La Estadística como una Herramienta en la Metodología Científica

©2022 Johanna Enith Aguilar Reyes
Nancy Elizabeth Chariguamán Maurisaca
Marlon Ernesto Moscoso Martínez
Segundo Hugo Calderón

La Estadística como una Herramienta en la Metodología Científica

Nancy Elizabeth Chariguamán Maurisaca
Marlon Ernesto Moscoso Martínez
Segundo Hugo Calderón

Escuela Superior Politécnica de Chimborazo (ESPOCH)
Riobamba – Ecuador
Panamericana Sur Km. 1½
Teléfono: 593 (03) 2998-200
Código Postal EC0600155

2022
Publicado por acuerdo con los autores.
Este libro se sometió a arbitraje bajo el sistema de doble ciego *(peer review)*

Corrección y diseño, respaldado por:
La Caracola Editores
Índice Científico Editorial

La Estadística como una Herramienta en la Metodología Científica
Riobamba, Ecuador
Dirección de Publicaciones Científicas, 2022
ISBN: 978-9942-40-980-5
Fecha de Publicación: 2022-02-11

DEDICATORIA

Queremos agradecer primero a Dios porque nos dió el don de la perseverancia para alcanzar nuestra meta.

A nuestra querida Escuela Superior Politécnica de Chimborazo quien nos ha permitido crecer cada día para ser mejores profesionales, y a la vez quien abrió sus puertas para permitirnos impartir nuestros conocimientos.

A nuestras autoridades que gracias a su constancia nos han enrumbado en el camino de la investigación y los saberes.

Joha
Nancy
Marlon
Hugo

INTRODUCCIÓN

La presente obra denominada "La estadística como una herramienta en la metodología científica" busca saciar el conocimiento de análisis estadístico, juicio crítico, de experimentar o tomar datos, de escribir; se ha concebido como una guía referencial para realizar investigación científica desde una perspectiva estadística, está destinado tanto a estudiantes de grado y posgrado, investigadores en general. Su contenido se fundamenta tanto en la estadística descriptiva como en la estadística inferencial donde se muestran ejemplos prácticos del procesamiento y análisis de datos reales en trabajos investigativos, lo cual permite un enfoque de conocimiento enriquecido que facilita la comprensión del desarrollo de los temas que incluye la obra. Las autoras se han propuesto aportar conocimientos que forjen un aporte para el éxito de este proyecto académico.

Para la comprensión y estudio del contenido se requiere un nivel intermedio de matemática, los temas que contiene la presente obra son: investigación científica y estadística, consideraciones de la estadística descriptiva, teoría de probabilidades y estadística inferencial; en el desarrollo de cada capítulo se muestran ejemplos prácticos que ilustran el contenido.

ÍNDICE

CAPÍTULO 1

Investigación Científica y estadística

1. Investigación Científica y estadística

1.1 Introducción a la Estadística

El origen de la estadística es tan antiguo como la civilización, alcanzó un desarrollo importante con el surgimiento de los Estados, en este acontecimiento se convirtió en un instrumento decisivo. La concepción de la estadística ha evolucionado a través de la historia, donde inicialmente se limitaba a la recopilación, ordenamiento de datos sobre aspectos de interés, en este contexto la estadística pasó a vincularse estrechamente con la teoría de las probabilidades, constituyéndose en una rama de la matemática aplicada, utilizando principios y modelos matemáticos aplicables en toda área.

El campo de la estadística se remonta a 1654, con raíces en los juegos de azar, se ha desarrollado en el campo de estudio de métodos y pruebas para definir cuantitativamente la variabilidad en los datos, la probabilidad de resultados, el erros y la incertidumbre asociado a los resultados, en este contexto los métodos estadísticos son utilizados extensivamente en el proceso científico desde el diseño de las interrogantes de la investigación, análisis de datos y la interpretación final de resultados.

En general el investigador científico diseña los estudios de investigación establecidos en la naturaleza de los argumentos a investigar, pulen el plan de investigación de acuerdo a los conceptos de estadísticos con el fin de incrementar la posibilidad de que los descubrimientos sean útiles.

