 – 2.

   Considerando que la variable X = "calificación de Mate" toma valores discretos:

   a) ***Calcula la tabla de frecuencias.***
   b) ***Dibuja el diagrama de barras y de frecuencias acumuladas.***
   c) ***Dibuja un diagrama de sectores con los suspensos, aprobados, notables y sobresalientes.***
   d) ***Calcula la media, la mediana y la moda.***
   e) ***Calcula la media geométrica y armónica.***
   f) ***Calcula las desviaciones medias y medianas.***
   g) ***Calcula el momento de orden tres respecto del origen y respecto de la media***

2. ***Los datos observados en un estudio sobre el tamaño de los huevos de cuco dieron los siguientes resultados:***

Anchura (mm.)	*Número de huevos*
14.75 – 15.25	*5*
15.25 – 15.75	*9*
15.75 – 16.25	*73*
16.25 – 16.75	*51*
16.75 – 17.25	*80*
17.25 – 17.75	*15*
17.75 – 18.25	*7*

Se pide:

a) ***Tabla de frecuencias***
b) ***Histograma y diagrama de frecuencias acumuladas***
c) ***Media, mediana y moda***

3. ***Los siguientes datos numéricos corresponden a la cantidad de veces que cada alumno de un grupo ha ido a un recital o concierto.***

   2 – 4 – 3 – 2 – 1 – 1 – 6 – 3 – 0 – 3 – 2 – 4 – 6 – 9 – 3 – 2 – 1 – 6

   Calcula, sin tabular, Media, moda, mediana, desviación, n, rango.

4. ***En un diagnóstico de educación física se pidió a los alumnos de los cuartos medios, que hicieran abdominales durante 3 minutos. Se obtuvieron los siguientes resultados:***

   4º A: 45 38 43 29 34 60 54 27 32 33 23 34 34 28 56 62 56 57 45 47 48 54 33 45 44 41 34 36 34 54

   4º B: 43 45 44 38 34 46 43 42 43 45 57 44 38 38 37 43 61 38 37 45 28 42 49 40 37 34 44 41 43

   ¿Cuál de los dos cursos tiene el rendimiento más parejo? ¿Qué distribución estadístico permite comparar la distribución de este tipo de datos?

5. ***A continuación se presentan los resultados de ambos cursos en la prueba de diagnóstico de salto largo.***

   4º A : 3.2 3.5 4.9 5.0 3.1 4.1 2.9 2.8 3.8 4.5 4.3 4.5 4.1 5.8 3.9 3.6 4.2 4.6 1.9 2.8 2.9 3.3 3.9 4.2 4.1 4.3 4.6 4.4 3.8 3.6

   4º B : 3.5 2.9 1.3 1.7 3.6 5.6 2.8 5.2 5.3 4.1 4.1 4.4 1.6 5.1 4.3 5.0 5.3 3.2 2.8 2.6 5.5 5.4 4.8 4.9 4.3 2.9 3.9 5.4 5.3 4.2

   a) Calcula el promedio de ambos cursos.
   b) Construye una tabla de frecuencias para cada curso
   c) ¿Cuál de los dos cursos tuvo un rendimiento más parejo?

6. ***Se han medido 75 alumnos, en centímetros, obteniéndose los siguientes datos:***

   175 156 172 159 161 185 186 192 179 163 164 170 164 167 168 174 172 168 176 166 167 169 182 170 169 167 170 162 172 171 174 171 155 171 171 170 157 170 173 173 174 168 166 172 172 158 159 163 163 168 174 175 150 154 175 160 175 177 178 180 169 165 180 166 184 183 174 173 162 185 189 169 173 171 173

   Agrupa estos resultados en 8 intervalos y confecciona una tabla de frecuencias y calcula las medidas de tendencia central y de dispersión. Además, grafica esta tabla.

7. ***A los mismos alumnos anteriores se les aplico una prueba de inteligencia, estos han sido:***

   171 170 157 170 173 173 174 168 166 172 172 158 159 163 163 168 174 175 150 154 175 160 175 177 178 180 169 165 180 166 184 183 174 173 162 185 189 169 173 171 173

   Agrupa los datos en intervalos de amplitud 8, y haz lo mismo que en problema anterior.

CAPITULO 4

Autores:

Ing. Roberto Basurto Q. MAE
UNIVERSIDAD ESTATAL DE MILAGRO

Ing. Glenda Blanc. MAE
UNIVERSIDAD ECOTEC

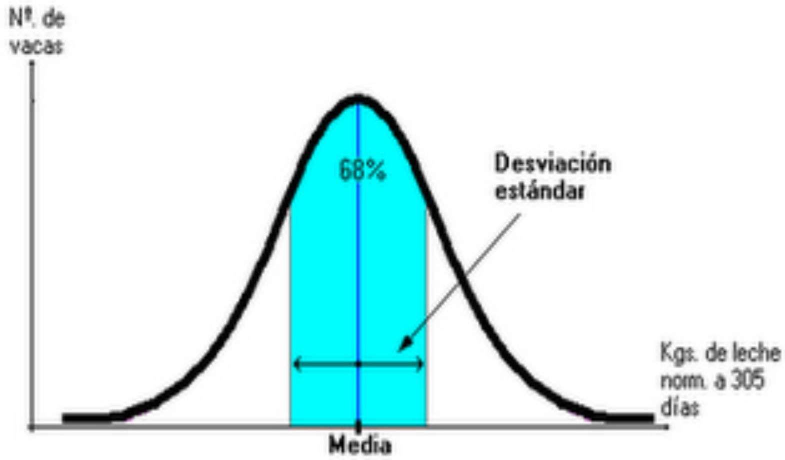
Ing. Virginia Zambrano Zambrano, MBA
UNIVERSIDAD DE GUAYAQUIL

El beso provoca tales palpitaciones que el corazón trabaja en 4 segundos más que en 3 minutos. Las estadísticas demuestran que 480 besos acortan la vida en un día, que 2.360 besos os privan de una semana y que 148. 071 besos, son sencillamente un año perdido. Paul Morand

CAPITULO 4

Medidas de Dispersión

OBJETIVOS:



Calcular e interpretar la amplitud de variación, la desviación media, la variancia, la desviación estándar de los datos originales. Calcular e interpretar la amplitud de variación, la variancia y la desviación estándar de datos agrupados. Explicar las características, usos, ventajas y desventajas de cada medida de dispersión. Calcular y explicar los cuartiles y la amplitud de variación intercuartílica. Elaborar e interpretar los diagramas de caja. Calcular y entender el coeficiente de variación y el coeficiente de asimetría.

Desarrollo:

Medidas de Dispersión Para datos Desagrupados

Varianza Poblacional:

Para datos no agrupados es la media aritmética de las desviaciones cuadráticas respecto a la media de la población(Crawley, 2014)**.**

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

Donde:

μ	***Es la Media de la Población.***
Xi	***Es cada valor en la población***
N	***Es el tamaño de la población***
$\sum$	***La suma de todos los datos***

EJEMPLO:

Para seguir la secuencia del ejemplo anterior usaremos los datos del ejemplo 1 del capítulo 3.

18, 29, 22, 24, 20, 23, 25, 19, 18, 17, 22,21, 20, 18, 19

$$\mu = \frac{\sum_{i=1}^{15} X_i}{15} = \frac{18+29+22+24+20+23+25+19+18+17+22+21+20+18+19}{15}$$

$$\mu = \frac{315}{15} = 21$$

$$\sigma^2 = \frac{\sum_{i=1}^{15}\left(X_i - 21\right)^2}{15} =$$

$$\sigma^2 = \frac{\begin{matrix}(18-21)^2+(29-21)^2+(22-21)^2+(24-21)^2+(20-21)^2+(23-21)^2+(25-21)^2+(19-21)^2+\\ +(18-21)^2+(17-21)^2+(22-21)^2+(21-21)^2+(20-21)^2+(18-21)^2+(19-21)^2\end{matrix}}{15}$$

$$\sigma^2 = \frac{\begin{matrix}(-3)^2+(8)^2+(1)^2+(3)^2+(-1)^2+(2)^2+(4)^2+(-2)^2+(-3)^2+\\ +(-4)^2+(1)^2+(0)+(-1)^2+(-3)^2+(-2)^2\end{matrix}}{15}$$

$$\sigma^2 = \frac{9+64+1+9+1+4+16+4+9+16+1+0+1+9+4}{15}$$

$$\sigma^2 = \frac{148}{15} = 9,87$$

Como se puede observar la Varianza es igual a 9,87, lo que quiere decir que existe una moderada Variabilidad en los datos.

Desviación Estándar de la Población.

La Desviación Estándar Poblacional (σ) es la raíz cuadrada de la variancia de la población. Su Fórmula es la siguiente(Mendenhall, Beaver, Beaver, & Velázquez Arellano, 2017)***:***

$$\sigma = \sqrt[2]{\sigma^2}$$

EJEMPLO:

Basándonos en el ejemplo anterior calcule la Desviación Estándar de la Población.

$$\sigma = \sqrt[2]{\sigma^2} = \sqrt[2]{9,87} = 3,1416$$

Por lo tanto la Desviación Estándar de la Población es igual a 3,1416.

Varianza de la Muestra S^2

La Varianza Muestral estima la variancia de la población.

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

Donde:

S^2	***La Varianza de la Muestra***
Xi	***Cada dato de la Muestra***
n	***El tamaño de la Muestra***
$\bar{X}$	***La Medía de la Muestra.***

EJEMPLO:

Si del ejemplo anterior el de los estudiantes de un curso de inglés, tomamos una muestra de 5 estudiantes para tener una idea de la edad promedio de los estudiantes de dicho curso. Calcule la Varianza de la Muestra, cuya muestra aleatoria está compuesta de los siguientes datos:

18, 25, 20, 19, 21

$$\overline{X} = \frac{\sum_{i=1}^{5} X_i}{5} = \frac{18+25+20+19+21}{5} = \frac{103}{5} = 20,6$$

$$\overline{X} \cong 21$$

$$S^2 = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{n-1} = \frac{\sum_{i=1}^{5}\left(X_i - 21\right)^2}{4} =$$

$$S^2 = \frac{\left(18-21\right)^2 + \left(25-21\right)^2 + (20-21)^2 + (19-21)^2 + (21-21)^2}{4}$$

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

$$S^2 = \frac{9+16+1+4+0}{4} = \frac{30}{4} = 7,5$$

Como podemos observar la Varianza de la Muestra es de 7,5; lo que indica que existe una moderada variabilidad de los datos.

Desviación Estándar de la Muestra. (S)

La Desviación Estándar Muestral, es la raíz cuadrada de la variancia muestral.

$$S = \sqrt[2]{S^2}$$

EJEMPLO:

Basándonos en el ejemplo anterior calcule la Desviación Estándar de la Muestra.

$$S = \sqrt[2]{S^2} = \sqrt[2]{7,5} = 2,74$$

Por lo tanto la Desviación Estándar de la Muestra es igual a 2,74.

El Rango:

Para datos no agrupados, el Rango es la diferencia entre los valores mayor y menor en un conjunto de datos.

Rango = Valor Mayor - Valor Menor

EJEMPLO:

Usando los datos anteriores de la población

18, 29, 22, 24, 20, 23, 25, 19, 18, 17, 22,21, 20, 18, 19

Rango = 29 – 17
Rango = 12

Quiere decir que la diferencia entre el Valor mayor y el Valor menor del conjunto de datos es 12.

Coeficiente de Variación:

Es la razón de la desviación estándar con respecto a la media aritmética, expresada como porcentaje. Su fórmula para calcularla es la siguiente:

$$CV = \frac{\sigma}{\mu} * 100\%$$

Si los datos se obtienen de una Población.

Donde:

σ *Es la Desviación Estándar de la Población.*
μ *Es la Media Aritmética de la Población.*

$$CV = \frac{S}{\overline{X}} * 100\%$$ ***Si los datos se obtienen de una Muestra.***

Donde:

S *Es la Desviación Estándar de la Muestra.*
$\overline{X}$ *Es la Media Aritmética de la Muestra.*

EJEMPLO:

Del ejercicio anterior sabemos que la σ = 3,1416 y μ = 21. Calcule el Coeficiente de Variación.

$$CV = \frac{\sigma}{\mu} * 100\% = \frac{3{,}1416}{21} * 100\% = 0{,}1496 * 100\% = 14{,}96\%$$

Asimetría (Sesgo) Sk:

Es la medida de la falta de simetría en una distribución. El coeficiente de asimetría se calcula mediante la siguiente fórmula:

Nota:

Si el Sesgo es positivo, se dice que la población está sesgada positivamente y la mayor cantidad de datos se encuentra a la derecha de la media.

Si el Sesgo es negativo, se dice que la población está sesgada negativamente y la mayor cantidad de datos se encuentra a la izquierda de la media.

Si el Sesgo es Cero, tanto la media, la mediana y la Moda son iguales.

$$Sk = \frac{3\left(\mu - \tilde{X}\right)}{\sigma}$$ ***Si los datos provienen de una Población***

Donde:

μ ***Es la Media de la Población***

$\tilde{X}$ ***Es la Mediana de la Población***

σ ***Es la Desviación Estándar de la Población***

$$Sk = \frac{3\left(\bar{X} - \tilde{X}\right)}{S}$$ ***Si los datos provienen de la Muestra***

Donde:

$\bar{X}$ ***Es la Media de la Muestra***

$\tilde{X}$ ***Es la Mediana de la Muestra***

S ***Es la Desviación Estándar de la Muestra***

EJEMPLO:

Si del ejemplo anterior la μ=21; $\tilde{X}$= 20; σ=3,1416. Calcule el Sesgo

$$Sk = \frac{3\left(\mu - \tilde{X}\right)}{\sigma} = \frac{3(21-20)}{3,1416} = \frac{3(1)}{3,1416} = 0,954$$

Por lo tanto podemos decir que la población se encuentra sesgada Positivamente y la mayor cantidad de datos se encuentra a la derecha de la Medía.

Cuartil 1

El primer cuartil es el valor correspondiente al punto debajo del cual se encuentra el 25% de las observaciones en un conjunto ordenado de datos. Su

Fórmula para calcularlo tanto para Población como para Muestra es la misma forma:

$$Q_1 = \frac{25}{100}(n+1)$$
$$Q_1 = \frac{1}{4}(n+1)$$

EJEMPLO:

Con los datos del Ejercicio anterior Calcule el Primer Cuartil.

18, 29, 22, 24, 20, 23, 25, 19, 18, 17, 22,21, 20, 18, 19, 29

Solución:

1) Debemos ordenar el conjunto de datos:

17, 18, 18, 18, 19, 19, 20, 20, 21, 22, 22, 23, 24, 25, 29, 29

2) Ahora procedemos a calcular el Primer Cuartil.

$$Q_1 = \frac{1}{4}(n+1) = \frac{1}{4}(16+1) = \frac{1}{4}(17) = 4,25^{\circ}$$

Este valor de 4,25 aun no es el resultado, lo que indica este valor es la posición 4,25 en el conjunto de datos ordenados. El valor 4 indica la posición 4 en mis datos ordenados la mima que sería el valor 18, y como podemos observar existe un sobrante de 0,25 el mismo que lo obtendremos tomando el siguiente valor que sigue a la posición 4 en este caso 19. De la siguiente forma: restamos el valor mayor del menor y al resultado lo multiplicamos por el sobrante en este caso 0,25 , el resultado que se obtenga se sumará al primer valor obtenido en este caso 18, y este dará como resultado el valor del Primer Cuartil.

19-18=1*0.25= 0.25+18= 18,25

Q1= 18,25

Por lo tanto el Primer cuartil es 18,25, lo que quiere decir que el 25% de los datos o de la información se encuentra por debajo de 18,25.

Nota:

Si el valor sale exacto ósea si por ejemplo hubiese sido 4 el resultado obtenido del primer cálculo entonces el primer cuartil sería solamente 18.

Cuartil 3

El tercer cuartil es el valor correspondiente al punto debajo del cual se encuentra 75% de las observaciones en un conjunto ordenado de datos. . Su Fórmula para calcularlo tanto para Población como para Muestra es la misma forma:

$$Q_3 = \frac{75}{100}(n+1)$$

$$Q_3 = \frac{3}{4}(n+1)$$

EJEMPLO:

Con los datos del Ejercicio anterior Calcule el Tercer Cuartil.

18, 29, 22, 24, 20, 23, 25, 19, 18, 17, 22,21, 20, 18, 19, 29

Solución:

1) Debemos ordenar el conjunto de datos:

17, 18, 18, 18, 19, 19, 20, 20, 21, 22, 22, 23, 24, 25, 29, 29

2) Ahora procedemos a calcular el Tercer Cuartil.

$$Q_3 = \frac{3}{4}(n+1) = \frac{3}{4}(16+1) = \frac{3}{4}(17) = 12,75^{\circ}$$

Este valor de 12,75 aun no es el resultado, lo que indica este valor es la posición 12,75 en el conjunto de datos ordenados. El valor 12 indica la posición

12 en mis datos ordenados la mima que sería el valor 23, y como podemos observar existe un sobrante de 0,75 el mismo que lo obtendremos tomando el siguiente valor que sigue a la posición 12 en este caso 24. De la siguiente forma: restamos el valor mayor del menor y al resultado lo multiplicamos por el sobrante en este caso 0,75 , el resultado que se obtenga se sumará al primer valor obtenido en este caso 23, y este dará como resultado el valor del Tercer Cuartil.

24-23 = 1*0,75= 0.75 + 23 = 23,75

Q3= 23,75

Por lo tanto el Primer cuartil es 23,75, lo que quiere decir que el 75% de los datos o de la información se encuentra por debajo de 23,75.

Amplitud Intercuartílica

Es la distancia entre el tercer cuartil Q3 y el primer cuartil Q1.

$$AI = Q_3 - Q_1$$

EJEMPLO:

Con los datos anteriores tenemos que:

Q3=23,75
Q1= 18,25

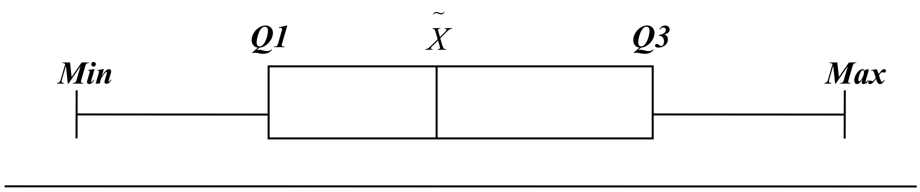
$$AI = Q_3 - Q_1$$
$$AI = 23,75 - 18,25$$
$$AI = 5,5$$

Por lo tanto podemos decir que la distancia entre el tercer y el primer Cuartil es de 5,5.

Diagrama de Caja:

Es una ilustración gráfica, basada en cuartiles, que ayuda a visualizar un conjunto de datos.