La estadística actual es la consecuencia de la alianza de dos disciplinas que han desplegado, la primera el cálculo probabilístico, cuyo origen se remonta al siglo XIX como la teoría de los juegos de azar, la segunda es la estadística o ciencia del Estado, que describe datos; las integraciones de estas dos líneas de pensamiento dan origen a la ciencia que estudia cómo generar conclusiones de la investigación empírica a través de la utilización de modelos matemáticos.

1.1.1 *Definición de Estadística*

La estadística se considera como un método científico, aplicable a una gran variedad de áreas del conocimiento, cuya utilidad es trascendental, porque va más allá de la simple descripción, llegando a descubrir leyes y tendencias, uno de los principales ejemplos que se puede citar es el caso del estadístico Ernesto Engel (1821-1896).

La estadística es una disciplina matemática cuya finalidad se enfoca en la interpretación de datos numéricos que se despegan de sucesos empíricos, se encarga del estudio de eventos o experimentos aleatorios, recoge y organiza un gran número de datos con el propósito de obtener alguna consecuencia (Martínez, 2020, Matus,2010).

La estadística estudia métodos científicos para recoger, organizar, resumir, analizar información, así como para obtener conclusiones válidas y tomar decisiones razonables fundamentadas en dicho análisis (Murray & Spiegel, 2005), representa un área de la ciencia que se encarga de diseñar experimentos, análisis de datos e inferencias sobre la población a partir de la información generada en una muestra.

Autores como Hernández y Oteyza (2015), Guerra (2003) y García et al., (2002), mencionan a la estadística como una disciplina encargada de recolectar, organizar, describir e interpretar datos, en este contexto se concluye que la estadística es una herramienta fundamental de análisis numérico que permite generar conocimiento sobre un evento.

Con lo antes expuesto la estadística brinda herramientas básicas para realizar investigación, esta disciplina aporta al descubrimiento de relaciones entre hechos y al fundamento de esos descubrimientos, es un modo de acercamiento al conocimiento de la realidad.

1.1.2 *Importancia de la estadística en la Investigación Científica*

La estadística es el vínculo usual que se presenta en casi todas las investigaciones científicas, en las cuales interviene el tratamiento, interpretación y predicción de datos; el estudio estadístico tiende a cobrar especial importancia al ejecutar investigaciones en campos de las ciencias como por ejemplo en la medicina, donde una mala interpretación genera consecuencias adversas para la población.

En la labor del investigador la estadística es una herramienta de apoyo fundamental en la investigación de fenómenos, es lógica con un fuerte ingrediente de procesos aritméticos, que crea material sobre el cual se basa la inferencia y se mide la incertidumbre, la estadística contribuye en el proceso investigativo en la etapa del diseño, plan de recolección de datos, análisis de resultados hasta la evaluación de la incertidumbre asociada a la inferencia extraída de ellos.

La estadística al ser una herramienta de análisis e interpretación de datos ha adquirido relevancia en todas las áreas de trabajo, siendo un factor claro en las predicciones y toma de decisiones en base a datos, la tarea más relevante de la estadística es la de suministrar alternativas cuantitativas, que se traduzcan en conclusiones objetivas, en este sentido las técnicas estadísticas permiten al investigador científico cuantificar la probabilidad de ocurrencia de un fenómeno.

*1.1.3 **Clasificación de la Estadística***

Una particularidad del trabajo estadístico práctico es el análisis o procesamiento de datos numéricos, toda conclusión generada por un procedimiento estadístico implica necesariamente el análisis de cantidades o características cuantitativas, razón por la cual es importante el uso de la estadística como una herramienta auxiliar que contribuye al proceso investigativo.

En la actualidad se han desarrollado una gama amplia de técnicas estadísticas para el análisis de datos, de manera general se dividen en dos grupos:

- **Estadística Descriptiva:** se encarga de reunir, presentar y organizar los datos, permite al científico adherir las propiedades más significativas de un conjunto de datos, aplicando medidas como el promedio, la media, desviación estándar; estas medidas proveen un sentido general del grupo de estudio. El objetivo esencial de este tipo de estadística es la caracterización de conjunto de datos numéricos, lo cual pone en manifiesto las propiedades del conjunto de datos (Cárdenas, 2014)

- **Estadística Inferencia:** se ocupa de analizar los datos de una muestra, con el fin de extraer conclusiones de una población de estudio, se utiliza para modelar patrones en datos, emitir juicios sobre los datos, identificar las relaciones entre las variables en el conjunto de datos e inferir sobre poblaciones más amplias en función de una muestra de datos, en la figura 1-1 se muestra el proceso de la inferencia estadística.