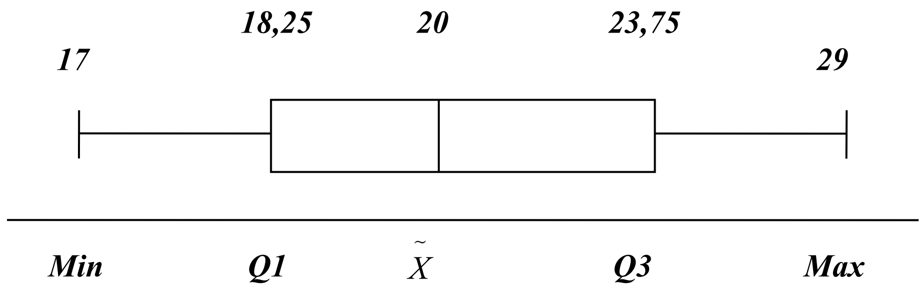
Se requieren cinco tipos de datos para construir un diagrama de caja: el valor mínimo, el primer cuartil, la mediana, el tercer cuartil, y el valor máximo.



EJEMPLO:

Con los datos del ejercicio anterior elabores el Diagrama de Caja correspondiente.

17, 18, 18, 18, 19, 19, 20, 20, 21, 22, 22, 23, 24, 25, 29, 29



Medidas de Dispersión Para datos Agrupados

Variancia Muestral para Datos Agrupados

La fórmula de la variancia para datos agrupados usada como estimador de la varianza poblacional es:

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

Donde:

f	***Es la frecuencia de la clase***
X	***Es la Marca de clase***
n	***Es el número de observaciones***

EJEMPLO:

Retomando el ejemplo 2 del capítulo 3, donde se registran las ventas de un departamento en una empresa. Calcule Varianza:

Clases	Frecuencias(f)
[20,30]	6
[30,40]	7
[40,50]	4
[50,60]	8
[60,70]	5
	30

Solución:

Lo Primero que tenemos que hacer es elevar al cuadrado la Marca de clase, y luego multiplicar la Frecuencia de clase por la Marca de clase al cuadrado.

Clases	Frecuencias(f)	Marca de Clases (X)	$f*X$	FA	X^2	$f*X^2$
[20,30]	6	25	150	6	625	3750
[30,40]	7	35	245	13	1225	8575
[40,50]	4	45	180	17	2025	8100
[50,60]	8	55	440	25	3025	24200
[60,70]	5	65	325	30	4225	21125
	30		1340			65750

$$S^2 = \frac{\sum_{i=1}^{n}(f_i * X_i^2) - \frac{\sum_{i=1}^{n}(f_i * X_i)^2}{n}}{n-1} = \frac{65750 - \frac{(1340)^2}{30}}{30-1} = \frac{65750 - \frac{1795600}{30}}{29}$$

$$S^2 = \frac{65750 - 59853,33}{29} = \frac{5896,67}{29} = 203,33$$

Podemos ver que la varianza es 203,33 lo que indica que existe una alta variabilidad en los datos.

Desviación Estándar de los Datos Agrupados:

Su fórmula para calcular es la siguiente:

$$S = \sqrt[2]{S^2}$$

EJEMPLO:

Con la Varianza obtenida en el ejemplo anterior calcule la Desviación Estándar:

$$S = \sqrt[2]{S^2} = \sqrt[2]{203,3} = 14,26$$

Por lo tanto la Desviación Estándar de la Varianza es 14,26.

El Rango:

Para datos no agrupados, el Rango es la diferencia entre los valores mayor y menor en un conjunto de datos.

Rango = Valor Mayor - Valor Menor

EJEMPLO:

Usando los datos de la tabla anterior es:

Clases	Frecuencias(f)	Marca de Clases (X)	$f*X$	FA	X^2	$f*X^2$
[20,30]	6	25	150	6	625	3750
[30,40]	7	35	245	13	1225	8575
[40,50]	4	45	180	17	2025	8100
[50,60]	8	55	440	25	3025	24200
[60,70]	5	65	325	30	4225	21125
	30		1340			65750

Rango = 70-20
Rango = 50

Entonces la Diferencia entre el valor mayor y el valor menor es de 50.

Coeficiente de Variación:

Es la razón de la desviación estándar con respecto a la media aritmética, expresada como porcentaje. Su fórmula para calcularla es la siguiente:

$$CV = \frac{S}{\overline{X}} * 100\%$$

Donde:

S *Es la Desviación Estándar de la Muestra.*
$\bar{X}$ *Es la Media Aritmética de la Muestra.*

EJEMPLO:

Del ejercicio Anterior sabemos que la S= 14,26 y $\bar{X}$ ***= 44,67. Calcule el Coeficiente de Variación.***

$$CV = \frac{S}{\bar{X}} * 100\% = \frac{14,26}{44,67} * 100\% = 0,3192 * 100\% = 31,92\%$$

Asimetría (Sesgo) Sk:

Es la medida de la falta de simetría en una distribución. El coeficiente de asimetría se calcula mediante la siguiente fórmula:

Nota:

Si el Sesgo es positivo, se dice que la población está sesgada positivamente y la mayor cantidad de datos se encuentra a la derecha de la media.

Si el Sesgo es negativo, se dice que la población está sesgada negativamente y la mayor cantidad de datos se encuentra a la izquierda de la media.

Si el Sesgo es Cero, tanto la media, la mediana y la Moda son iguales.

$$Sk = \frac{3\left(\bar{X} - \tilde{X}\right)}{S}$$

Donde:

$\bar{X}$ ***Es la Media de la Muestra***
$\tilde{X}$ ***Es la Mediana de la Muestra***
S ***Es la Desviación Estándar de la Muestra***

EJEMPLO:

Basado en los resultados del ejercicio anterior, $\bar{X}$ ***= 44,67;*** $\tilde{X}$ ***= 35;*** S ***=14,26.***

$$Sk = \frac{3\left(\overline{X} - \tilde{X}\right)}{S} = \frac{3(44,67-35)}{14,26} = \frac{3(9,67)}{14,26} = \frac{29,01}{14,26}$$

$$Sk = 2,03$$

Por lo tanto podemos decir que la población se encuentra sesgada Positivamente y la mayor cantidad de datos se encuentra a la derecha de la Medía.

Cuartil 1:

El primer cuartil es el valor correspondiente al punto debajo del cual se encuentra el 25% de las observaciones en un conjunto ordenado de datos. Su Fórmula para calcularlo es la siguiente:

$$Q_1 = L + \left(\frac{n/4 - FA}{f} \right) * i$$

Donde:

L	*Es el límite inferior de la clase que contiene a la Mediana*
n	*Es el tamaño de la Muestra o Población*
FA	*Es la Frecuencia Acumulada de la clase que contiene a la Mediana*
f	*Es la Frecuencia de la clase que contiene a la Mediana*
i	*Es el tamaño del intervalo.*

EJEMPLO:

Siguiendo con el ejemplo anterior de la tabla. Calcule el Primer Cuartil.

Clases	Frecuencias(f)	Marca de Clases (X)	$f*X$	FA	X^2	$f*X^2$
[20,30]	6	25	150	6	625	3750
[30,40]	7	35	245	13	1225	8575
[40,50]	4	45	180	17	2025	8100
[50,60]	8	55	440	25	3025	24200
[60,70]	5	65	325	30	4225	21125
	30		1340			65750

Para determinar la clase que contiene al Primer cuartil de datos agrupados debemos:

- ***Divida el número total de datos entre 4. Determine qué clase contiene este valor. Por ejemplo, si n=30, 30/4 = 7,5, después determine qué clase contiene el 7,5° valor (la clase del Primer Cuartil) . El mismo que se encuentra en la frecuencia Acumulada donde la suma llega a 13, ya que aquí esta contenido la posición 7,5 de los datos ordenados.***

Entonces lo siguiente es trabajar solamente con a fila que contiene al Primer cuartil, en este caso la que está señalada de amarillo.

$$Q_1 = L + \left(\frac{n/4 - FA}{f} \right) * i = 30 + \left(\frac{7,5 - 13}{7} \right) * 10 = 30 + \left(\frac{-5,5}{7} \right) * 10$$

$$Q_1 = 30 + (-0,785) * 10 = 30 - 7,85 = 22,15$$

Lo que indica que el 25% de los Datos se encuentran por debajo de 22,15.

Cuartil 3:

El tercer cuartil es el valor correspondiente al punto debajo del cual se encuentra 75% de las observaciones en un conjunto ordenado de datos. Su Fórmula para calcularlo es la siguiente:

$$Q_3 = L + \left(\frac{3n/4 - FA}{f} \right) * i$$

Donde:

L	*Es el límite inferior de la clase que contiene a la Mediana*
n	*Es el tamaño de la Muestra o Población*
FA	*Es la Frecuencia Acumulada de la clase que contiene a la Mediana*
f	*Es la Frecuencia de la clase que contiene a la Mediana*
i	*Es el tamaño del intervalo.*

EJEMPLO:

Siguiendo con el ejemplo anterior de la tabla. Calcule el Tercer Cuartil.

Clases	*Frecuencias(f)*	*Marca de Clases (X)*	$f*X$	*FA*	X^2	$f*X^2$
[20,30]	*6*	*25*	*150*	*6*	*625*	*3750*
[30,40]	*7*	*35*	*245*	*13*	*1225*	*8575*
[40,50]	*4*	*45*	*180*	*17*	*2025*	*8100*
[50,60]	*8*	*55*	*440*	*25*	*3025*	*24200*
[60,70]	*5*	*65*	*325*	*30*	*4225*	*21125*
	30		*1340*			*65750*

Para determinar la clase que contiene al Tercer cuartil de datos agrupados debemos:

Primero multiplique el número total de datos por 3, y luego divida el número total de datos entre 4. Determine qué clase contiene este valor. Por ejemplo, si n=30, (3)30/4 = 22,25, después determine qué clase contiene el 22,25° valor (la clase del Tercer Cuartil). El mismo que se encuentra en la frecuencia Acumulada donde la suma llega a 25, ya que aquí esta contenido la posición 22,25 de los datos ordenados.

Entonces lo siguiente es trabajar solamente con a fila que contiene al Tercer cuartil, en este caso la que está señalada.

$$Q_3 = L + \left(\frac{3n/4 - FA}{f} \right) * i = 50 + \left(\frac{22,25 - 25}{8} \right) * 10 = 50 + \left(\frac{-2,75}{8} \right) * 10$$

$$Q_1 = 50 + (-0,344) * 10 = 50 - 3,44 = 46,56$$

Lo que indica que el 75% de los Datos se encuentran por debajo de 46,56.

Amplitud Intercuartílica:

Es la distancia entre el tercer cuartil Q3 y el primer cuartil Q1.

$$AI = Q_3 - Q_1$$

EJEMPLO:

Con los datos anteriores tenemos que:

Q3= 46,56
Q1= 22,15

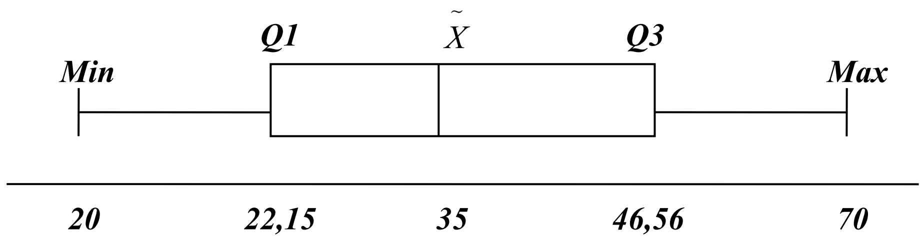
$$AI = Q_3 - Q_1$$
$$AI = 46,56 - 22,15$$
$$AI = 24,41$$

Por lo tanto podemos decir que la distancia entre el tercer y el primer Cuartil es de 24,41.

Diagrama de Caja:

Es una ilustración gráfica, basada en cuartiles, que ayuda a visualizar un conjunto de datos.

Se requieren cinco tipos de datos para construir un diagrama de caja: el valor mínimo, el primer cuartil, la mediana, el tercer cuartil, y el valor máximo.



- ### *Ejercicio Completo de Datos Desagrupados*

Se tiene 15 estudiantes en un curso de inglés, y se desea saber el promedio de Edad de los estudiantes que están en dicha clase. Las edades son las siguientes:

18, 29, 22, 24, 20, 23, 25, 19, 18, 17, 22,21, 20, 18, 19

Medía Aritmética

$$\mu = \frac{\sum_{i=1}^{15} X_i}{15} = \frac{18+29+22+24+20+23+25+19+18+17+22+21+20+18+19}{15}$$

$$\mu = \frac{315}{15} = 21$$

Entonces podemos decir que el promedio de edad de los estudiantes en dicho curso es de 21 años.

Mediana

1) Lo primero que debemos hacer es ordenar los datos:

17, 18, 18, 18, 19, 19, 20, 20, 21, 22, 22, 23, 24, 25, 29,

2) Ahora debemos ver si la población o la muestra es Par o Impar, como podemos observar la población es IMPAR. Entonces utilizamos la fórmula cuando es IMPAR.

$$\tilde{X} = X'_{\frac{n+1}{2}} = X'_{\frac{15+1}{2}} = X'_{16/2} = X'_{8}$$

Ahora debemos buscar en el conjunto de datos ordenados, la posición 8 de los datos y esa es la mediana

$$\tilde{X} = 20$$

Como podemos ver la mediana es 20

- ***La Moda*** $\hat{X}$ ***:***

$$\hat{X} = 18$$

Como el valor que más veces se repite es el 18 está es la moda.

Varianza de la Población

$$\sigma^2 = \frac{\sum_{i=1}^{15}(X_i - 21)^2}{15} =$$

$$\sigma^2 = \frac{\begin{array}{l}(18-21)^2+(29-21)^2+(22-21)^2+(24-21)^2+(20-21)^2+(23-21)^2+(25-21)^2+(19-21)^2+ \\ +(18-21)^2+(17-21)^2+(22-21)^2+(21-21)^2+(20-21)^2+(18-21)^2+(19-21)^2\end{array}}{15}$$

$$\sigma^2 = \frac{\begin{array}{l}(-3)^2+(8)^2+(1)^2+(3)^2+(-1)^2+(2)^2+(4)^2+(-2)^2+(-3)^2+ \\ +(-4)^2+(1)^2+(0)+(-1)^2+(-3)^2+(-2)^2\end{array}}{15}$$

$$\sigma^2 = \frac{9+64+1+9+1+4+16+4+9+16+1+0+1+9+4}{15}$$

$$\sigma^2 = \frac{148}{15} = 9{,}87$$

Como se puede observar la Varianza es igual a 9,87, lo que quiere decir que existe una moderada Variabilidad en los datos.

Desviación Estándar de la Población

$$\sigma = \sqrt[2]{\sigma^2} = \sqrt[2]{9{,}87} = 3{,}1416$$

Por lo tanto la Desviación Estándar de la Población es igual a 3,1416.

Rango

Rango = 29 – 17
Rango = 12

Quiere decir que la diferencia entre el Valor mayor y el Valor menor del conjunto de datos es 12.

Coeficiente de Variación

$$CV = \frac{\sigma}{\mu} * 100\% = \frac{3{,}1416}{21} * 100\% = 0{,}1496 * 100\% = 14{,}96\%$$

Sesgo o Coeficiente de Asimetría

$$Sk = \frac{3\left(\mu - \tilde{X}\right)}{\sigma} = \frac{3\left(21-20\right)}{3,1416} = \frac{3(1)}{3,1416} = 0,954$$

Por lo tanto podemos decir que la población se encuentra sesgada Positivamente y la mayor cantidad de datos se encuentra a la derecha de la Media.

Primer Cuartil

$$Q_1 = \frac{1}{4}(n+1) = \frac{1}{4}(16+1) = \frac{1}{4}(17) = 4,25^{\circ}$$

Este valor de 4,25 aun no es el resultado, lo que indica este valor es la posición 4,25 en el conjunto de datos ordenados. El valor 4 indica la posición 4 en mis datos ordenados la mima que sería el valor 18, y como podemos observar existe un sobrante de 0,25 el mismo que lo obtendremos tomando el siguiente valor que sigue a la posición 4 en este caso 19. De la siguiente forma: restamos el valor mayor del menor y al resultado lo multiplicamos por el sobrante en este caso 0,25 , el resultado que se obtenga se sumará al primer valor obtenido en este caso 18, y este dará como resultado el valor del Primer Cuartil.

19-18=1*0.25= 0.25+18= 18,25

Q1= 18,25

Por lo tanto el Primer cuartil es 18,25, lo que quiere decir que el 25% de los datos o de la información se encuentra por debajo de 18,25.

Tercer Cuartil

$$Q_3 = \frac{3}{4}(n+1) = \frac{3}{4}(16+1) = \frac{3}{4}(17) = 12,75^{\circ}$$

Este valor de 12,75 aun no es el resultado, lo que indica este valor es la posición 12,75 en el conjunto de datos ordenados. El valor 12 indica la posición 12 en mis datos ordenados la mima que sería el valor 23, y como podemos observar existe un sobrante de 0,75 el mismo que lo obtendremos tomando el siguiente valor que sigue a la posición 12 en este caso 24. De la siguiente forma: restamos el valor mayor del menor y al resultado lo multiplicamos por el sobrante en este caso 0,75 , el resultado que se obtenga se sumará al primer *valor obtenido en este caso 23, y este dará como resultado el valor del Tercer Cuartil.*

24-23 = 1*0,75= 0.75 + 23 = 23,75

Q3= 23,75

Por lo tanto el tercer cuartil es 23,75, lo que quiere decir que el 75% de los datos o de la información se encuentra por debajo de 23,75.

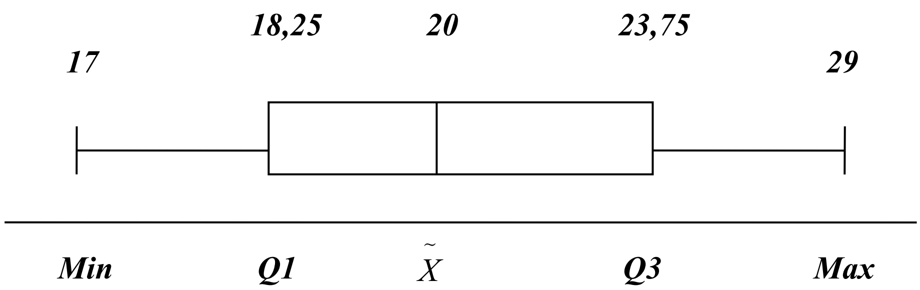
Amplitud Intercuartílica

Q3=23,75
Q1= 18,25

$$AI = Q_3 - Q_1$$
$$AI = 23,75 - 18,25$$
$$AI = 5,5$$

Por lo tanto podemos decir que la distancia entre el tercer y el primer Cuartil es de 5,5.

Diagrama De Caja



- ### *Ejercicio Completo de Datos Agrupados*

Los siguientes datos representan las ventas de un departamento en miles de dólares.

Clases	*Frecuencias(f)*
[20,30]	*6*

[30,40]	*7*
[40,50]	*4*
[50,60]	*8*
[60,70]	*5*
	30

Solución:

Tenemos que calcular la Marca de Clases (X), y luego multiplicamos la Columna de la frecuencia por la Columna de la Marca de clases. A continuación se muestra como quedaría el cuadro:

- ***Media Aritmética.***

$$\overline{X} = \frac{\sum_{i=1}^{30} X_i * f_i}{30} = \frac{1340}{30} = 44,67$$

Como podemos ver la Media es de 44,67 (en miles de dólares), en Ventas de dicho departamento.

Clases	*Frecuencias(f)*	*Marca de Clases (X)*	*f * X*
[20,30]	*6*	*25*	*150*
[30,40]	*7*	*35*	*245*
[40,50]	*4*	*45*	*180*
[50,60]	*8*	*55*	*440*
[60,70]	*5*	*65*	*325*
	30		*1340*

- ***Mediana***

Para determinar la clase de la mediana de datos agrupados:

- ***Elabore una distribución de frecuencias acumulada.***

- ***Divida el número total de datos entre 2. Determine qué clase contiene este valor. Por ejemplo, si n=30, 30/2 = 15, después determine qué clase contiene el 15° valor (la clase de la mediana). El mismo que se encuentra en la frecuencia Acumulada donde la suma llega a 17, ya que aquí esta contenido la posición 15 de los datos ordenados.***

Clases	Frecuencias(f)	Marca de Clases (X)	$f*X$	Frecuencia Acumulada (FA)
[20,30]	6	25	150	6
[30,40]	7	35	245	13
[40,50]	4	45	180	17
[50,60]	8	55	440	25
[60,70]	5	65	325	30
	30		1340	

Por lo tanto la Mediana es de 35 lo que quiere decir que el valor medio es este.

Moda

Clases	Frecuencias (f)	Marca de Clases (X)	$f*X$	Frecuencia Acumulada (FA)
[20,30]	6	25	150	6
[30,40]	7	35	245	13
[40,50]	4	45	180	17
[50,60]	8	55	440	25
[60,70]	5	65	325	30
	30		1340	

Como podemos ver la frecuencia Mayor es 8, y por lo tanto la moda $\hat{X} = 55$ ***ya que la Marca de clase es 55.***

Varianza

Lo Primero que tenemos que hacer es elevar al cuadrado la Marca de clase, y luego multiplicar la Frecuencia de clase por la Marca de clase al cuadrado.

Clases	Frecuencias(f)	Marca de Clases (X)	$f*X$	FA	X^2	$f*X^2$
[20,30)	6	25	150	6	625	3750
[30,40)	7	35	245	13	1225	8575
[40,50)	4	45	180	17	2025	8100
[50,60)	8	55	440	25	3025	24200
[60,70)	5	65	325	30	4225	21125
	30		1340			65750

$$S^2 = \frac{\sum_{i=1}^{n}(f_i * X_i^2) - \frac{\sum_{i=1}^{n}(f_i * X_i)^2}{n}}{n-1} = \frac{65750 - \frac{(1340)^2}{30}}{30-1} = \frac{65750 - \frac{1795600}{30}}{29}$$

$$S^2 = \frac{65750 - 59853,33}{29} = \frac{5896,67}{29} = 203,33$$

Podemos ver que la varianza es 203,33 lo que indica que existe una alta variabilidad en los datos.

- ***Desviación Estándar***

$$S = \sqrt[2]{S^2} = \sqrt[2]{203,3} = 14,26$$

Por lo tanto la Desviación Estándar de la Varianza es 14,26.

- ***Rango***

Rango = 70-20
Rango = 50

Entonces la Diferencia entre el valor mayor y el valor menor es de 50.

- ***Coeficiente de Variación***

$$CV = \frac{S}{\overline{X}} * 100\% = \frac{14,26}{44,67} * 100\% = 0,3192 * 100\% = 31,92\%$$

- ***Sesgo o Coeficiente de Asimetría***

$$Sk = \frac{3\left(\overline{X} - \tilde{X}\right)}{S} = \frac{3(44,67-35)}{14,26} = \frac{3(9,67)}{14,26} = \frac{29,01}{14,26}$$

$$Sk = 2,03$$

Por lo tanto podemos decir que la población se encuentra sesgada Positivamente y la mayor cantidad de datos se encuentra a la derecha de la Media.

- ***Primer Cuartil***

Clases	Frecuencias(f)	Marca de Clases (X)	$f*X$	FA	X^2	$f*X^2$
[20,30]	6	25	150	6	625	3750
[30,40]	7	35	245	13	1225	8575
[40,50]	4	45	180	17	2025	8100
[50,60]	8	55	440	25	3025	24200
[60,70]	5	65	325	30	4225	21125
	30		1340			65750

Para determinar la clase que contiene al Primer cuartil de datos agrupados debemos:

- ***Divida el número total de datos entre 4. Determine qué clase contiene este valor. Por ejemplo, si n=30, 30/4 = 7,5, después determine qué clase contiene el 7,5° valor (la clase del Primer Cuartil). El mismo que se encuentra en la frecuencia Acumulada donde la suma llega a 13, ya que aquí esta contenido la posición 7,5 de los datos ordenados.***

Entonces lo siguiente es trabajar solamente con a fila que contiene al Primer cuartil, en este caso la que está señalada de amarillo.

$$Q_1 = L + \left(\frac{n/4 - FA}{f} \right) * i = 30 + \left(\frac{7,5 - 13}{7} \right) * 10 = 30 + \left(\frac{-5,5}{7} \right) * 10$$

$$Q_1 = 30 + (-0,785) * 10 = 30 - 7,85 = 22,15$$

Lo que indica que el 25% de los Datos se encuentran por debajo de 22,15.

- ***Tercer Cuartil***

Clases	Frecuencias(f)	Marca de Clases (X)	$f*X$	FA	X^2	$f*X^2$
[20,30]	6	25	150	6	625	3750
[30,40]	7	35	245	13	1225	8575
[40,50]	4	45	180	17	2025	8100
[50,60]	8	55	440	25	3025	24200
[60,70]	5	65	325	30	4225	21125
	30		1340			65750

Para determinar la clase que contiene al Tercer cuartil de datos agrupados debemos:

- ***Primero multiplique el número total de datos por 3, y luego divida el número total de datos entre 4. Determine qué clase contiene este valor.***

Por ejemplo, si n=30, (3)30/4 = 22,25, después determine qué clase contiene el 22,25° valor (la clase del Tercer Cuartil). El mismo que se encuentra en la frecuencia Acumulada donde la suma llega a 25, ya que aquí esta contenido la posición 22,25 de los datos ordenados.

Entonces lo siguiente es trabajar solamente con a fila que contiene al Tercer cuartil, en este caso la que está señalada.

$$Q_3 = L + \left(\frac{{}^{3n}\!/_{4} - FA}{f} \right) * i = 50 + \left(\frac{22,25 - 25}{8} \right) * 10 = 50 + \left(\frac{-2,75}{8} \right) * 10$$

$$Q_1 = 50 + (-0,344) * 10 = 50 - 3,44 = 46,56$$

Lo que indica que el 75% de los Datos se encuentran por debajo de 46,56.

- ***Amplitud Intercuartílica***

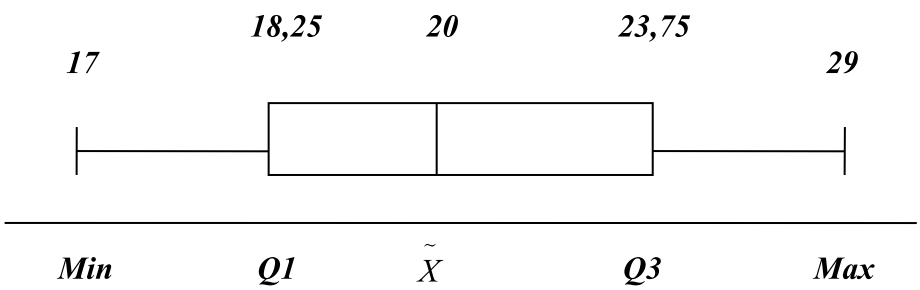
Con los datos anteriores tenemos que:

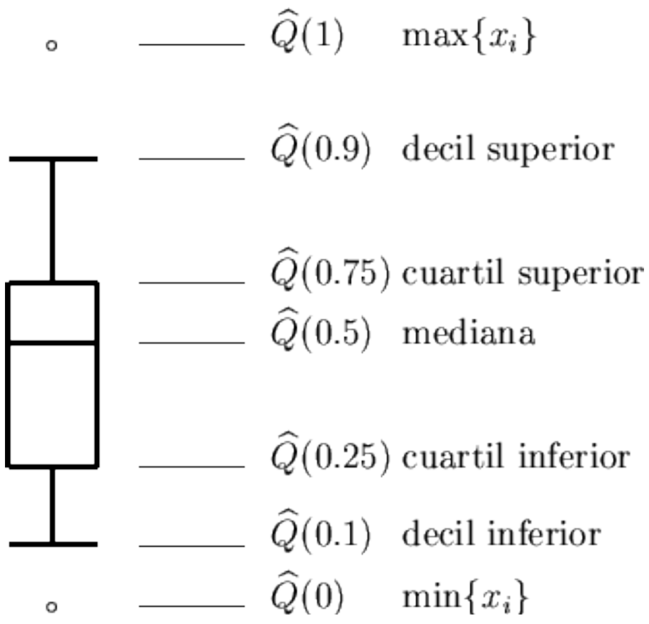
Q3= 46,56
Q1= 22,15

$$AI = Q_3 - Q_1$$
$$AI = 46,56 - 22,15$$
$$AI = 24,41$$

Por lo tanto podemos decir que la distancia entre el tercer y el primer Cuartil es de 24,41.

- ***Diagrama De Caja***





AUTOEVALUACIÓN # 4

REALICE LOS SIGUIENTES EJERCICIOS:

1. ***Las alturas de los jugadores de un equipo de baloncesto vienen dadas por la tabla:***

Altura	*[170, 175]*	*[175, 180]*	*[180, 185]*	*[185, 190]*	*[190, 195]*	*[195, 2.00]*
Nº de jugadores	*1*	*3*	*4*	*8*	*5*	*2*

Calcular:

*1. La **media**;*

*2. La **moda**;*

*3. La **desviación típica**;*

*4. La **Varianza**;*

*5. ¿Cuántos jugadores se encuentran por encima de la **media más una desviación típica**?*

2. ***Las calificaciones de 50 alumnos en Matemáticas han sido las siguientes:***

5, 2, 4, 9, 7, 4, 5, 6, 5, 7, 7, 5, 5, 2, 10, 5, 6, 5, 4, 5, 8, 8, 4, 0, 8, 4, 8, 6, 6, 3, 6, 7, 6,

Calcular:

*1. La **media**,*

*2. La **moda**,*

*3. La **desviación típica**,*

*4. La **Varianza***

*5. Calcular la **Tabla de Frecuencias.***

3. Se ha aplicado test a los empleados de una fábrica, obteniéndose las siete tablas:

	f_i	***MARCA DE CLASE (xi)***	***f_i (xi)***	***f_i (xi)^2***
[38, 44]	*7*			
[44, 50]	*8*			
[50, 56]	*15*			
[56, 62]	*25*			
[62, 68]	*18*			
[68, 74]	*9*			
[74, 80]	*6*			

Calcular:
1. Las ***frecuencias acumuladas;***
2. Dibujar el ***histograma.***

4. Las puntuaciones obtenidas por un grupo en una prueba han sido:

15, 20, 15, 18, 22, 13, 13, 16, 15, 19, 18, 15, 16, 20, 16, 15, 18, 16.

*Construir la **tabla de distribución de frecuencias** y dibuja el **polígono de frecuencias**.*

5. El número de estrellas de los hoteles de una ciudad viene dado por la siguiente serie:

3, 3, 4, 3, 4, 3, 1, 3, 4, 3, 3, 3, 2, 1, 3, 3, 3, 2, 3, 2, 2, 3, 3, 3, 2, 2,

Construir la tabla de distribución de frecuencias y dibuja el diagrama de barras.

6. Las calificaciones de 50 alumnos en Matemáticas han sido las siguientes:

5, 2, 4, 9, 7, 4, 5, 6, 5, 7, 7, 5, 5, 2, 10, 5, 6, 5, 4, 5, 8, 8, 4, 0, 8, 4.

*Construir la **tabla de distribución de frecuencias** y dibuja el **diagrama de barras**.*

7. Los pesos de los 65 empleados de una fábrica vienen dados por la siguiente tabla:

Peso	***[50, 60]***	***[60, 70]***	***[70, 80]***	***[80,90]***	***[90, 100]***	***[100, 110]***	***[110, 120]***
f_i	***8***	***10***	***16***	***14***	***10***	***5***	***2***

***1.** Construir la **tabla de frecuencias**,*

***2.** Representar el **histograma** y el **polígono de frecuencias**.*

8. Los 40 alumnos de una clase han obtenido las siguientes puntuaciones, sobre 50, en un examen de Física.

3, 15, 24, 28, 33, 35, 38, 42, 23, 38, 36, 34, 29, 25, 17, 7, 34, 36, 39, 44, 31, 26, 20, 11, 13, 22, 27, 47, 39, 37, 34, 32, 35, 28, 38, 41, 48.

***1.** Construir la **tabla de frecuencias**,*

***2.** Dibujar el **histograma** y el **polígono de frecuencias**.*

9. Sea una distribución estadística que viene dada por la siguiente tabla:

x_i	61	64	67	70	73
f_i	5	18	42	27	8

Calcular:

***1.** La **moda, mediana y media**,*

***2.** El **rango, desviación media, varianza y desviación típica**.*

10. ***Calcular la media, la mediana y la moda de la siguiente serie de números:***

5, 3, 6, 5, 4, 5, 2, 8, 6, 5, 4, 8, 3, 4, 5, 4, 8, 2, 5, 4

17. ***Hallar la varianza y la desviación típica de la siguiente serie de datos:***

12, 6, 7, 3, 15, 10, 18, 5

18. ***Hallar la media, mediana y moda de la siguiente serie de números:***

3, 5, 2, 6, 5, 9, 5, 2, 8, 6

19. ***Hallar la desviación media, la varianza y la desviación típica de la series de números siguientes:***

2, 3, 6, 8, 11.

12, 6, 7, 3, 15, 10, 18, 5.

20. Se ha aplicado un test a los empleados de una fábrica, obteniéndose la siguiente tabla:

	f_i
[38, 44)	*7*
[44, 50)	***8***
[50, 56)	***15***
[56, 62)	***25***
[62, 68)	***18***
[68, 74)	***9***
[74, 80)	***6***

Dibujar el ***histograma*** *y el* ***polígono de frecuencias acumuladas****.*

21. Dadas las series estadísticas:

3, 5, 2, 7, 6, 4, ,9, 9.

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

Calcular:

*1. La **moda**, la **mediana** y la **media**,*

*2. La **desviación media, la varianza** y la **desviación típica**,*

3. Los cuartiles 1º y 3º,

*4. Los **deciles** 2º y 7º,*

*5. Los **percentiles** 32 y 85.*

22. Una distribución estadística viene dada por la siguiente tabla:

	[10, 15]	[15, 20]	[20, 25]	[25, 30]	[30, 35]
f_i	3	5	7	4	2

Hallar:

*1. La **moda, mediana** y **media**,*

*2. El **rango**, **desviación media** y **varianza**,*

*3. Los **cuartiles** 1º y 3º,*

*4. Los **deciles** 3º y 6º,*

*5. Los **percentiles** 30 y 70.*

17. Dada la distribución estadística:

	[0, 5]	[5, 10]	[10, 15]	[15, 20]	[20, 25]	[25, ∞]
f_i	3	5	7	8	2	6

Calcular:

*1. La **mediana y moda**,*

*2. **Cuartil** 2º y 3º,*

*3. **Media**.*

CAPITULO 5

Autores:

Ing. Luis Palacios Torres. Mgs
UNIVERSIDAD ESTATAL DE MILAGRO

Mdc. Carlos Alcívar Trejo
UNIVERSIDAD ECOTEC

Msc. Roberto Israel Soria Vera
UNIVERSIDAD DE GUAYAQUIL

La esencia de la vida es la improbabilidad estadística a escala colosal. *"**El relojero ciego**" (1986)*, ***Richard Dawkins***

CAPITULO 5

Conceptos Probabilísticos

OBJETIVOS

Definir lo que es probabilidad. Describir los enfoques clásico, empírico y subjetivo para la probabilidad. Entender los conceptos: experimento, evento, resultado, permutaciones y combinaciones. Definir los conceptos: probabilidad condicional y probabilidad conjunta. Calcular probabilidades aplicando las reglas de adición y multiplicación. Utilizar un diagrama de árbol para organizar y calcular probabilidades. Determine el número de permutaciones y el número de combinaciones.

Desarrollo:

Probabilidad: La probabilidad es un valor entre cero y uno, incluidos, que describe la posibilidad relativa de que ocurra un evento. A medida que el valor se acerque a uno significa que tiene más posibilidades de ocurrir, mientras que si se acerca a cero es menos probable que ocurra(Anderson & Peralta Rosales, 2014).

Experimento: proceso que conduce a la ocurrencia de una de varias observaciones posibles.

Resultado: lo que resulta en particular de un experimento.

Evento: ***conjunto de uno o más resultados de un experimento.***

Enfoques de la Probabilidad:

Probabilidad clásica:

Se basa en la consideración de que los resultados de un experimento son igualmente posibles(Devore, 2016)**.**

$$P = \frac{\text{Numero de resultados favorables}}{\text{Numero de resultados posibles}}$$

EJEMPLO:

Considere el experimento de lanzar dos monedas al mismo tiempo. ¿Calcule la probabilidad de obtener una cara en dichos lanzamientos?

El espacio muestral S = {CC, CS, SC, SS}

$$P=\frac{\text{Numero de resultados favorables}}{\text{Numero de resultados posibles}}$$

$$P=\frac{2}{4}=0,5$$

Como podemos ver hay 2 casos en que ocurre una cara, entre 4 posibles resultados.

Entonces la probabilidad de que se obtenga una cara en dos lanzamientos es de 0,5; que tiene exactamente la mitad de posibilidades que ocurra.

Calcule la probabilidad de obtener al menos una cara en dichos lanzamientos.

Si nos dicen que obtengamos al menos una cara, es posible obtener una o dos caras en dos lanzamientos.

$$P=\frac{3}{4}=0,75$$

Como podemos ver hay 3 casos en que ocurre al menos una cara, entre 4 posibles resultados. Entonces la probabilidad de que se obtenga al menos una cara en dos lanzamientos es de 0,75, lo que quiere decir que es muy probable que ocurra.

Calcule la probabilidad de obtener exactamente 2 caras en dichos lanzamientos.

$$P=\frac{1}{4}=0,25$$

Como podemos ver hay 1 caso en que ocurre 2 caras, entre 4 posibles resultados. Entonces la probabilidad de que se obtenga 2 caras en dos lanzamientos es de 0,25, lo que quiere decir que no es muy probable que ocurra pero puede hacerlo.

Calcule la probabilidad de no obtener ninguna cara en dichos lanzamientos.

$$P=\frac{1}{4}=0,25$$

Como podemos ver hay 1 caso en que no ocurre ninguna cara, entre 4 posibles resultados. Entonces la probabilidad de que no se obtenga ninguna cara en dos lanzamientos es de 0,25, lo que quiere decir que no es muy probable que ocurra pero puede hacerlo.

Eventos Mutuamente Excluyentes:

La ocurrencia de cualquier evento implica que ningún otro puede ocurrir al mismo tiempo.

EJEMPLO:

En el Ejemplo anterior, los cuatro resultados posibles son mutuamente excluyentes.

Colectivamente Exhaustivos:

Por lo menos uno de los eventos debe ocurrir cuando se realiza un experimento.