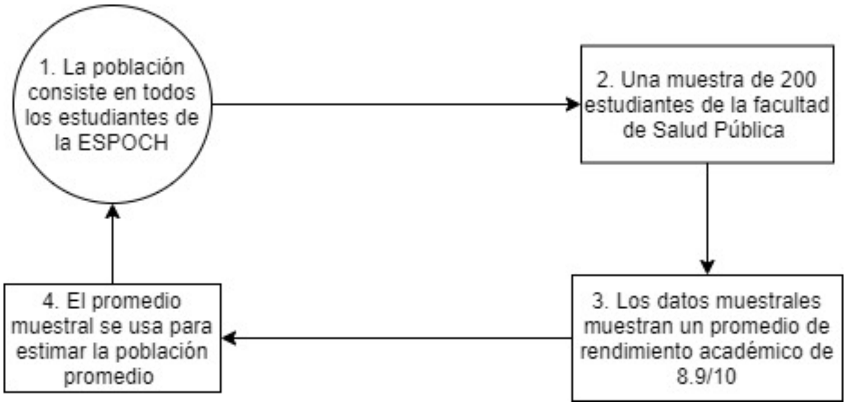


Figura 1-1. Proceso de inferencia estadística.

Sin embargo, dentro de la estadística se genera la teoría de probabilidades, que es una rama donde se utiliza el análisis de situaciones en las que interviene el azar, como, por ejemplo: juegos de cartas, tiro de moneda, en la mayoría de deportes.

1.1.4 ***Datos***

Un dato es un elemento de todo un conjunto que tiene un valor o peculiaridad con el cual se distingue de los demás (Hernández & Oteyza, 2015), el número de datos se puede contabilizar, sin embargo; los valores que adquieren los datos no siempre son expresados a través de una medida numérica, sino por un atributo, ejemplo: los colores (azul, rojo, amarillo, etc), nivel de educación (elemental, básica, secundaria, superior, posgrado), en este contexto los datos se clasifican en: cuantitativos y cualitativos.

- **Datos Cuantitativos:** son aquellos que se pueden contar o medir a través de una expresión numérica, ejemplo: edades de los estudiantes (pueden expresarse en años 19 años, 25 años, etc.), el precio de un bien o servicios (expresado en valor monetario 19 USD, 35.5 USD). Se subdividen en datos continuos que pueden tomar cualquier valor de los números reales (altura de una persona 1.70 metros) y discretos cuando solo toman valores enteros (5 mesas, 3 sillas).

- **Datos Cualitativos:** son aquellos cuyos valores no se pueden cuantificar, sino se expresan un nivel de calidad o señalan un atributo por el que se puede identificar a cada elemento, es decir; aquellos datos que pueden identificarlos entre sí como sexo (hombre, mujer), estado civil (soltero, casado, viudo, divorciado, unión libre)

*1.1.4.1 **Técnicas de recolección de datos***

La recolección de datos se refiere a un enfoque sistemático que permite reunir y medir información de distintas fuentes, permite al investigador responder a preguntas relevantes, evaluar resultados y anticipar mejor las tendencias.

Existen diferentes técnicas de recolección de datos, la selección de la técnica depende del tipo de variable, a precisión deseada, el punto de recolección y las habilidades del encuestador.

1. **Observación:** permite conocer el comportamiento del objeto de estudio de manera directa, la forma más adecuada de aplicar está técnica es registrar las observaciones en notas de campo o en alguna plataforma. Esta técnica se caracteriza por no ser instructiva y requiere evaluar el comportamiento del objeto de estudio continuamente sin intervenir.

2. **Encuestas:** esta técnica consiste en obtener directamente datos de los sujetos de estudio, con el fin de lograr resultados deseados es fundamental tener claro los objetivos de investigación, además es importante elaborar los cuestionarios para aplicar las encuestas cuidadosamente, definiendo que tipo de cuestionario es el más eficiente para la recolección de datos, en este sentido los cuestionarios más populares son:

    a. *Cuestionarios Abiertos:* son aplicados para conocer a profundidad la perspectiva de la persona sobre un tema específico, analizar sus opiniones y obtener información al detalle.

    b. *Cuestionarios Cerrados:* son aplicados para obtener una cantidad de información suficiente, las respuestas de los encuestados son limitados, pueden contener preguntas de opción múltiple o dicotómicas (si/no).