EJEMPLO:

En el Ejemplo anterior, los cuatro resultados posibles son colectivamente exhaustivos. En otras palabras, la suma de las probabilidades es igual a 1.

1= (.25 + .25 + .25 + .25).

Probabilidad Empírica

La probabilidad de que un evento ocurra a largo plazo se determina observando en qué fracción de tiempo sucedieron eventos semejantes en el pasado. Su forma de calcularlo es la siguiente (Véliz Capuñay, 2015)***:***

$$P = \frac{\text{Numero de veces que ocurrio el evento en el pasado}}{\text{Numero total de obsevaciones}}$$

EJEMPLO:

A lo largo de su carrera, la profesora Jones ha otorgado 186 calificaciones de "A" entre sus 1200 estudiantes. ¿Cuál es la probabilidad de que un estudiante de su clase en este semestre reciba una A?

$$P = \frac{\text{Numero de veces que ocurrio el evento en el pasado}}{\text{Numero total de obsevaciones}}$$

$$P = \frac{186}{1200} = 0{,}155$$

Como se puede observar la probabilidad de que dicha maestra otorgue una calificación de A, a algún estudiante en este semestre fue de 0,155; lo que indica que no es muy probable que esto ocurra en este semestre.

Probabilidad Subjetiva:

La posibilidad (probabilidad) de que suceda un evento específico que asigna una persona con base en cualquier información disponible.

EJEMPLOS:

- ***Un de probabilidad subjetiva son estimar la probabilidad de que los Redskins de Washington ganen el Super Bowl el próximo año.***

- ***Estimar la probabilidad de que ocurra un terremoto en Los Ángeles este año.***

Reglas Básicas de Probabilidad:

Si los eventos son mutuamente excluyentes, la ocurrencia de cualquier evento impide que otro evento ocurra.

Regla de la Adición:

Si dos eventos A y B son mutuamente excluyentes, la regla especial de adición indica que la probabilidad de que ocurra A o B es igual a la suma de sus probabilidades respectivas(Mendenhall et al., 2016)***:***

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

EJEMPLO:

Una caja contiene 10 libros de los cuales 4 son novelas y 4 de un matemático y 2 de historia. Se elige al azar un libro ¿Cuál es la probabilidad de tomar un libro que sea una novela o de historia?

A: El libro es una novela $P(A) = 4/10 = 0{,}4$
B: El libro es de Historia $P(B) = 2/10 = 0{,}2$

P(A U B) = P(A) + P(B)
P(A U B) = 0,4+ 0,2
P(A U B) = 0,6

Entonces la probabilidad de tomar un libro de la caja y que este sea novela o historia, es 0,6 es decir existe un 60% de posibilidades de que esto ocurra.

Regla del Complemento:

Se utiliza para determinar la probabilidad de que ocurra un evento restando del número 1 la probabilidad de que un evento no ocurra.

Si P(A) es la probabilidad del evento A y P(~A) es el complemento de A(que no ocurra A). Entonces:

$$P(A) + P(\sim A) = 1$$
$$P(A) = 1 - P(\sim A)$$

EJEMPLO:

Si la probabilidad de que Barcelona gané el próximo campeonato es 0,60. ¿Cuál es la probabilidad de que esto no ocurra?

A: Barcelona gané el próximo campeonato.

$$P(A) + P(\sim A) = 1$$
$$0{,}60 + P(\sim A) = 1$$
$$P(\sim A) = 1 - 0{,}60$$
$$P(\sim A) = 0{,}40$$

Entonces la probabilidad de que no gané es de 0,40.

Regla general de adición

Si A y B son dos eventos que no son mutuamente excluyentes, entonces P(A o B) se calcula con la siguiente fórmula:

P(A U B) = P(A) + P(B) - P(A y B)

EJEMPLO:

En una muestra de 500 estudiantes, 320 dijeron tener un estéreo, 175 dijeron tener una TV y 100 dijeron tener ambos:

A: Tienen un Estéreo
B: Tienen un TV

P(A U B) = P(A) + P(B) - P(A y B)

$$\text{P(A U B)} = \frac{320}{500} + \frac{175}{500} - \frac{100}{500}$$
$$\text{P(A U B)} = \frac{395}{500} = 0,79$$

Como podemos observar es muy probable que al elegir a un estudiante de dicha muestra este tenga, un TV ó un Estéreo en su cuarto es de 0,79.

Probabilidad conjunta

Es una probabilidad que mide la posibilidad de que dos o más eventos ocurran juntos. Un ejemplo sería el hecho de que un estudiante tenga tanto un estéreo como una TV en su habitación(Crawley, 2014)**.**

La Regla Especial de la Multiplicación:

Requiere que dos eventos A y B sean independientes. Dos eventos A y B son independientes si la ocurrencia de una no afecta la probabilidad de ocurrencia del otro. La regla especial se escribe(Elorza Pérez Tejada, 2015)***:***

P(A y B) = P(A) * P(B)

EJEMPLO:

Peter posee dos inventarios independientes uno de otro. La probabilidad de que el inventario A aumente su valor el próximo año es 0,5. La probabilidad de que el B aumente el suyo es 0,7. ¿Cuál es la probabilidad de que ambos aumenten su valor el próximo año?

P(A y B) = (.5)(.7) = .35.

Entonces la probabilidad de que ambos aumenten su valor el próximo año es de 0,35.

¿Cuál es la probabilidad de que al menos uno aumente su valor el próximo año (esto implica que cualquiera de los dos o ambos aumenten)?

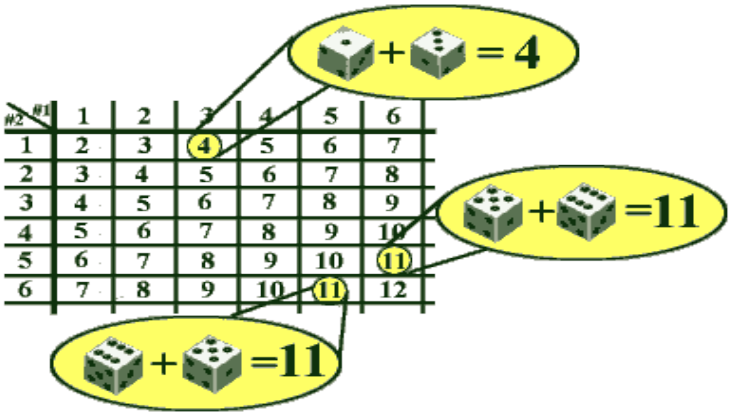
P(A U B) = (.5)(.3) + (.5)(.7) + (.7)(.5) = .85.

Como podemos observar que es más probable que al menos uno de los dos inventarios aumente su valor el próximo año es 0,85, lo que indica que es muy posible que esto ocurra.

Probabilidad Condicional

Es la probabilidad de que ocurra un evento en particular, dado que ocurrió otro evento(Fischetti, 2015)**.**

Nota: ***La probabilidad de que ocurra el evento A dado que ya ocurrió B se denota como P(A|B).***



La Regla General de la Multiplicación:

Se utiliza para determina la probabilidad conjunta de que ocurran dos eventos y establece: para dos eventos A y B, la probabilidad conjunta que ambos ocurran se encuentra multiplicando la probabilidad de A por la probabilidad condicional de B dado que A ocurrió. La probabilidad conjunta, P(A y B) está dada por la siguiente fórmula(I. Espejo Miranda et al., 2014)***:***

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

o

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

EJEMPLO:

Se tiene una caja con 10 focos de los cuales 6 están en buen estado. Si se extraen al azar 2 focos; calcule la probabilidad que ambos estén en buen estado.

A: Se extrae el primer foco en buen estado
B: Se extrae el segundo foco en buen estado

$$P(A) = \frac{6}{10}$$

$$P(B/A) = \frac{5}{9}$$

$$P(AyB) = P(A) * P(B/A)$$

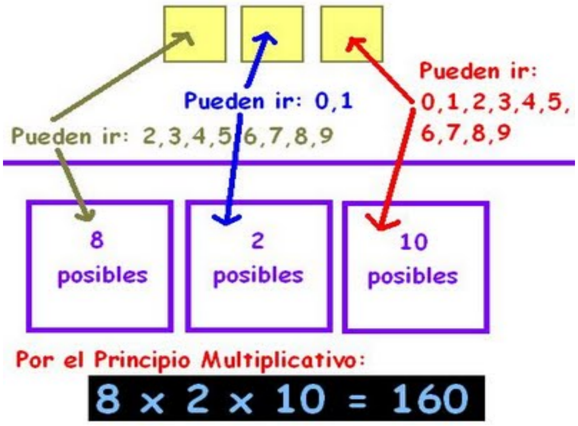
$$P(AyB) = \frac{6}{10} * \frac{5}{9} = \frac{30}{90} = 0,33$$

Por lo tanto la probabilidad de sacar de la caja los dos focos en buen estados uno después del otro es de 0,33.

Principios del Conteo

Fórmula de la multiplicación:

Si hay m modos de hacer una cosa y n formas de hacer otra, existen m x n formas de hacer ambas(Laverty & Kelly, 1995)**.**



EJEMPLO:

Si el doctor Pérez tiene 10 camisas y 8 corbatas. ¿Cuántos conjuntos de camisas y de corbatas tiene?

$$(10)(8) = 80.$$

Podemos ver que el Dr. Pérez puede combinar de 80 formas diferentes una camisa y una corbata.

Permutación:

Un arreglo de "r" objetos seleccionados a partir de un grupo único de "n" objetos posibles(Boslaugh & Watters, 2008)**.**

Nota: ***el orden del arreglo es importante en las permutaciones.***

Su Fórmula es la siguiente:

$${}_nP_r = \frac{n!}{(n-r)!}$$

EJEMPLO:

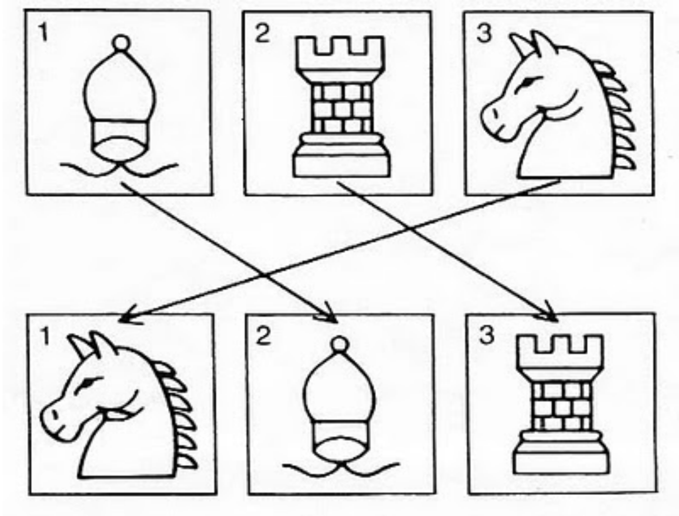
En un curso de 10 estudiantes, se desea elegir una directiva compuesta de un Presidente, un Tesorero, y un Secretario. ¿De cuántas formas diferentes se puede elegir una persona para cada dignidad?

$$_{n}P_{r} = \frac{n!}{(n-r)!}$$

$$_{10}P_{3} = \frac{10!}{(10-3)!} = \frac{10!}{7!} = \frac{10*9*8*7*6*5*4*3*2*1}{7*6*5*4*3*2*1}$$

$$_{10}P_{3} = 720$$

Podemos observar que se puede elegir de 720 formas diferentes de elegir a 3 personas de 10 para estas dignidades.



Combinación:

El número de modos para elegir "r" objetos de un grupo de "n" objetos sin considerar el orden. Su fórmula es la siguiente(GARG, 2018)***:***

$$_{n}C_{r} = \frac{n!}{r!*(n-r)!}$$

EJEMPLO:

Se tienen 5 Ingenieros y 10 Tecnólogos. Se requieren 2 Ingenieros y 4 Tecnólogos para un trabajo. ¿De cuántas formas diferentes se puede elegir?

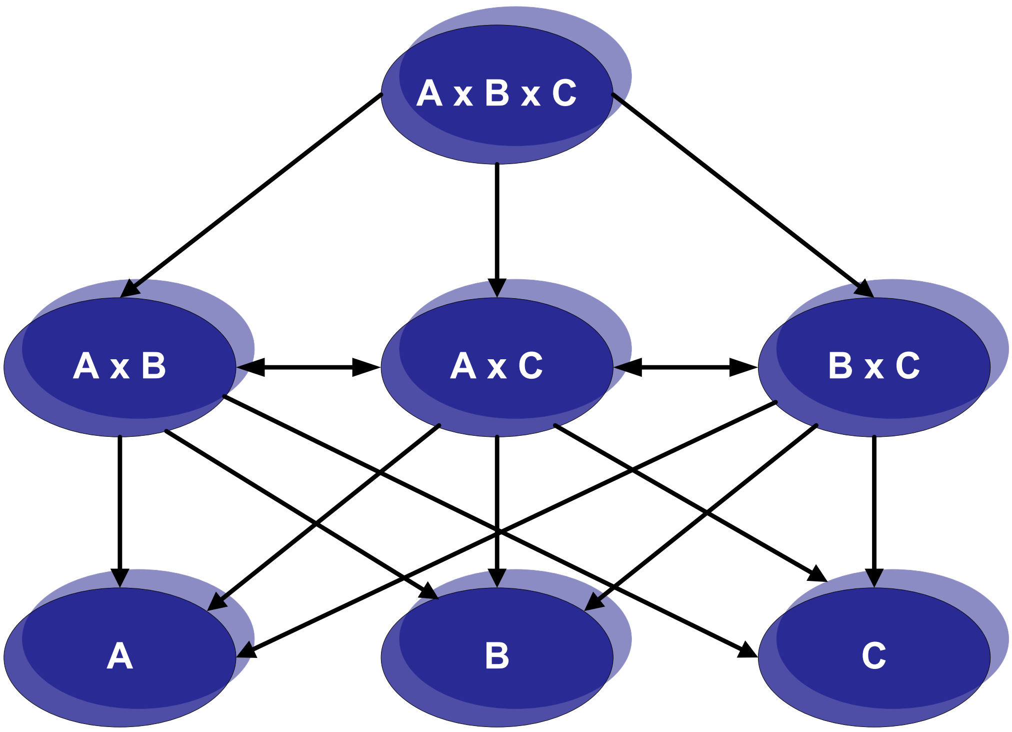
$$_5C_2 = \frac{5!}{2!*(5-2)!} = \frac{5!}{2!*3!} = \frac{5*4*3*2*1}{2*1*(3*2*1)} = 10$$

$$_{10}C_4 = \frac{10!}{4!*(10-4)!} = \frac{10!}{4!*6!} = \frac{10*9*8*7*6*5*4*3*2*1}{4*3*2*1*(6*5*4*3*2*1)} = 210$$

Entonces utilizando el principio del conteo:

10* 210 = 2100

Existen 2100 formas diferentes de elegir a 2 Ingenieros y 4 Tecnólogos para un trabajo.



AUTOEVALUACIÓN # 5

CONTESTE:

1. ***¿Es posible poner un ejemplo de dos sucesos que sean contrarios y a la vez compatibles?***

2. ***En un armario de cocina hay 6 refrescos de cola, 12 de naranja y 5 de limón. Cuando Ana iba a coger un refresco se fue la luz, y por tanto, lo tomó al azar.***

   Hallar las siguientes probabilidades:

   a) *Que sea de cola.*
   b) *Que sea de limón.*
   c) *Que sea de naranja.*
   d) *Que sea de cola o limón.*
   e) *Que no sea de limón.*

3. ***Luis tiene 10 cartas con los números siguientes: 1 2 2 3 3 4 5 8 8 9***

   Las pone hacia abajo y después las baraja. Su amigo Carlos coge una carta. Halla la probabilidad de que la carta escogida sea:

   a) *El 4.*
   b) *Mayor que 3.*
   c) *Divisible por 3.*
   d) *Múltiplo de 4.*
   e) *Impar.*
   f) *Menor que 7.*
   g) *Menor o igual que 7*

4. ***Sean A y B dos sucesos aleatorios con:***

$$p(A) = \frac{3}{8} \qquad p(B) = \frac{1}{2} \qquad p(A \cap B) = \frac{1}{4}$$

Hallar:

1. $p(A \cup B)$
2. $p(\overline{A})$
3. $p(\overline{B})$
4. $p(\overline{A} \cap \overline{B})$
5. $p(\overline{A} \cup \overline{B})$
6. $p(A \cap \overline{B})$
7. $p(B \cap \overline{A})$

5. ***Sean A y B dos sucesos aleatorios con:***

$$p(\overline{A}) = \frac{2}{3} \qquad p(A \cup B) = \frac{3}{4} \qquad p(A \cap B) = \frac{1}{4}$$

Hallar:

1. $p(A)$
2. $p(B)$
3. $p(A \cap \overline{B})$
4. $p(B \cap \overline{A})$

6. ***Se sacan dos bolas de una urna que se compone de una bola blanca, otra roja, otra verde y otra negra. Escribir el espacio muestral cuando:***

   a) La primera bola se devuelve a la urna antes de sacar la segunda.
   b) La primera bola no se devuelve.

7. ***Una urna tiene ocho bolas rojas, cinco amarillas y siete verdes. Si se extrae una bola al azar calcular la probabilidad de:***

   a) Sea roja.
   b) Sea verde.
   c) Sea amarilla.
   d) No sea roja.

8. ***Una urna contiene tres bolas rojas y siete blancas. Se extraen dos bolas al azar. Escribir el espacio muestral y hallar la probabilidad de los sucesos:***

   a. *Con reemplazamiento.* ***b.*** *Sin reemplazamiento.*

9. ***Se extrae una bola de una urna que contiene 4 bolas rojas, 5 blancas y 6 negras, ¿Cuál es la probabilidad de que la bola sea roja o blanca?, ¿Cuál es la probabilidad de que no sea blanca?***

10. ***En una clase hay 10 alumnas rubias, 20 morenas, 5 alumnos rubios y 10 morenos. Un día asisten 45 alumnos, encontrar la probabilidad de que un alumno:***

    1. *Sea hombre.*
    2. *Sea mujer morena.*
    3. *Sea hombre o mujer.*

11. Un dado está trucado, de forma que las probabilidades de obtener las distintas caras son proporcionales a los números de estas. Hallar:

1. *La probabilidad de obtener el 6 en un lanzamiento.*
2. *La probabilidad de conseguir un número impar en un lanzamiento.*

12. Se lanzan dos dados al aire y se anota la suma de los puntos obtenidos. Se pide:

1. *La probabilidad de que salga el 7.*
2. *La probabilidad de que el número obtenido sea par.*
3. *La probabilidad de que el número obtenido sea múltiplo de 3.*

13. Se lanzan tres dados. Encontrar la probabilidad de que:

1. *Salga 6 en todos.*
2. *Los puntos obtenidos sumen 7.*

14. Hallar la probabilidad de que al levantar unas fichas de dominó se obtenga un número de puntos mayor que 9 o que sea múltiplo de 4.

15. Busca la probabilidad de que al echar un dado al aire, salga:

1. *Un número par.*
2. *Un múltiplo de tres.*
3. *Mayor que cuatro.*

16. Hallar la probabilidad de que al lanzar al aire dos monedas, salgan:

1. *Dos caras.*
2. *Dos cruces.*
3. *Una cara y una cruz.*

17. En un sobre hay 20 papeletas, ocho llevan dibujado un coche las restantes son blancas. Hallar la probabilidad de extraer al menos una papeleta con el dibujo de un coche:

1. *Si se saca una papeleta.*
2. *Si se extraen dos papeletas.*
3. *Si se extraen tres papeletas.*

18. Los estudiantes A y B tienen respectivamente probabilidades 1/2 y 1/5 de suspender un examen. La probabilidad de que suspendan el examen simultáneamente es de 1/10. Determinar la probabilidad de que al menos uno de los dos estudiantes suspenda el examen.

19. Dos hermanos salen de caza. El primero mata un promedio de 2 piezas cada 5 disparos y el segundo una pieza cada 2 disparos. Si los dos disparan al mismo tiempo a una misma pieza, ¿cuál es la probabilidad de que la maten?

CAPITULO 6

Autores:

Lcda. Imelda Arias Montero, MSc.
UNIVERSIDAD ESTATAL DE MILAGRO

Ing. Juan Gonzalez Vaca, MSc.
ASESOR FINANCIERO

Ing. Marcos Guerrero Zambrano. Mgs
UNIVERSIDAD ESTATAL DE MILAGRO

"Nuestra generación ha entregado el alma a los contables y todas las pasiones que hoy nos conmueven se derivan de las estadísticas: para saber si somos felices, ahora se hacen encuestas." MANUEL VICENT

CAPITULO 6

Distribuciones Probabilísticas Discretas

OBJETIVOS

Definir los términos de distribución de probabilidad y variable aleatoria. Distinguir entre una distribución de probabilidad discreta y una continua. Calcular la media, la variancia y la desviación estándar de una distribución probabilística discreta. Describir las características y calcular las probabilidades usando la distribución de probabilidad binomial. Describir las características y calcular las probabilidades usando la distribución hipergeométrica. Describir las características y calcular las probabilidades usando la distribución de Poisson.

Desarrollo:

Variable Aleatoria

Una variable aleatoria es un valor numérico determinado por el resultado de un experimento(Huertas, Mellado, Trechera, & Patiño, 2015)**.**

EJEMPLO:

Considere un experimento aleatorio en el que se lanza tres veces una moneda. Sea X el número de caras. Sea C el resultado de obtener una cara y S el de obtener una sello.

Solución:

El espacio muestral para este experimento será:

S={SSS, SSC, SCS, SCC, CSS, CSC, CCS, CCC}

Entonces, los valores posibles de X (número de caras) son x = 0, 1, 2, 3.

- *El resultado "cero caras" ocurrió una vez.*
- *El resultado "una cara" ocurrió tres veces.*
- *El resultado "dos caras" ocurrió tres veces.*
- *El resultado "tres caras" ocurrió una vez.*

De la definición de variable aleatoria, la X definida en este experimento, es una variable aleatoria.

Distribución Probabilística Discreta:

Una distribución probabilística es la enumeración de todos los resultados de un experimento junto con las probabilidades asociadas(Batanero & Díaz, 2017)**.**

Numero de Caras	*Prob. Resultados*
0	*1/8*
1	*3/8*
2	*3/8*
3	*1/8*
	1

La probabilidad de un resultado siempre debe estar entre 0 y 1. La suma de todos los resultados mutuamente excluyentes siempre es 1.

Variable Aleatoria Discreta

Una variable aleatoria discreta es una variable que puede tomar sólo ciertos valores diferentes que son el resultado de la cuenta de alguna característica de interés(Aaron & Morales, 2016)**.**

EJEMPLO:

Sea "X" el número de caras obtenidas al lanzar 3 veces una moneda. Aquí los valores de "X" son x = 0, 1, 2, 3.

Variable Aleatoria Continua

Una variable aleatoria continua es una variable que puede tomar un número infinito de valores(Díaz-levicoy & Arteaga Cezón Dra Carmen Batanero Bernabeu, 2014)**.**

EJEMPLO:

La altura de un jugador de básquetbol o el tiempo que dura una siesta.

Media de una Distribución Probabilística Discreta

La media:

- ***Indica la ubicación central de los datos.***

- ***Es el promedio, a la larga, del valor de la variable aleatoria.***
- ***También se conoce como el valor esperado, E(x), en una distribución de probabilidad.***
- ***Es un promedio ponderado.***

La media se calcula con la fórmula:

$$\mu = E(X) = \sum_{i=1}^{n} \left(X_i * P\left(X_i \right) \right)$$

Donde:

μ Representa la media
P(x) Es la probabilidad de los diferentes resultados x.

EJEMPLO:

Con la distribución anterior calcule la media de la Distribución

$$\mu = E(X) = \sum_{i=1}^{n} \left(X_i * P\left(X_i \right) \right)$$

$$\mu = \sum_{i=1}^{4} \left(X_i * P\left(X_i \right) \right) = (0)*\left(\frac{1}{8}\right) + (1)*\left(\frac{3}{8}\right) + (2)*\left(\frac{3}{8}\right) + (3)*\left(\frac{1}{8}\right)$$

$$\mu = 0 + \frac{3}{8} + \frac{6}{8} + \frac{3}{8} = \frac{12}{8} = 1,5$$

Variancia de una distribución probabilística discreta

La variancia mide la cantidad de dispersión (variación) de una distribución. La variancia de una distribución discreta se denota por la letra griega (sigma cuadrada)(José & Castillo, 2016)**.**

$$\sigma^2$$

La variancia de una distribución de probabilidad discreta se calcula a partir de la fórmula:

$$\sigma^2 = \sum_{i=1}^{n} [\left(X_i - \mu \right)^2 * P(X_i)]$$

EJEMPLO:

$$\sigma^2 = \sum_{i=1}^{4}[(X_i - 1,5)^2 * P(X_i)]$$

$$\sigma^2 = [(0-1,5)^2 * \frac{1}{8}] + \left[(1-1,5)^2 * \frac{3}{8}\right] + \left[(2-1,5)^2 * \frac{3}{8}\right] + \left[(3-1,5)^2 * \frac{1}{8}\right]$$

$$\sigma^2 = 0,28 + 0,094 + 0,094 + 0,28 = 0,748$$

Desviación Estándar

La desviación estándar se obtiene tomando la raíz cuadrada de la variancia.

$$\sigma = \sqrt[2]{\sigma^2}$$

EJEMPLO:

$$\sigma = \sqrt[2]{\sigma^2} = \sqrt[2]{0,748} = 0,864$$

Distribución Probabilística Binomial

La distribución binomial tiene las siguientes características(Díaz-levicoy, Batanero, Arteaga, & Del, 2015)***:***

1. ***Se realiza una cantidad finita de Ensayos (n)***
2. ***Los ensayos son independientes***
3. ***Cada ensayo puede tener 2 posibles resultados (éxito, o fracaso)***
4. ***La probabilidad de que un ensayo sea éxito es constante (p)***
5. ***La probabilidad de que un ensayo sea fracaso es constante (1-p)***

Su forma de calcular es la siguiente:

$$P = \binom{n}{x} p^x (1-p)^{n-x}$$

EJEMPLO:

Un vendedor coloca una venta cada 10 intentos. Cierto día realiza 20 intentos. Calcule la probabilidad de:

a) ***Coloque exactamente 3 ventas.***
b) ***Coloque al menos 1 venta.***
c) ***Coloque no más de 3 ventas.***

Solución:

n = 20
x: cantidad de intentos exitosos
p = 1/10 = 0,1
1-p = 1 – 0,1= 0,9

a)

$$P(x=3)=\binom{20}{3}0,1^3*(1-0,1)^{20-3}=\binom{20}{3}0,001*(0,9)^{17}$$

$$P(x=3)=1140*0,001*0,17=0,19$$

b) $P(x\geq 1)=P(1)+P(2)+\cdots\cdots+P(20)$

Como podemos observar calcular esta probabilidad es muy larga, se recomienda usar la regla complemento para hacer el cálculo un poco más fácil.

$$P(x\geq 1)+P(x<1)=1$$
$$P(x<1)=P(0)$$
$$P(x<1)=\binom{20}{0}*(0,1)^0*(0,9)^{20-0}=\binom{20}{0}*(1)*(0,9)^{20}$$
$$P(x<1)=(1)*(1)*(0,12)=0,12$$
$$P(x\geq 1)=1-P(x<1)$$
$$P(x\geq 1)=1-0,12=0,88$$

Entonces podemos ver que la probabilidad que un vendedor coloque al menos 1 venta es de 0,88.

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

$$P(x<3)=0,12+\binom{20}{1}*(0,1)^{1}*(0,9)^{19}+\binom{20}{2}*(0,1)^{2}*(0,9)^{18}$$

c) $P(x<3)=0,12+((20)*0,1*0,14)+(190*0,01*0,15)$

$$P(x<3)=0,12+0,28+0,285$$

$$P(x<3)=0,685$$

Por lo tanto la probabilidad que el vendedor haga menos de tres ventas es 0,685

Población Finita:

Una población finita es una población que consiste en un número fijo de individuos, objetos o medidas conocidos.

EJEMPLOS:

a. El número de estudiantes en esta clase.
b. El número de automóviles en el estacionamiento.

Distribución Hipergeométrica

Use la distribución hipergeométrica para encontrar la probabilidad de un número específico de éxitos o fracasos si(Antonio Rustom J., 2012)***:***

- *La muestra se selecciona de una población finita sin reemplazo (recuerde que un criterio para la distribución binomial es que la probabilidad de éxito es la misma de un ensayo a otro).*

- *El tamaño de la muestra "n" es mayor que 5% del tamaño de la población "N".*

Su forma de calcular es la siguiente:

$$P(x)=\left(\frac{\binom{k}{x}*\binom{N-k}{n-x}}{\binom{N}{n}}\right)$$

Donde:

N	*Es el tamaño de la población*
n	*Es el tamaño de la muestra*
k	*Es el número de éxitos en la población*
x	*Es el número de éxitos en la muestra*

EJEMPLO:

La National Air Safety Board tiene una lista de 10 violaciones a la seguridad reportadas por ValueJet. Suponga que sólo 4 de ellas son en realidad violaciones y que el Safety Board sólo podrá investigar cinco de las violaciones. ¿Cuál es la probabilidad de que tres de las cinco violaciones seleccionadas al azar para investigarlas sean en realidad violaciones?

N = 10
n = 5
k = 4
x = 3

$$P(x=3)=\left(\frac{\binom{4}{3}*\binom{10-4}{5-3}}{\binom{10}{5}}\right)=\left(\frac{\binom{4}{3}*\binom{6}{2}}{\binom{10}{5}}\right)=\frac{(4)*(15)}{252}$$

$$P(x=3)=0,24$$

La probabilidad que se cometan 3 violaciones de seguridad es de 0,24

Distribución de Poisson

La distribución de probabilidades binomial se hace cada vez más sesgada a la derecha conforme la probabilidad de éxitos disminuye(Rey, María, & Díaz, 2010)**.**

La forma límite de la distribución binomial donde la probabilidad de éxito p es muy pequeña y n es grande se llama distribución de probabilidades de Poisson.

La distribución de Poisson se puede describir matemáticamente por la fórmula:

$$P(x)=\frac{\mu^{x}*e^{-\mu}}{x!}$$

Donde:

μ *Es la media aritmética del número de ocurrencias en un intervalo específico de tiempo*

e *Es la constante 2.71828*

x *Es el número de ocurrencias.*

El número medio de éxitos μ se puede determinar en situaciones binomiales por (n* p), donde "n" es el número de ensayos y "p" la probabilidad de éxito.

La varianza de la distribución de Poisson también es igual a (n* p).

EJEMPLO:

La Sylvania Urgent Care se especializa en el cuidado de lesiones menores, resfriados y gripe. En las horas de la tarde de 6-10 PM el número medio de llegadas es 4.0 por hora. ¿Cuál es la probabilidad de 4 llegadas en una hora?

$$P(x=4)=\frac{\mu^{x}*e^{-\mu}}{x!}=\frac{4^{4}*e^{-4}}{4!}=\frac{256*0{,}018}{4*3*2*1}$$

$$P(x=4)=\frac{4{,}608}{24}=0{,}192$$

AUTOEVALUACIÓN # 6

DESARROLLE LOS SIGUIENTES EJERCICIOS:

1. ***Una ruleta tiene 37 agujeros numerados del 0 al 36. ¿Cuál es la probabilidad de que la bola se introduzca en el agujero 19?***

2. ***Tengo en la mano 6 cartas con los números 1, 2, 3, 5, 6 y 7. Mi amigo toma una carta al azar.***

    a. ***¿Cuál es la probabilidad de que obtenga un número menor que 4?***
    b. ***¿Cuál es la probabilidad de que el número que obtenga sea divisible entre 2?***

3. ***De una urna con 8 bolas rojas, 5 bolas amarillas y 6 bolas verdes, se extrae una bola al azar. Halla la probabilidad de que:***

    a) Sea roja; b) Sea verde; c) Sea amarilla; d) No sea roja; e) No sea verde; f) No sea amarilla.

4. ***En una bolsa se han introducido 25 postales de ciudades españolas, de las cuales 3 son de Segovia:***

   a) ¿Cuál es la probabilidad de obtener una tarjeta de Segovia?

   b) ¿Cuál es la probabilidad de obtener una tarjeta que no sea de Segovia?

5. ***En unas pruebas de alcoholemia se ha observado que el 5% de los conductores controlados dan positivo en la prueba y que el 10% de los conductores controlados no llevan puesto el cinturón de seguridad. También se ha observado que las dos infracciones son independientes.***

   Un guardia de tráfico para cinco conductores al azar. Si tenemos en cuenta que el número de conductores es suficientemente importante como para estimar que la proporción de infractores no varía al hacer la selección.

   a. *Determinar la probabilidad de que exactamente tres conductores hayan cometido alguna de las dos infracciones.*

   b. *Determine la probabilidad de que al menos uno de los conductores controlados haya cometido alguna de las dos infracciones.*

6. ***Un laboratorio afirma que una droga causa efectos secundarios en una proporción de 3 de cada 100 pacientes. Para contrastar esta afirmación, otro laboratorio elige al azar a 5 pacientes a los que aplica la droga. ¿Cuál es la probabilidad de los siguientes sucesos?***

   a. *Ningún paciente tenga efectos secundarios.*

   b. *Al menos dos tengan efectos secundarios.*

   c. *¿Cuál es el número medio de pacientes que espera laboratorio que sufran efectos secundarios si elige 100 pacientes al azar?*

7. ***Se lanza una moneda cuatro veces. Calcular la probabilidad de que salgan más caras que cruces.***

8. ***Un agente de seguros vende pólizas a cinco personas de la misma edad y que disfrutan de buena salud. Según las tablas actuales, la probabilidad de que una persona en estas condiciones viva 30 años o más es 2/3. Hállese la probabilidad de que, transcurridos 30 años, vivan:***

   a. *Las cinco personas.*

   b. *Al menos tres personas.*

   c. *Exactamente dos personas.*

9. Si de seis a siete de la tarde se admite que un número de teléfono de cada cinco está comunicando, ¿cuál es la probabilidad de que, cuando se marquen 10 números de teléfono elegidos al azar, sólo comuniquen dos?

10. La probabilidad de que un hombre acierte en el blanco es 1/4. Si dispara 10 veces ¿cuál es la probabilidad de que acierte exactamente en tres ocasiones? ¿Cuál es la probabilidad de que acierte por lo menos en una ocasión?

11. En una urna hay 30 bolas, 10 rojas y el resto blancas. Se elige una bola al azar y se anota si es roja; el proceso se repite, devolviendo la bola, 10 veces. Calcular la media y la desviación típica.

CAPITULO 7

Autores:

Eco. Yecenia Escobar De la Cuadra. Mgs
UNIVERSIDAD ESTATAL DE MILAGRO

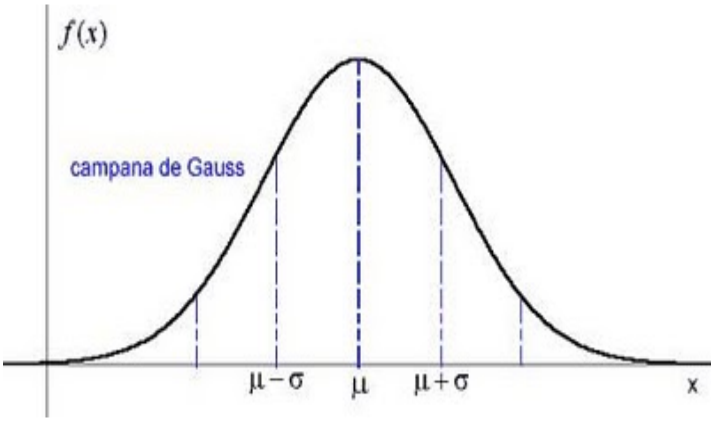
Lcdo. Bolívar R. Duchi Ortega Mg.
UNIVERSIDAD ECOTEC

Econ. Ernesto Rangel Luzuriaga. M.Sc. Ph.D (c)
UNIVERSIDAD TECNOLÓGICA EMPRESARIAL DE GUAYAQUIL UTEG

"La estadística es una ciencia que demuestra que, si mi vecino tiene dos coches y yo ninguno, los dos tenemos uno." GEORGE BERNARD SHAW

CAPITULO 7

Distribución Probabilística Normal



OBJETIVOS:

Enumerar las características de la distribución probabilística normal. Definir y calcular valores de *z*. Determinar la probabilidad que una observación esté entre dos puntos utilizando la distribución normal estándar. Determinar la probabilidad que una observación esté por arriba (o por abajo) de un valor determinado, utilizando la distribución normal estándar.

Desarrollo:

Características de la Distribución Probabilística Normal

La curva normal tiene forma de campana con un solo pico justo en el centro de la distribución.

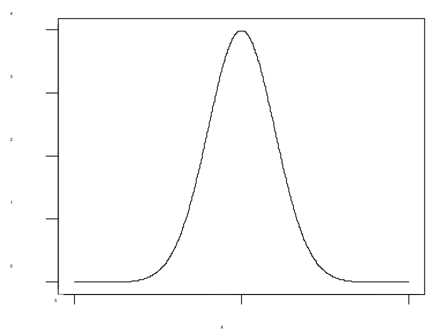
La media, mediana y moda de la distribución aritmética son iguales y se localizan en el pico.

La mitad del área bajo la curva está a la derecha del pico, y la otra mitad está a la izquierda.

La distribución normal es simétrica respecto a su media.

La distribución normal es asintótica - la curva se acerca cada vez más al eje x pero en realidad nunca llega a tocarlo.