3. **Focus group:** es una técnica cualitativa, consiste en llevar a cabo una reunión donde las personas dan su opinión y buscan resolver un problema específico, una de las cualidades de esta técnica es la posibilidad de obtener varios puntos de vista sobre un mismo tema para llegar a una óptima solución.

4. **Entrevista:** este método consiste en recopilar información aclarando una pregunta, por medio de la comunicación interpersonal, el emisor tiene respuestas verbales del recepto sobre un tema o problema específico.

*1.1.4.2 **Tipos de Datos***

En estadística los datos de pueden diferencias en agrupados y no agrupados, en estadística no existe estándares establecidos para definir el uso apropiado de datos agrupados o no agrupados, sin embargo; la sugerencia es que cuando el total de datos (N) es igual o superior a 20, se utilice una distribución de datos agrupados.

- **Datos agrupados:** son aquellos datos que tienen como característica principal la frecuencia con la que se presentan, es decir; aquellos datos que se encuentran contados y clasificados, ya sea por sus cualidades cuantitativas o cualitativas, razón por la cual pueden agruparse.

- **Datos no agrupados:** son aquellos datos que carecen de frecuencia.

*1.1.5 **Variables***

Una variable estadística representa una característica de una muestra o población de datos que puede adoptar diferentes valores, las variables se clasifican cuantitativas y cualitativas, en la figura 1-2 se muestra la clasificación de las variables.

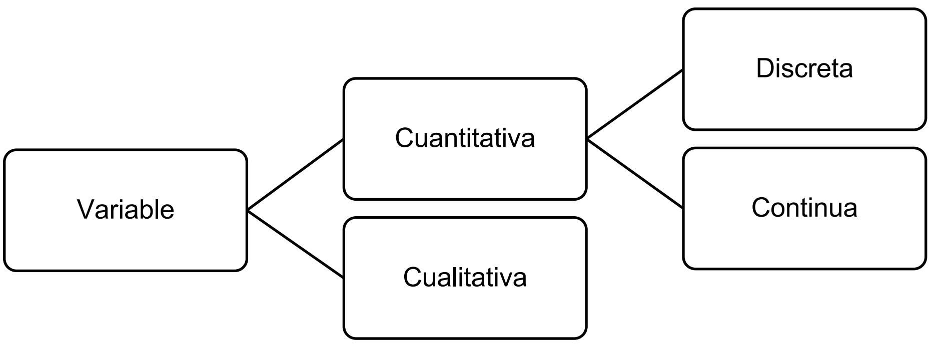


Figura 1-1. Clasificación de Variables.

Variables Cuantitativas

- *Variable continua:* son aquellas que surgen mediante el proceso de medición, aceptan números decimales.

- *Variable discreta:* surgen por el proceso de conteo, acepta únicamente valores enteros.

Variables Cualitativas

- *Variable continua:* son aquellas que surgen mediante el proceso de medición, aceptan números decimales.

- *Variable discreta:* surgen por el proceso de conteo, acepta únicamente valores enteros.

Desde la perspectiva de las pruebas estadísticas las variables se clasifican en:

- *Variable Independiente:* es explicativa de la dependiente, un ejemplo los puntajes de un test de memoria puede ser explicativo del estado cognitivo de normalidad o deterioro.

- *Variable Dependiente:* es la variable explicada por la independiente, por ejemplo, el estado cognitivo de una persona puede explicarse por un conjunto de variables independientes como, edad, nivel de estudio, nivel económico social, entre otros.

- *Variables Intervinientes:* conocidas como de control, ejemplo la presencia de depresión puede influir en el resultado del test, desviando los resultados reales.

*1.1.6 **Escalas de medición***

La recolección de datos requiere una de las escalas de medición sea nominal, ordinal, de intervalo, de razón, la escala de medición determina la cantidad de información contenida en los datos e indica la manera más apropiada de resumen y análisis de daros estadísticamente (Anderson et al., 2012).

En este sentido se puede decir que las escalas o niveles de medición se refieren a la relación entre los valores que se asignan a los atributos de una variable.

- **Escala nominal:** cuando los datos de una variable se componen de etiquetas o nombres utilizados para identificar un atributo del elemento.