- ***Característica de una Distribución Normal***



La media, mediana y moda son iguales

Distribución Normal Estándar:

Una distribución normal que tiene media igual a 0 y desviación estándar igual a 1 se denomina distribución normal estándar(Ramón Pérez Juste, 2012)**.**

Valor Z: ***Es la distancia entre un valor seleccionado, designado como X, y la población media μ, dividida entre la desviación estándar de la población σ.***(Gorgas, Nicolás, López, & Calvo, 2011)

$$Z = \frac{X - \mu}{\sigma}$$

EJEMPLO:

El ingreso mensual que una corporación grande ofrece a los graduados en MBA tiene una distribución normal con media de $2000 y desviación estándar de $200. ¿Cuál es el valor z para un ingreso de $2200? y ¿cuál para uno de $1700?

$$Z = \frac{X - \mu}{\sigma} = \frac{2200 - 2000}{200} = 1$$
$$Z = \frac{X - \mu}{\sigma} = \frac{1700 - 2000}{200} = -1,5$$

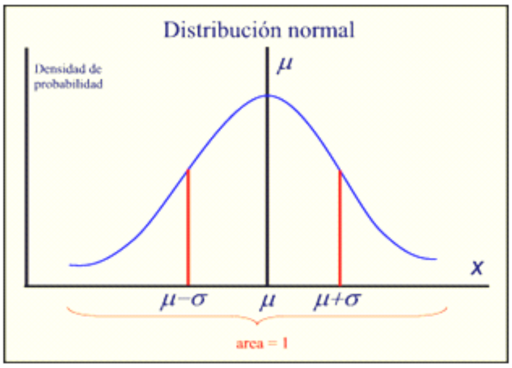
Esto quiere decir que un valor z igual a 1 indica que el valor de $2200 es mayor que la desviación estándar de la media de $2000, así como el valor z igual a -1.5 indica que el valor de $1700 es menor que la desviación estándar de la media de $2000.

Áreas Bajo la Curva Normal

Cerca de 68% del área bajo la curva normal está a menos de una desviación estándar respecto a la media.

Alrededor de 95% está a menos de dos desviaciones estándar de la media.

99.74% está a menos de tres desviaciones estándar de la media.



EJEMPLO:

La media de un grupo de ingresos semanales con distribución normal para un gran conjunto de gerentes de nivel medio, es de $1000, y la desviación estándar es de $100.

a) ***¿Calcule la probabilidad de que un ingreso semanal específico seleccionado al azar esté entre $790 y $1000?***

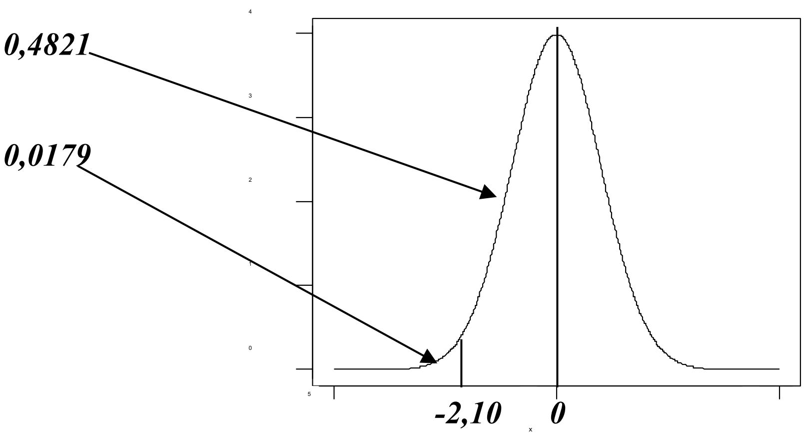
b) ***¿Calcule la probabilidad de que el ingreso sea menor de $790?***

Solución:

$$Z_1 = \frac{X-\mu}{\sigma} = \frac{790-1000}{100} = -2{,}10$$

a)
$$Z_2 = \frac{X-\mu}{\sigma} = \frac{1000-1000}{100} = 0$$

$$P(Z_1 \le Z \le Z_2) = P(-2{,}10 \le Z \le 0) = 0{,}4821$$



Como podemos observar el área bajo la curva comprendida entre los valores de Z -2,10 y 0 es de 0,4821, el mismo que se puede observar en la tabla que se adjunta.

b)

$$Z = \frac{X - \mu}{\sigma} = \frac{790 - 1000}{100} = -2,10$$

$$P(Z \le Z) = P(Z \le -2,10) = 0,0179$$

El área bajo la curva de la gráfica a la izquierda de la media, vale 0,5; y el área que se encuentra a la derecha de la medía también vale 0,5. Como el área bajo la curva entre $790 y $ 1000 es 0,4821, el área por debajo de $790 se determina restando 0,4821 de 0,5. De esta forma, 0,5 – 0,4821= 0,0179.

EJEMPLO:

El profesor Mann determinó que el promedio final en su curso de estadística tiene una distribución normal con media de 72 y desviación estándar de 5. Decidió asignar las calificaciones del curso de manera que 15% de los alumnos reciban una calificación de A. ¿Cuál es el promedio más bajo que un alumno puede tener para obtener una A?

Solución:

Sea X el promedio más bajo. Encuentre X de manera que P(X > X) = 0,15. El valor Z correspondiente es 1.04. Así se tiene (X - 72) / 5 = 1.04, o X = 77.2

$$Z = \frac{X - \mu}{\sigma}$$

$$1,04 = \frac{X - 72}{5}$$

$$(1,04) * 5 = X - 72$$

$$X = 5,20 + 72$$

$$X = 77,20$$

Entonces podemos observar que el promedio más bajo que un alumno puede tener para obtener una A es 77,20.

AUTOEVALUACIÓN # 7

DESARROLLE LOS SIGUIENTES EJERCICIOS:

1. ***Los resultados del examen de Matemáticas en un instituto arrojaron una puntuación media de 5,3 (puntuando sobre 10) y una desviación típica de 1,9 sobre un total de 355 alumnos presentados. Suponiendo una distribución normal en los resultados, determinar:***

   a) La probabilidad de seleccionar a un alumno con x mayor o igual a 8.
   b) Número de alumnos con x menor o igual a 5.
   c) Valor del primer cuartil.

2. ***En una universidad española que tiene en nómina a 725 profesores, la media de años de trabajo es de 7,3 y la desviación típica de 4,2 Suponiendo que se distribuyan normalmente:***

   a) ¿A cuántas unidades de desviación típica se encuentra un profesor que lleva 10 años en la universidad respecto a la media del colectivo?
   b) ¿Qué número de profesores lleva menos de tres años en la universidad?
   c) ¿Cuál será el menor número de años trabajados por los 100 profesores que llevan el máximo número de años en la universidad?

3. ***Si X es una variable aleatoria de una distribución N(μ, σ), hallar:***

$p(\mu - 3\sigma \leq X \leq \mu + 3\sigma)$

4. ***En una distribución normal de media 4 y desviación típica 2, calcular el valor de "a" para que:***

$$P(4 - a \leq x \leq 4 + a) = 0.5934$$

5. ***En una ciudad se estima que la temperatura máxima en el mes de junio sigue una distribución normal, con media 23° y desviación típica 5°. Calcular el número de días del mes en los que se espera alcanzar máximas entre 21° y 27°.***

6. ***La media de los pesos de 500 estudiantes de un colegio es 70 kg y la desviación típica 3 kg. Suponiendo que los pesos se distribuyen normalmente, hallar cuántos estudiantes pesan:***

    a. *Entre 60 kg y 75 kg.*
    b. *Más de 90 kg.*
    c. *Menos de 64 kg.*
    d. *64 kg.*
    e. *64 kg o menos.*

7. ***Se supone que los resultados de un examen siguen una distribución normal con media 78 y varianza 36. Se pide:***

    a. *¿Cuál es la probabilidad de que una persona que se presenta el examen obtenga una calificación superior a 72?*
    b. *Calcular la proporción de estudiantes que tienen puntuaciones que exceden por lo menos en cinco puntos de la puntuación que marca la frontera entre el Apto y el No-Apto (son declarados No-Aptos el 25% de los estudiantes que obtuvieron las puntuaciones más bajas).*
    c. *Si se sabe que la calificación de un estudiante es mayor que 72 ¿cuál es la probabilidad de que su calificación sea, de hecho, superior a 84?*

8. ***Tras un test de cultura general se observa que las puntuaciones obtenidas siguen una distribución una distribución N(65, 18). Se desea clasificar a los examinados en tres grupos (de baja cultura general, de cultura general aceptable, de excelente cultura general) de modo que hay en el primero un 20% la población, un 65% el segundo y un 15% en el tercero. ¿Cuáles han de ser las puntuaciones que marcan el paso de un grupo al otro?***

9. ***Varios test de inteligencia dieron una puntuación que sigue una ley normal con media 100 y desviación típica 15.***

   a. *Determinar el porcentaje de población que obtendría un coeficiente entre 95 y 110.*
   b. *¿Qué intervalo centrado en 100 contiene al 50% de la población?*
   c. *En una población de 2500 individuos ¿cuántos individuos se esperan que tengan un coeficiente superior a 125?*

10. ***En una ciudad una de cada tres familias posee teléfono. Si se eligen al azar 90 familias, calcular la probabilidad de que entre ellas haya por lo menos 30 tengan teléfono.***

11. En un examen tipo test de 200 preguntas de elección múltiple, cada pregunta tiene una respuesta correcta y una incorrecta. Se aprueba si se contesta a más de 110 respuestas correctas. Suponiendo que se contesta al azar, calcular la probabilidad de aprobar el examen.

12. Un estudio ha mostrado que, en un cierto barrio, el 60% de los hogares tienen al menos dos televisores Se elige al azar una muestra de 50 hogares en el citado barrio. Se pide:

a. *¿Cuál es la probabilidad de que al menos 20 de los citados hogares tengan cuando menos dos televisores?*

b. *¿Cuál es la probabilidad de que entre 35 y 40 hogares tengan cuando menos dos televisores?*

CAPITULO 8

Autores:

Ing. Salomón Arias Montero, Mgs.
UNIVERSIDAD TECNICA DE MACHALA

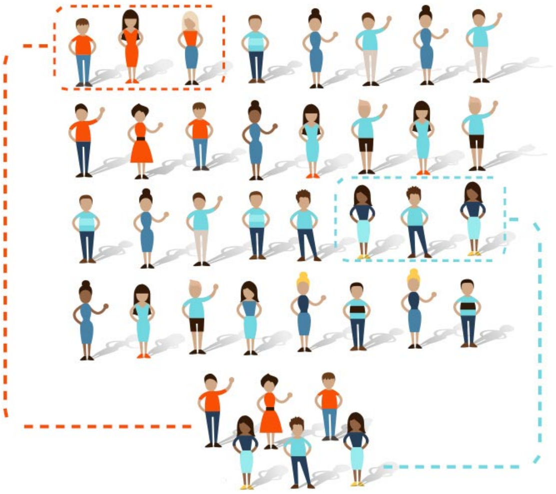
Ing. Zoila Noemi Merino Acosta, PhD
UNIVERSIDAD TÉCNICA DE BABAHOYO

Lcdo. Víctor Hugo Del Pozo R. Mgs.
UNIVERSIDAD DE GUAYAQUIL

"Las estadísticas me producen la misma impresión que las minifaldas: Muestran lo atractivo, ocultan lo vital." Doris Band

CAPITULO 8

Distribución Muestral



OBJETIVOS:
Enumerar las características de la Población y los parámetros que las describen. Estas poblaciones son discretas o continuas y empleamos la probabilidad como herramienta para determinar qué tan probables podrían ser ciertos resultados muéstrales. Este capítulo comprende muestreo y distribuciones muéstrales, que describen el comportamiento de estadísticas muéstrales.

Desarrollo:

En estadística, la distribución muestral es lo que resulta de considerar todas las muestras posibles que pueden ser tomadas de una población. Su estudio permite calcular la probabilidad que se tiene, dada una sola muestra, de acercarse al parámetro de la población.

Características de la Distribución Muestral

Población: *es el conjunto de todos los individuos sujetos a estudio.*

Muestra: *es el subconjunto finito de elementos seleccionados de la población.*

Para que las inferencias sean válidas, las muestras deben ser representativas de la población.

Muestreo: *procedimiento de obtención de una muestra. Podemos describir los siguientes tipos:*

Muestreo No Aleatorio: *la selección de los elementos muéstrales se realiza según el criterio del investigador. la muestra no es representativa de la población.*

Muestreo Aleatorio: *se selecciona de forma que cada elemento de la población tiene una probabilidad positiva de ser elegido.*

Definiendo la población

Antes de tomar una muestra, primero debemos definir la población a la que queremos generalizar nuestros resultados. La población de interés puede diferir para cada *estudio que realicemos. Podría ser la población de jugadores de fútbol profesional en los Ecuador o los votantes registrados en la Ciudad de Guayaquil. Podría también ser todos los estudiantes universitarios en una universidad determinada sea esta publica o privada, o todos los estudiantes de segundo año en esa institución. El punto debe ser claro; la muestra debe ser extraída de la población a la que desea generalizar: la población en la que está interesado en realizar la investigación.*

Conceptos de muestreo

Una muestra es un subconjunto de una población más grande. La población, o universo, es el conjunto de elementos sobre los cuales le gustaría sacar conclusiones. Si una muestra captura a toda la población, se llama censo.

Un muestreo aleatorio perfecto es difícil de obtener en la práctica. Si el tamaño N de la población es pequeño, se podría escribir cada uno de los N números en una ficha, mezclar las fichas y seleccionar una muestra de n fichas. Los números que seleccione corresponden a las n mediciones que aparecen en la muestra. Como este método no siempre es práctico, un método más sencillo y confiable utiliza números aleatorios, es decir, dígitos generados de modo que los valores de 0 a 9 se presentan al azar y con igual frecuencia.

Formas de error de investigación en muestreo

Cuando trabaja con datos de muestra, normalmente desea que su muestra proporcione una representación precisa de la población más amplia. Por ejemplo, si realiza una encuesta para apoyar la campaña política de un amigo, y sus datos muestran que el 64% de los encuestados desea que el distrito educativo ponga más énfasis en las habilidades de lectura en los estudiantes, desea que esta cifra del 64% sea un reflejo preciso del sentir entre la población en general.

Hay tres razones por las que podría no ser:

- *Error de no muestreo.*
- *Error de muestreo, también conocido como varianza muestral.*
- *Sesgo de muestra.*

El error de no muestreo

Consiste en todas las fuentes de error no relacionadas con el muestreo de los encuestados. Estos incluyen:
1.-Errores que ocurren porque las personas dan respuestas incorrectas
2.- Errores que ocurren debido a fallas o fluctuaciones en los instrumentos de medición física (encuestas, Formularios, Test, etc.)
3.- Errores que ocurren en la codificación o entrada de datos. Estas fuentes de error de no muestreo se controlan, en la medida de lo posible, utilizando buenos principios de medición y ejerciendo un buen control sobre la codificación, la entrada de datos y el análisis de datos.

El error de muestreo

Se refiere al hecho de que las muestras no siempre reflejan las verdaderas características de una población, incluso si las muestras se extraen utilizando procedimientos justos. Por ejemplo, tomamos en cuenta la teoría de los grandes números, en cualquier muestra dada se verá afectado por la variación aleatoria. De la misma manera, si pregunta a 100 personas si el distrito escolar debería recortar otros programas para poner más énfasis en la lectura, el porcentaje que responde afirmativamente se verá afectado por la variación aleatoria en la composición de la muestra.

Muestras Probabilísticas vs. Muestras no Probabilísticas

Hay dos tipos generales de muestras: muestras probabilísticas y muestras no probabilísticas. Las muestras de probabilidad, también llamadas muestras aleatorias, usan algún proceso aleatorio para seleccionar elementos de población para la muestra y dan a cada elemento de población una posibilidad conocida, no nula. Las muestras no probabilísticas no utilizan un proceso aleatorio; En cambio, los elementos son seleccionados por juicio o conveniencia.

Los dos tipos de muestras dependen de diferentes mecanismos para controlar el sesgo de selección. Las muestras probabilísticas dependen del azar. La idea es que, si las selecciones se hacen por casualidad, una muestra grande contendrá naturalmente una sección representativa de la población. Las muestras no probabilísticas, por el contrario, se basan en el juicio o en la nada. La idea es que las muestras pueden ser controladas de manera judicial para producir una sección representativa de la población, o que no se necesita una sección representativa para los fines de la investigación.

Con muestreo probabilístico, un investigador puede especificar la probabilidad de que un elemento (participante) sea Incluido en la muestra. Con el muestreo no probabilístico, no hay forma de estimar la probabilidad de un elemento está incluido en una muestra. Si el interés del investigador es generalizar los hallazgos derivado de la muestra a la población general, entonces el muestreo probabilístico es mucho más útil y preciso. Desafortunadamente, también es mucho más difícil y costoso que el muestreo no probabilístico.

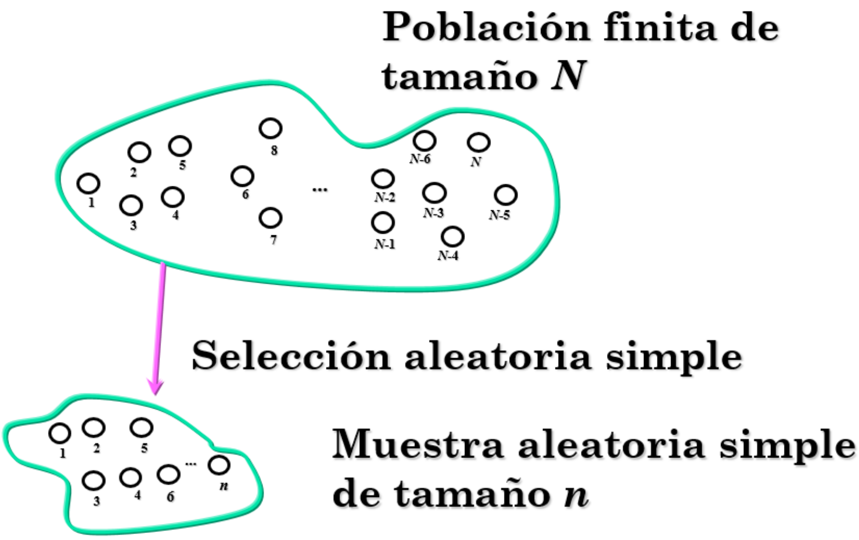
Tipos de muestras probabilísticas

Hay tres tipos amplios de muestras probabilísticas. El muestreo aleatorio simple es la versión básica del muestreo probabilístico. El muestreo estratificado y el muestreo por conglomerados son formas especiales de muestreo probabilístico que pueden mejorar la rentabilidad de la investigación en ciertas circunstancias.

Muestreo aleatorio simple

El muestreo probabilístico también se conoce como muestreo aleatorio o muestreo representativo. La palabra aleatorio describe el procedimiento utilizado para seleccionar elementos (participantes, autos, elementos de prueba) de una población.

Cuando se utiliza un muestreo aleatorio, cada elemento de la población tiene las mismas posibilidades de ser seleccionado (muestreo aleatorio simple) o una probabilidad conocida de ser seleccionado (muestreo aleatorio estratificado), una advertencia antes de comenzar nuestra descripción de muestreo aleatorio simple: el muestreo aleatorio es muy diferente de la asignación aleatoria. La asignación aleatoria describe el proceso de colocar a los participantes en diferentes grupos experimentales.