Ejemplo 1.1

¿Cuál es el grado de incomodidad con el ruido generado por el movimiento vehicular? 1. Leve 2. Moderado 3. Severo

Ejemplo 1.2

¿Cuál es tu género?
H – hombre.
M – mujer.

Ejemplo 1.3

¿Vive en Riobamba?
Si
No

Este subtipo de escala nominal se conoce como escala nominal dicotómica.

- **Escala ordinal:** cuando los datos exhiben las propiedades de los datos nominales y su orden de clasificación.

Ejemplo 1.4

En la calificación del servicio excelente, bueno o malo, dando a cada uno de ellos un número donde que 3 es excelente, 2 bueno, 1 es malo.

Lo que hacen los encuestados es elegir entre las opciones de satisfacción, pero claro la respuesta a la pregunta "¿cuánto exactamente?" permanece sin respuesta. Comprender las diversas escalas de medición ayudan a los investigadores a obtener datos que pueden ser aplicados a favor en el futuro. Por lo tanto, se utiliza una escala ordinal como parámetro para comprender si las variables son mayores o menores. La tendencia central de la escala ordinal es mediana.

- **La escala de Likert:** es un ejemplo de porque la diferencia de intervalo entre las variables ordinales no se puede concluir. En esta escala de hecho, las opciones de respuesta suelen ser polares, como, por ejemplo, algo como "totalmente satisfecho" o "totalmente insatisfecho".

Ejemplo 1.5

¿Qué tan satisfecho estás con nuestros productos?

1. Totalmente satisfecho
2. Satisfecho
3. Neural
4. Insatisfecho
5. Totalmente insatisfecho

- **Escala de intervalo:** presentan todas las propiedades de los datos ordinales y el intervalo entre los valores se expresan en términos de una unidad de medida fija, los datos de intervalo son siempre numéricos. La escala de intervalo es el tipo de pregunta que se utiliza con mayor frecuencia en un estudio o investigación. Para obtener cualquier tipo de respuesta, es indispensable que la pregunta solicitada requiera que los encuestados respondan en una escala numérica donde la diferencia entre los dos números sea la misma.

Ejemplo 1.6
¿Cómo fue su experiencia con la comida en el restaurante ESPOCH?

	Mucho	Poco	Neutral
Picante			
Aburrida			
Agradable			

- **Escala de razón:** tiene todas las propiedades de los datos de intervalo y la razón de dos valores son significativos. Para la medición de variables como distancia, estatura, peso tiempo se usa la escala de razón.

*1.1.7 **Población y muestra***

Población

La población es definida como el total de elementos o datos asociados en un estudio, conjunto de elementos o conjuntos que presentan una característica en común (Lind et al, 2012). En estadística la población es el total de individuos o conjunto de ellos que presenta un rasgo característico que se pretende estudiar, en este contexto existen dos tipos de poblaciones:

- **Población estadística finita:** número de valores que tienen un fin, ejemplo la cantidad de árboles de la ciudad de Riobamba, el número de estudiantes de la Escuela Superior Politécnica de Chimborazo.

- **Población estadística infinita:** se trata de aquella población que no tiene fin, ejemplo el número de estrellas.

El tamaño de la población representa el número total de entes, fenómenos, eventos que la conforman, se simboliza con la letra **N.**

Se llama elemento a cada uno de los entes, fenómenos o eventos que integran la población.

Muestra

La muestra es una porción o parte de la población de interés, la toma de muestras sirve para aprender algo de una población es de uso frecuente en la administración, agricultura, política, acciones de gobierno, es un subgrupo de la población.

En estadística el término muestra es usado para denominar a cualquier subconjunto de una población, existen dos tipos de muestreo el aleatorio o probabilístico y no probabilístico, cada uno de ellos incluye distintas clases de muestreo que distinguen de acuerdo a los factores característicos de la población.

En una investigación no siempre se utiliza la muestra, pero de manera general se aplica una muestra al estudio, solo cuando se efectúa un censo no se aplica muestra, porque en ese contexto se incluyen todos los casos del universo o población (Hernández et al., 2016), en la figura 1-3 se muestra los tipos de muestra que se puede aplicar.