- ***Los pasos a seguir son:***
    - *Listar todos los elementos*
    - *Calcular el tamaño muestral necesario*
    - *Listado de números aleatorios*
    - *Extraer las unidades muéstrales del universo poblacional.*

- ***Lo positivo:***
    - *Sencillez.*
    - *Se basa en métodos probabilísticos*
    - *Sirve de base para otros métodos más complejos*

- ***Lo negativo:***
    - *Todas las unidades poblacionales han de ser listadas.*
    - *No unifica las distintas características de los elementos.*
    - *Algún grupo poblacional puede no ser representado*

La lista de población de la cual se pretende muestrear debe ser reciente y exhaustiva. Si no, los problemas pueden ocurrir. Por una Lista exhaustiva, queremos decir que todos los miembros de la población deben aparecer en la lista. Registro de votantes a veces se usan listas, directorios telefónicos, listas de propietarios y directorios escolares, pero estas listas pueden tener limitaciones. Deben estar actualizados y completos si las muestras elegidas de ellos deben ser verdaderamente representativo de la población. Además, dichas listas pueden proporcionar muestras muy sesgadas para algunos.

Para el muestreo aleatorio se utiliza las funciones especiales de los programas informáticos. Las listas de población ahora están disponibles como bases de datos de software (como Excel, SQL, MySql) o pueden ser importado a dicha base de datos. Muchos de estos programas de base de datos tienen una función para generar una serie de Números aleatorios y una función para seleccionar una muestra aleatoria de un rango de entradas en la base de datos. También existen numerosos sitios de internet pueden generar números aleatorios.

Ejemplo: Una población está formada por N 5 números: 3, 6, 9, 12, 15. Si una muestra aleatoria de tamaño n 3 se selecciona sin reemplazo, encuentre las distribuciones muéstrales para la media muestral x y la mediana m.

Solución Contiene cinco números distintos y cada uno es igualmente probable, con probabilidad p(x) 1/5. Con facilidad puede hallar la media poblacional y mediana como

$$\mu = \frac{3 + 6 + 9 + 12 + 15}{5} = 9 \quad \text{y} \quad M = 9$$

Hay 10 posibles muestras aleatorias de tamaño n 3 y cada una de ellas es igualmente probable, con probabilidad 1/10. Estas muestras, junto con los valores calculados de x y m para cada una.

Como la población de N 5 valores es simétrica cerca del valor de x 9, la media poblacional y la mediana son iguales a 9. Parecería razonable, por tanto, considerar usar ya sea x o m como estimador posible de M m 9. ¿Cuál estimador escogería usted? al usar m como estimador, estaría en un error de 9- 6=3 con probabilidad .3 o de 9 - 12=- 3 con probabilidad .3. Esto es, el error en estimación usando m sería 3 con probabilidad .6. Al usar x, no obstante, un error de 3 ocurriría con probabilidad de sólo .2. Sólo en estos casos, podría usarse x como estimador en preferencia sobre m.

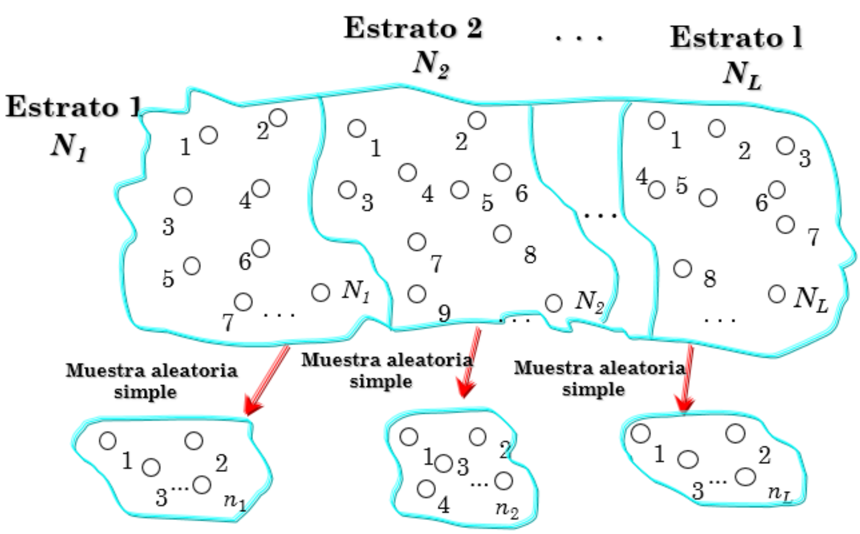
Muestreo aleatorio estratificado

Este procedimiento conocido como muestreo aleatorio estratificado es también una forma de muestreo probabilístico. Estratificar significa clasificar o separar a las personas en grupos de acuerdo con algunas características, como la posición, Rango, ingreso, educación, sexo, u origen étnico. Estas agrupaciones separadas se denominan subconjuntos o subgrupos Para una muestra aleatoria estratificada, la población se divide en grupos o estratos. Cualquiera la muestra se selecciona de cada estrato en función del porcentaje que cada subgrupo representa en la población.

Las muestras aleatorias estratificadas son generalmente más precisas para representar a la población que las muestras aleatorias simples. También requieren más esfuerzo, y hay un límite práctico para el número de estratos utilizados. Debido a que los participantes deben elegirse al azar de cada estrato, una lista completa de la población dentro de cada estrato debe ser construida. El muestreo estratificado se utiliza generalmente en dos diferentes formas. En uno, el interés principal está en la representatividad de la muestra para propósitos de comentar la población. En el otro, el foco de interés es la comparación entre y entre los estratos.

Las muestras estratificadas se utilizan a veces para optimizar las comparaciones de grupo. En esto caso, no nos preocupa la representación de la población total. En cambio, nuestro enfoque está en las comparaciones, involucrando dos o más estratos. Si los grupos involucrados en nuestras comparaciones están igualmente representados en la población, una sola muestra aleatoria podría ser utilizada. Cuando este no es el caso, un procedimiento diferente es necesario. Por ejemplo, si estuviéramos interesados en hacer comparaciones entre las etnias de personas blancas y negras, una muestra aleatoria simple de 100 personas puede incluir de 85 a 90 blancos y solo de 10 a 15 negros. Esta no es una muestra satisfactoria para hacer comparaciones. Con una muestra aleatoria estratificada, podríamos elija al azar 50 personas blancas y 50 personas negras y optimice así nuestra comparación. Cada vez que los estratos en lugar de la población es nuestro principal interés, podemos muestrear en diferentes proporciones de cada estrato.
Aunque el muestreo aleatorio es óptimo desde el punto de vista metodológico, no siempre es posible desde un punto de vista práctico. Veamos las ventajas y desventajas de varias otras muestras técnicas.

Población finita de tamaño $N_1 + N_2 + \ldots + N_L = N$



Muestra aleatoria estratificada de tamaño
$$n = n_1 + n_2 + \ldots + n_L$$

- ***Los pasos a seguir son:***

Se eligen unas características de la población diana y se realizan subgrupos de la misma a partir de éstos, intentando que sea un fiel reflejo del universo muestral.

Estrato: *Subconjunto de la población agrupado por una o más características.*

- ***Lo positivo:***
    - *Consigue una mayor precisión que el MAS y el MAST.*
    - *Se puede obtener información de cada estrato.*
    - *Es más beneficioso logística y administrativamente hablando.*
    -
- ***Lo negativo:***
    - *Es más complejo que el MAS.*
    - *Es más costoso económicamente y de recursos utilizados.*

En este proceso se divide la población en estratos o subconjuntos disjuntos y exhaustivos para la posterior extracción de una MAS en cada uno de ellos, con el objeto de conseguir una representación de cada uno de ellos. Se debe conseguir una homogeneidad intra-estrato y heterogeneidad inter-estrato.

Podemos elegir dos tipos de reparto al realizar los estratos:

Simple. - *Cada estrato contiene el mismo número de unidades muéstrales.*

$$n_e = n_{muestral} / L \qquad L = n^o \text{ de estratos}$$

Proporcional. - *Cada estrato es proporcional en la muestra a la proporción que le corresponde en la población diana.*

$$n_e = {N_e}/{N_{población}} \times n_{muestra}$$

N=600 n=60

Camas/Hospitales	N_e	Simple 60/3=20	Proporcional	
>1000	50	20	5	(50/600)x60
1000-500	200	20	20	(200/600)x60
<500	350	20	35	(350/600)x60
	N=600	n=60	n=60	

Simple

L=nº de estratos=3

$$n_e = n_{muestral}/L = 60/3 = 20$$

Proporcional

$$n_e = N_e/N_{población} \times n_{muestra}$$

La siguiente tabla resume el tamaño de muestra requerido al emplear cada técnica, en función del error máximo que estamos dispuestos a aceptar y de las características del propio universo, que consideraremos de tamaño infinito (si fuese finito, debe aplicarse un factor de corrección).

	Tamaño de muestra para estimar una proporción	Tamaño de muestra para estimar una media
Muestreo aleatorio simple	$\frac{Z^2 p(1-p)}{e^2}$	$\frac{Z^2\sigma^2}{e^2}$
Muestreo estratificado proporcional	$\frac{Z^2 \sum_{h=1}^{L} W_h p_h (1-p_h)}{e^2}$	$\frac{Z^2 \sum_{h=1}^{L} W_h \sigma_h^2}{e^2}$
Muestreo estratificado óptimo	$\frac{Z^2 (\sum_{h=1}^{L} W_h \sqrt{p_h(1-p_h)})^2}{e^2}$	$\frac{Z^2 (\sum_{h=1}^{L} W_h \sigma_h)^2}{e^2}$

Z = *Es la desviación del valor medio que aceptamos para lograr el nivel de confianza deseado. En función del nivel de confianza que busquemos, usaremos un valor determinado que viene dado por la forma que tiene la distribución de Gauss. Los valores más frecuentes son:*

Nivel de confianza 90% -> Z=1,645

Nivel de confianza 95% -> Z=1,96

Nivel de confianza 99% -> Z=2,575

Ejemplo: Supongamos que se trata de elegir una muestra de n = 1.000 personas, de una población de 68.000 habitantes, en la que hay 37.400 mujeres y 30.600 hombres, los tamaños de los estratos muéstrales serán:

Estratos	Nº de Personas en la Población	% en la Población	Nº de elementos en cada estrato
N_1: Mujeres	37.400	55.0	n_1: 550
N_2: Hombres	30.600	45.0	n_2: 450
N : Total	68.000	100.0	n : 1.000

Es decir: $n_j = \frac{Nj}{N} \times n$ donde **j** = 1, 2 = número de estratos.

Muestreo por Conglomerados

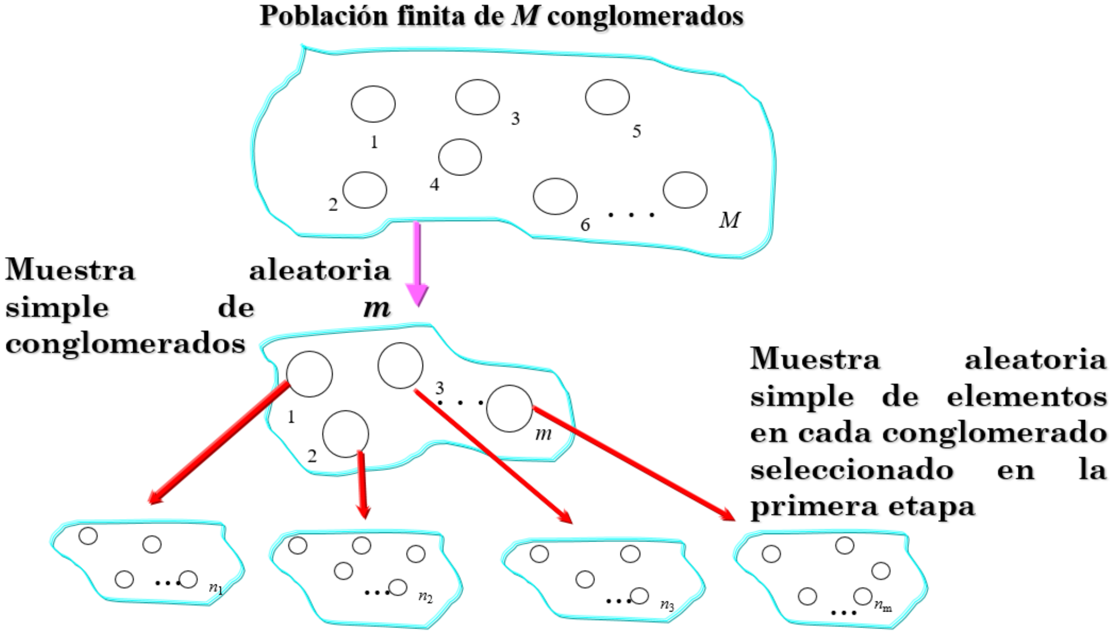
El muestreo de conglomerados se refiere a un tipo de método de muestreo. Con el muestreo de conglomerados, el investigador divide a la población en grupos separados, llamados conglomerados o cluster. Luego, se selecciona una muestra aleatoria simple de grupos de la población. El investigador lleva a cabo su análisis de los datos de los grupos muestreados.

En comparación con el muestreo aleatorio simple y el muestreo estratificado, el muestreo por conglomerados tiene ventajas y desventajas. Por ejemplo, dados los tamaños de muestra iguales, el muestreo por grupos generalmente proporciona menos precisión que el muestreo aleatorio simple o el muestreo estratificado. Por otro lado, si los costos de viaje entre grupos son altos, el muestreo de grupos puede ser más rentable que los otros métodos.

Una muestra de grupo es una muestra de probabilidad en la que cada unidad de muestreo es una colección o un grupo de elementos.
Es útil cuando:

- *No está disponible una lista de elementos de la población, pero es fácil obtener una lista de grupos.*

- *El costo de obtener observaciones aumenta a medida que la distancia que separa los elementos.*



- *Primero seleccionamos una muestra de m conglomerados usando muestreo aleatorio simple.*
- *De cada conglomerado seleccionado obtenemos un marco de las Ni unidades i=1,2,...,m.*
- *Seleccionamos una muestra aleatoria de tamaño ni donde i=1,2,...,m de cada uno de los conglomerados.*

Así la muestra total será de tamaño n=n1+n2+...nm. La forma de determinar ni para cada conglomerado puede ser hecha por separado o bien determinar n y después distribuirla sobre los m conglomerados.

Aquí hay que determinar n y m.

El propósito de la teoría del muestreo es hacer que el muestreo sea más eficiente. Intenta desarrollar métodos de selección de muestras y de estimaciones que proporcionen, al menor costo posible, estimaciones que sean lo suficientemente precisas para nuestro propósito. Este principio de precisión especificada al costo mínimo se repite repetidamente en la presentación de la teoría.

Para aplicar este principio, debemos poder predecir, para cualquier procedimiento de muestreo que esté bajo consideración, la precisión y el costo que se espera. En lo que se refiere a la precisión, no podemos predecir exactamente cuán grande será el error presente en una estimación en una situación específica, ya que esto requeriría un conocimiento del valor real para la población. En cambio, la precisión de un procedimiento de muestreo se juzga examinando la distribución de frecuencia que se genera para la estimación 'si el procedimiento se aplica una y otra vez a la misma población. Esta es, por supuesto, la técnica estándar por la cual se juzga la precisión Teoría estadística.

Se introduce una simplificación adicional. Con muestras de los tamaños que son comunes en la práctica, a menudo hay una buena razón para suponer que las *estimaciones de la muestra se distribuyen de manera aproximadamente normal. En consecuencia, la varianza muestral de la estimación se utiliza para proporcionar, en términos inversos, una medida de su precisión. Una parte considerable de la teoría se ocupa del cálculo de fórmulas para las varianzas muéstrales de estimaciones obtenidas mediante diversos procedimientos.*

Ejemplo: Se tienen 3000 elementos divididos en 30 conglomerados de diferentes tamaños. Si se toma una muestra de 20 conglomerados, cada uno tiene 20/30 oportunidades de ser seleccionado.

Si seleccionamos 1/3 de los elementos dentro de cada conglomerado, la probabilidad total de cada elemento de ser seleccionado es ***20/30 x 1/3 = 0,22***

De esta manera los conglomerados más grandes tienen más elementos seleccionados y el tamaño muestral es: ***0,22 x 3000 = 660 elementos***

La probabilidad de que un elemento sea seleccionado es igual a la fracción muestral
n/N = 660 / 3000 = 0,22

AUTOEVALUACIÓN # 8

REALICE LOS SIGUIENTES EJERCICIOS:

1.- Los pesos de una población en la ciudad de Milagro siguen una distribución desconocida con una media de 74 Kg. y una desviación típica de 4.

Halla la esperanza matemática (media) y la varianza del estadístico "Media aritmética" ($\overline{X}$), para una muestra aleatoria simple de tamaño 9 y otra de tamaño 30

2.-Los pesos de una población en la ciudad de Milagro siguen una distribución normal de media 75 Kg. y desviación típica 15 Kg. Se extrae una muestra aleatoria simple de tamaño 10.

a) Halla la probabilidad de que la media muestral se encuentre entre 72 y 80 Kg.

b) Halla la probabilidad de que la media de la muestra sea mayor que 77

3.-El peso de una población en la ciudad de Milagro siguen una distribución desconocida de media 80 Kg. y desviación típica 14 Kg.

a) Se extrae una muestra aleatoria simple de tamaño 49. Halla la probabilidad de que la media muestral sea mayor que 83 Kg.

b) Con la misma muestra del apartado anterior, halla la probabilidad de que la media de la muestra se encuentre entre 77 y 83 Kg.

c) Halla el tamaño que ha de tener la muestra para que la probabilidad de que la diferencia entre la media muestral y la media poblacional sea menor que 3 sea igual a 0,95 (es decir, para que $P(\left|\overline{X} - 80\right| < 3) = 0{,}95$)

4.-Supongamos que el cociente intelectual de la población se distribuye normalmente con media 100 y desviación típica 15. Se extrae una muestra aleatoria simple.

a) Si el tamaño de la muestra es 51, halla la probabilidad de que la varianza muestral sea mayor que 297,79.

b) Halla la probabilidad de que la varianza de la muestra del apartado anterior esté entre 297,79 y 335,96

5.-Supongamos que el 30% de los hogares Milagreños tiene conexión a Internet. Se extrae una muestra aleatoria simple de 400 hogares,

a) Halla la probabilidad de que la proporción de los que tienen Internet en dicha muestra esté entre el 28% y el 35%.

b) Halla la probabilidad de que el porcentaje de viviendas en la muestra con Internet supere el 33%.

c) Halla cuál ha de ser el tamaño de la muestra para que exista una probabilidad de 0,99 de que la proporción muestral esté entre 0,28 y 0,32 (es decir, para que el 99% de las muestras tengan su media entre 0,28 y 0,32.

6- Un analista político desea seleccionar una muestra de n 20 personas de entre una población de 2000. Use la tabla numérica aleatoria para identificar las personas a ser incluidas en la muestra.

7-Una población contiene 50 mil votantes. Use la tabla numérica aleatoria para identificar los votantes a ser incluidos en una muestra aleatoria de n 15.

8-Una pequeña ciudad contiene 20 mil votantes. Use la tabla numérica aleatoria para identificar los votantes a ser incluidos en una muestra aleatoria de n 15.

9-Cada décima persona Una muestra aleatoria de opinión pública, en una pequeña ciudad, se obtuvo al seleccionar cada décima persona que pasara por la esquina de mayor movimiento en el centro de la ciudad. ¿Esta muestra tendrá las características de una muestra aleatoria seleccionada de entre los ciudadanos? Explique.