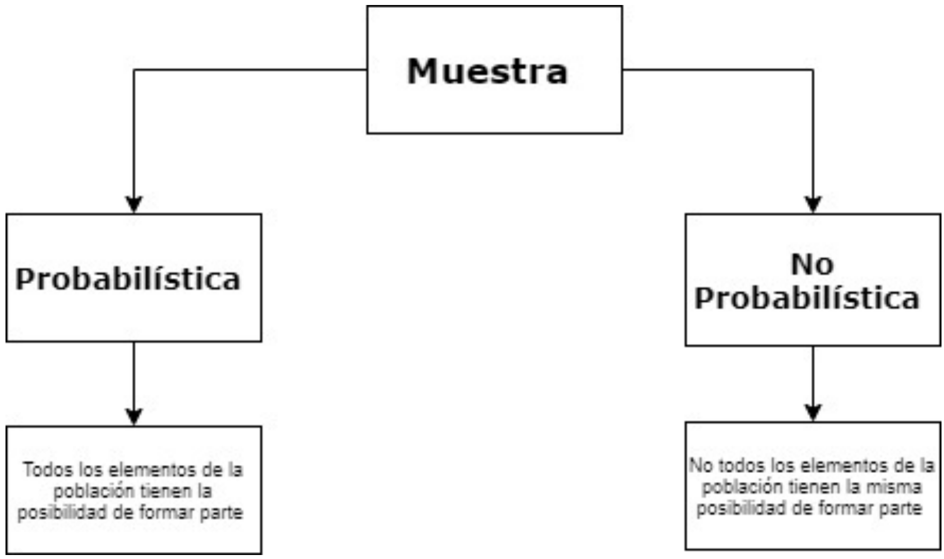


Figura 1-1. Muestra Estadística.

Muestreo Probabilístico se subdivide en cuatro tipos principales:

a. **Muestro aleatorio simple:** es un procedimiento de muestreo probabilístico en el cual cada elemento de la población objetivo tiene la misma probabilidad de ser seleccionado, este tipo de muestreo no se utiliza en investigaciones del consumidor.

Ejemplo 1.7

Cobertura de la vacuna COVID – 19 en la escuela ABC entre 1500 estudiantes, muestra 60 niños, se realiza un listado de todos los niños, se los numera del 1 al 1500 y se selecciona aleatoriamente 60.

b. **Muestreo estratificado:** procedimiento de muestreo en el que el objetivo de la población se separa en segmentos exclusivos, homogéneos (estratos) y luego se aplica una muestra aleatoria simple se selecciona de cada segmento o estrato. Las muestras seleccionadas de los distintos estratos se combinan en una sola muestra, este procedimiento se denomina como muestreo de cuota aleatoria.

Ejemplo 1.8

Para obtener una muestra de 100 individuos de una población de 1000, e divide a la población en los siguientes estratos:

Estrato 1: 300 individuos
Estrato 2: 500 individuos
Estrato 3: 200 individuos

Mediante el muestreo estratificado proporcionado, la muestra obtenida de cada estrato será representativa de cada uno de estos y otorgará los siguientes resultados:

Estrato	Individuos	Porcentaje	Muestra
Estrato 1:	300	30%	30
Estrato 2:	500	50%	50
Estrato 3:	200	20%	20

De esta manera se conforma la muestra de 100 personas con resultados equitativos, sin embargo, dicha muestra no puede ser considerada completamente probabilística para todo estrato, ya que los individuos del grupo con menor cantidad de personas poseen más posibilidades de ser seleccionados para la muestra que los otros estratos.

c. **Muestreo Sistemático:** consiste en una selección aleatoria del primer elemento para la muestra, luego se selecciona elementos posteriores utilizando intervalos fijos o sistemáticos hasta logar el tamaño de la muestra deseada. Desde un punto de vista técnico, este muestreo no crea una muestra verdaderamente aleatoria, solo la selección del primer elemento de muestreo sistemático es una selección de probabilidad, algunos elementos tendrán una probabilidad cero de selección.

Ejemplo 1.9

Extraer una muestra de 10 personas a partir de una población total de 100 y el primer individuo seleccionado para la muestra es el número 3. A partir de este, mediante un intervalo de 4 decidido por el investigador, se seleccionarán los próximos individuos hasta completar la muestra, de manera que serán los números 7, 11, 15, etc.

d. **Muestro por conglomerado:** denominado muestreo por racimos, donde que los elementos de la población se seleccionan al azar por agrupaciones (clusters[1]), los elementos del muestreo se seleccionan de la población de manera individual, uno a la vez.