10-Parques y recreación Se envió por correo un cuestionario a mil votantes municipales registrados seleccionados al azar. Sólo 500 cuestionarios fueron devueltos y, de éstos, 360 que contestaron se oponían rotundamente a un recargo propuesto para sostener al Departamento de Parques y Recreación de la ciudad. ¿Está usted dispuesto a aceptar la cifra de 72% como estimación válida del porcentaje en la ciudad que se oponen al recargo? ¿Por qué sí o por qué no

RECURSOS ONLINE:

La Sociedad Ecuatoriana de Estadística (SEE) *es una organización de la Sociedad Civil, que desde el año 1998, surge como producto de la necesidad de crear una institución nacional que organice, promueva, y apoye la investigación estadística en el país, así como la formación de profesionales de otras áreas que tenían interés o utilizaban de manera frecuente las técnicas y herramientas estadísticas. Desde ese entonces, la SEE ha promovido la gestión del conocimiento a través de la utilización de la estadística, así como fomentando las mejores prácticas y actualización permanente de los conocimientos estadísticos de quienes aplican esta herramienta para orientar la toma de decisiones en los ámbitos público y privado.*
http://see-ec.org/node/67
https://es-la.facebook.com/sococuest/

Escuela de R-Estadística

En el 2013 tuvo lugar la primera Escuela de R-Estadística siendo un éxito rotundo al difundir el uso de las técnicas estadísticas mediante el uso de software libre, desde su primera edición se ha caracterizado por contar con un staff de facilitadores del más alto nivel. Ahora en su 3ra edición nuestro principal interés es mantener el mismo estándar de ediciones anteriores y reiterar nuestro compromiso de brindar a la comunidad un servicio de calidad que contribuya en sus diversos campos profesionales.

http://rusersgroup.com/blog/escuelar3/

La Sociedad Ecuatoriana de Matemática, SEdeM, *es una organización científica no gubernamental, sin fines de lucro, conformada por personas afines al ámbito de la matemática, y cuyo objetivo es promover el desarrollo de esta ciencia en Ecuador. La SEdeM fue fundada en Quito en 1967 y, desde entonces, ha organizado y apoyado congresos nacionales e internacionales, encuentros y olimpiadas matemáticas, entre otros eventos. Actualmente, la SEdeM es reconocida como miembro asociado de la International Mathematical Union, IMU, y de la Unión Matemática de Latinoamérica y el Caribe, UMALCA.*

http://www.sedem.org.ec/index.html

19 herramientas de análisis de datos para convertirte en un ninja del analytics

https://conceptosclaros.com/19-herramientas-de-analisis-de-datos/

Software estadístico

PSPP

PSPP es una aplicación de software libre para el análisis de datos. Se presenta en modo gráfico y está escrita en lenguaje de programación C. Desarrollada por GNU Operating System. Proporciona funcionalidades básicas como: frecuencias, tablas cruzadas, comparación de media, regresión lineal, fiabilidad, reordenamiento de datos, pruebas no paramétricas, factor de análisis, entre otras características. Los formatos de salida pueden ser: en ASCII, PDF, PostScript o HTML, así como algunos gráficos estadísticos: histogramas, gráfico circulares y gráficos de distribución normal.

http://www.gnu.org/software/pspp/

R

R es una implementación de software libre del lenguaje S pero con soporte de alcance estático. Se trata de uno de los lenguajes más utilizados en investigación por la comunidad estadística, siendo además muy popular en el campo de la minería de datos, la investigación biomédica, la bioinformática y las matemáticas financieras. Además permite la posibilidad de cargar diferentes bibliotecas o paquetes con funcionalidades de cálculo y gráficas. Desarrollado por R Development Core Team.

https://www.r-project.org/

BioStatFLOSS 2.0

BioStatFLOSS 2.0, una recopilación de software de libre acceso para la realización de estudios bioestadísticos, epidemiológicos y de salud en general. Esta plataforma ha sido desarrollada por el Servicio de Epidemiología de la Dirección Xeral de Innovación e Xestión da Saúde Pública de Galicia.

https://www.sergas.es/Saude-publica/BioStatFLOSS

InfoStat

Cubre tanto las necesidades elementales para la obtención de estadísticas descriptivas y gráficos para el análisis exploratorio, como métodos avanzados de modelación estadística y análisis multivariado. Una de sus fortalezas es la sencillez de su interfaz combinada con capacidades profesionales para el análisis estadístico y el manejo de datos. Debido al origen universitario, el programa tiene muchas facilidades para la enseñanza de la estadística que no son fáciles en encontrar en otros programas *similares. La versión en español es muy valorada por los usuarios, especialmente por los estudiantes. Una propiedad casi única entre el software estadístico es la habilidad de InfoStat se conectarse con R, una plataforma de desarrollo de algoritmos estadísticos de dominio público de gran crecimiento. InfoStat de conecta con R de dos maneras: mediante un intérprete integrado que permite ejecutar script de R sin salir del ambiente de trabajo de InfoStat y mediante el desarrollo de aplicaciones utilizando el motor de cálculo de R pero con la interfaz amigable que los usuarios esperan. Ese es el caso del la inclusión de modelos lineales mixtos y generalizados mixtos en InfoStat. Estos han sido siempre modelos difíciles de especificar por su complejidad, pero la interfaz lograda en InfoStat nos ha permitido incorporar sus contenidos en cursos de postgrado y capacitaciones a técnicos de empresas que de otra manera hubieran sido imposibles de abordar.*

http://www.infostat.com.ar/

SPSS

- Programa estadístico para las ciencias sociales: SPSS

SPAD

- SPAD: Système Pour l'Analyse des Données, software estadístico francés
- Manual de introducción al SPAD del Servei d'Estadística de la UAB (versión 4.5)
- Manual de SPAD (versión 3)

SAS

- SAS: Statistical Analysis System, programari estadístic
- Manuals d'introducció a SAS del Servei d'Estadística de la UAB)

Altres

- R: Software estadístico libre
- Stata: Data Analysis and Statistical Software, software estadístico
- Statsoft STATISTICA Data Analysis Software and Services

Software de muestreo

- SOTAM Sistema de optimización para tamaños muestrales
- POSDEM. Software para escoger entre diferentes planes de muestreo probabilístico
- Ene: programa para el cálcul0 del tamaño muestral

Análisis de datos y Estadística

http://pagines.uab.cat/plopez/content/estad%C3%ADstica

Enlaces a recursos

- Servei d'Estadística de la UAB
- International Statistical Institute | Glosario de término de estadística: Glossary (català | castellano)
- estadistico.com Portal de servicios y recursos de estadística
- StatLib. Data, Software and News from the Statistics Community. Carnegie Mellon University.
- KDnuggets. Portal de Datamining
- Material Docent d'Estadística. E. Garcia Berthou
- Enlaces del Grupo de Investigación sobre Educación Estadística de la Universidad de Granada
- Links. International Association for Statistical Education
- Statistical Books, Manuals and Journals. StatPages.net
- Statistics. The World Wide Web Virtual Library
- Statsci. Recursos diversos y enlaces de estadística
- Web Pages that Perform Statistical Calculations. StatPages.net
- Radical Statistics Group
- Mactutor The MacTutor History of Mathematics archive University of St. Andrews Scotland
- History of Statistics Materials for the History of Statistics. York University

Cursos y manuales en línea

- Aula Virtual de Bioestadística. Facultad de Biología, Universidad Complutense de Madrid
- Comprensión y uso de la estadística Fernando Valdes. Universidad Romulo Gallegos
- Curso de Estadística AulaFacil.com. Introducción a la estadística: descriptiva y probabilidades
- Curso de Bioestadística Universidad de Málaga, Facultad de Medicina
- Statsoft. Statistica Electronic Statistics Textbook
- Statistical Computing (UCLA)
- Social Research Methods Knowledge Base William M. Trochim, Cornell University. El proceso de investigación
- Rice Virtual Lab in Statistics: HyperStat Online Textbook | Simulations/Demonstrations
- Emtech. Recursos de estadística para la enseñanza
- Calculadora del test de khi-quadrat 1 | 2

Datos estadísticos: Bases de Datos

Revistas estadísticas

- Índice. Revista de Estadística y Sociedad de l'INE
- Revista Estadística Española
- SORT Revista d'Estadística i Investigació Operativa de l'Idescat
- Econometrica
- Journals of the Royal Statistical Society. Series B (Statistical Methodology)
- Journal of the American Statistical Association
- The American Statistician
- Journal of the Royal Statistical Society. Series A (Statistics in Society)
- Annals of Statistics
- Biometrika
- Biometrics
- Statistics and Computing
- Oxford Bulletin of Economics & Statistics

Recursos en R para la docencia de Probabilidad y Estadística
Angelo Santana & Carmen Nieves Hernández,

http://www.dma.ulpgc.es/profesores/personal/stat/cursoR4ULPGC/17-recursos.html

Webs

- StatsTeachR: Recursos hechos por y para docentes de estadística y cálculo de probabilidades.
- IPSUR: Introduction to Statistics and Probability using R
- Rpubs.com: documentos elaborados con R Markdown
- Shiny: Herramienta para el desarrollo de aplicaciones web con R
- Show Me Shiny: galería de aplicaciones Web desarrolladas con Shiny
- swirl: paquete para enseñar estadística y R simultáneamente de manera interactiva
- Project Mosaic: paquete para enseñar estadística, modelos y cálculo con R.
- The Statistical Sleuth (*el detective estadístico*)
- Creación de cursos de R con DataCamp
- Advanced R Programming

Herramientas interactivas para el aprendizaje *online* de R

- TryR
- Datacamp

Libros

- Learning Statistics with R: Libro de Daniel Navarro del *Computational cognitive Science Lab* en la *Universidad de Adelaida*(Australia)
- The elements of Statistical Learning: Libro de T. Hastie, R. Tibshirani y J. Friedman, de la Universidad de Stanford.
- An introduction to Statistical Learning: de G. James, D. Witten, T. Hastie y R. Tibshirani
- Free statistics e-books for download
- Statistics with R: libro en pdf y html de Vincent Zoonekynd.
- Rtips Paul E. Johnson
- Three free books on R for Statistics: Análisis multivariante, bioestadística, series temporales
- Libros y documentos en CRAN
- Introduction to data science, Jeffrey Stanton, Syracuse University
- Introduction to Statistical Thinking, with R, without Calculus, Benjamin Yakir, Dpt. of Statistics, Hebrew University of Jerusalem.
- Multivariate statistics with R, Paul j. Hewson
- R programming Wiki Book
- A littel book of R for bioinformatics

Tutoriales

- R tutorial
- Michael Clark, Center of Social Research, Univ. of Notre Dame, Indiana. En esta web se ofrecen tutoriales de R (básico y avanzado), así como de modelos aditivos generalizados, estadística bayesiana y Machine Learning, con orientación a las ciencias sociales.

Blogs

- Análisis y Decisión: blog de R. Vaquerizo
- Econometría, estadística, clases y R: blog de G.R. Serrano
- Datanalytics: blog de C. Gil Bellosta

Cursos de R y Análisis de Datos

- codeSchool
- Coursera
- edX
- Udacity

BIBLIOGRAFÍA

1. *CHRIS, Leach. Fundamentos de estadística. Editorial Limusa. Mexico.1998*
2. *NIETO DE ALBA, Ubaldo. Introducción a la estadística. Editorial Aguilar. Madrid. 2001*
3. *MONTGOMERY, Douglas y RUNGER, George. Probabilidad y Estadística. McGraw-Hill. Arizona.2003*
4. *R.E. Walpole y R.H. Myers. Probabilidad y estadística. McGraw-Hill,2007*
5. *Douglas C. Montgomery y George C. Runger Probabilidad y Estadística aplicadas a la Ingeniería. 2006*
6. *Anderson D., Sweeney D., Williams T. Estadística para la administración y economía. Décima edición. Cengage Learning. 2008*
7. *Berenson M., Levine D., Krehbiel T. Estadística para administración. Segunda edición. Prentice Hall. 2000*
8. *Blair C., Taylor R. Bioestadística. Peason. Prentice Hall. 2008*
9. *Daniel W. Bioestadística. Cuarta edición. Limusa Wiley. 2006*
10. *Devore J. Probabilidad y Estadística para ingeniería y ciencias. Séptima edición. Cengage Learning. 2008*
11. *Johnson R. Probabilidad y Estadística para ingenieros. Octava edición. Pearson. 2012*
12. *Kuehl R., Diseño de experimentos. Segunda edición. Thomson Learning. 2001*

13. *Meyer P. Probabilidad y Aplicaciones estadísticas. Edición revisada. Addison Wesley Logman. 1998*

14. *Montgomery D., Diseño y análisis de experimentos. Segunda edición. Limusa Wiley. 2006*

15. *Newbold P., Carlson W., Thorne B. Estadística para la administración y economía. Sexta edición. Pearson. Prentice Hall. 2008*

16. *Aaron, G., & Morales, E. (2016). ESTADISTÍCA Y PROBABILIDADES. Retrieved from http://www.x.edu.uy/inet/EstadisticayProbabilidad.pdf*
17. *Anderson, D. R. (David R., & Peralta Rosales, L. (2014). Estadísica para negocios y economía. Retrieved from http://latinoamerica.cengage.com/ls/estadistica-para-negocios-y-economia/*
18. *Antonio Rustom J. (2012). ESTADÍSTICA DESCRIPTIVA, PROBABILIDAD E INFERENCIA. La Pintana, Santiago, Chile. . Retrieved from http://repositorio.uchile.cl/bitstream/handle/2250/120284/Rustom_Antonio_Estadistica_descriptiva.pdf?sequence=1*
19. *Batanero, C., & Díaz, C. (2017). ESTADÍSTICA CON PROYECTOS. Retrieved from https://www.ugr.es/~batanero/pages/ARTICULOS/Libroproyectos.pdf*
20. *Boslaugh, S., & Watters, P. A. (2008). Statistics in a nutshell. O'Reilly.*
21. *Brown, R. D. (2018). Business case analysis with R : simulation tutorials to support complex business decisions.*
22. *Bruce, P. C., & Bruce, A. (2017). Practical statistics for data scientists : 50 essential concepts.*
23. *Crawley, M. J. (2014). Statistics : an introduction using R.*
24. *Devore, J. L. (2016). Probabilidad y estadística para ingeniería y ciencias. Cengage Learning. Retrieved from http://latinoamerica.cengage.com/ls/probabilidad-y-estadistica-para-ingenieria-y-ciencias/*
25. *Díaz-levicoy, D., & Arteaga Cezón Dra Carmen Batanero Bernabeu, P. (2014). UN ESTUDIO EMPÍRICO DE LOS GRÁFICOS ESTADÍSTICOS EN LIBROS*

DE TEXTO DE EDUCACIÓN PRIMARIA ESPAÑOLA. Retrieved from http://funes.uniandes.edu.co/6385/1/TFMDanilo.pdf

26. *Díaz-levicoy, D., Batanero, C., Arteaga, P., & Del, M. (2015). ANÁLISIS DE LOS GRÁFICOS ESTADÍSTICOS PRESENTADOS EN LIBROS DE TEXTO DE EDUCACIÓN PRIMARIA CHILENA 1 ANALYSIS OF STATISTICAL GRAPHICS INTRODUCED IN CHILEAN PRIMARY EDUCATION TEXTBOOKS. Retrieved from http://funes.uniandes.edu.co/8332/1/23446-68077-1-PB.pdf*
27. *Elorza Pérez Tejada, H. (2015). Estadistica para las ciencias sociales, del comportamiento y de la salud. Retrieved from https://www.uv.mx/rmipe/files/2015/09/Estadistica-para-las-ciencias-sociales-del-comportamiento-y-de-la-salud.pdf*
28. *Fischetti, T. (2015). Data analysis with R : load, wrangle, and analyze your data using the world's most powerful statistical programming language.*
29. *GARG, H. (2018). MASTERING EXPLORATORY ANALYSIS WITH PANDAS : build an end-to-end data analysis workflow with python. PACKT PUBLISHING LIMITED.*
30. *Gorgas, J., Nicolás, G., López, C., & Calvo, J. Z. (2011). ESTADISTICA BASICA PARA ESTUDIANTES DE CIENCIAS. Madrid. Retrieved from http://webs.ucm.es/info/Astrof/users/jaz/ESTADISTICA/libro_GCZ2009.pdf*
31. *Huertas, R. R., Mellado, A. G., Trechera, L. M., & Patiño, S. F. (2015). ESTADÍSTICA INDUSTRIAL (Temas de estadística para Ingenieros). Retrieved from http://taylor.us.es/componentes/mcalle/MetodosEstadisticosIngenieria/Libros/Libro.EstadisticaIndustrialParaIngenieros.pdf*
32. *I. Espejo Miranda, F. Fern´andez Palac´ın, M. A. L´opez S´anchez, M. Mu˜noz M´arquez, A. M. Rodr´ıguez Ch´ıa, A. S´anchez Navas, & C. Valero Franco. (2014). Estadıstica Descriptiva y Probabilidad. (Servicio de Publicaciones de la Universidad de Cadiz, Ed.). Cadiz. Retrieved from http://www.uca.es/publicaciones*
33. *José, A., & Castillo, S. (2016). Apuntes de Estadística para Ingenieros. Retrieved from http://www4.ujaen.es/~ajsaez/recursos/EstadisticaIngenieros.pdf*

34. *Laverty, W. H., & Kelly, I. W. (1995). An Examination of Male and Female Monthly Employment Rates over Time in Canada and the United States Using Hidden Markov Probability Models. Open Journal of Statistics, 8, 837–845. https://doi.org/10.4236/ojs.2018.85055*

35. *Mendenhall, W., Beaver, R. J., Beaver, B. M., Pérez Tejada, H. E., Velázquez Arellano, J. A., & Bañuelos Saucedo, A. L. (2016). Probabilidad y estadística : para las ciencias sociales del comportamiento y la salud. Retrieved from http://latinoamerica.cengage.com/ls/probabilidad-y-estadistica-para-las-ciencias-sociales-y-de-la-salud/*

36. *Mendenhall, W., Beaver, R. J., Beaver, B. M., & Velázquez Arellano, J. A. (2017). Introducción a la probabilidad y estadística. Retrieved from http://latinoamerica.cengage.com/ls/introduccion-a-la-probabilidad-y-estadistica-2/*

37. *Ramón Pérez Juste. (2012). Estadista aplicada a las Ciencias Sociales. Madrid. Retrieved from http://www.intecca.uned.es/portalavip/grabacion.php?ID_Grabacion=61218&ID_Sala=65481&hashData=0745b3f95043cc887d78*

38. *Rey, C., María, G., & Díaz, R. (2010). Introducción a la Estadística Descriptiva 2a Edición. Retrieved from https://ruc.udc.es/dspace/bitstream/handle/2183/11897/8497451678.pdf?sequence=2&isAllowed=y*

39. *Véliz Capuñay, C. (2015). Análisis multivariante : métodos estadísticos multivariantes para la investigación. Retrieved from http://latinoamerica.cengage.com/ls/analisis-multivariante-1a-ed/*

ISBN: 978-9942-33-286-8

@grupocompas.ec
compasacademico@icloud.com

compAs
Grupo de capacitación e investigación pedagógica