[1] Cluster representa una concentración de empresas interconectadas en la actividad económica que desarrollan.

Las unidades de muestreo o grupos pueden ser espaciados, tal como ocurre en las unidades físicas o geográficas, ejemplo: estados, provincias, cantones, distritos, parroquias; en base a una organización instituciones de educación superior, nivel escolar; la heterogeneidad del grupo es fundamental para un buen diseño del muestreo por conglomerado, los elementos dentro de cada grupo deben ser tan heterogéneos como la población objetivo misma, las dimensiones de este tipo de muestreo se basan en el número de etapas del diseño de la muestra y en la representación proporcional de los grupos en la muestra.

1.1.8 ***Tamaño de la muestra***

La muestra es una selección de encuestados elegidos y que representan a la población total, su tamaño se traduce en una representación significativa de la población, que cumple con peculiaridades relacionadas a la investigación. Para determinar el tamaño de la muestra dentro de una investigación es importante tener en cuenta los objetivos y circunstancias que se desarrollan el trabajo investigativo.

a. **Muestra para una población finita**

$$n = \frac{N * Z^2 * p * q}{e^2 * (N-1) + Z^2 * p * q} \tag{1.1}$$

Dónde:
n = tamaño de la muestra.
N = población o universo.
Z = Nivel de confianza.
p = probabilidad a favor.
q = probabilidad en contra.
e = error muestral.

El nivel de confianza Z, es un valor constante necesario, los niveles de confianza más comunes son:

Nivel de Confianza	Valor de Z	Error Muestral
90%	1.645	10%
95%	1.960	5%
99%	2.576	1%

Ejemplo cálculo de muestra finita

Una institución de educación superior tiene 6530 estudiantes, el investigador para un estudio de satisfacción de la calidad de la educación, toma una muestra del universo a investigar, asignando un nivel de confianza del 95 %, se desconoce la probabilidad p.

En este caso el valor de Z es de 1.96, e = 5%; N 6530, entonces se aplica la fórmula 1.1, al no conocer la probabilidad de ocurrencia p, se asume que tanto p como q representan el 50 %

$$n = \frac{6530 * 1.96^2 * 0.5 * 0.5}{0.05^2 * (6530 - 1) + 1.96^2 * 0.5 * 0.5}$$
$$n = 363$$

b. **Muestra infinita**

Cuando la población es desconocida (número), el investigador debe aplicar la siguiente fórmula:

$$n = \frac{Z^2 * p * q}{e^2} \tag{1.2}$$

Dónde:
n = tamaño de la muestra buscada.
e= error de estimación máximo permitido
p = probabilidad a favor.
q = probabilidad en contra.

1.2 *Investigación Científica*

Investigar significa una serie de métodos encaminados a la adquisición individual de conocimiento no poseído, implica la búsqueda de nuevas relaciones entre elementos ya conocidos, aspectos novedosos, es decir; la investigación constituye la generación de conocimiento.

La investigación es considerada como la base que permite construir una guía, que proporciona a la ciencia no solamente esquemas descriptivos sino causales y conceptuales del entorno (Pérez 2008; Sautu et al., 2014), en otras palabras es la guía en relación con los eventos y sus conexiones causales recíprocas, la naturaleza de la ciencia y por consecuencia la investigación han sido explicadas por la rama de la filosofía denominada Filosofía de la ciencia, que constituye una disciplina del razonamiento humano para comprender cuál es el fundamento de la ciencia.

Desde la perspectiva académica, la investigación es una actividad que se ejecuta de manera sistemática, controlada y crítica. La finalidad de la investigación científica de acuerdo a los autores Cruz et al., (2014); Mora & Sepúlveda (1999) es descubrir, describir, interpretar hechos o fenómenos, así como establecer relaciones entre los hechos o fenómenos, generar, divulgar conocimiento, producir teorías y resolver problemas prácticos.

La investigación científica es un procedimiento dinámico, caracterizado por su rigurosidad y conducta en la adquisición de nuevos conocimientos (Monroy & Nava, 2018), su principal función se enfoca en describir, comprender, controlar, predecir fenómenos, comportamientos, hechos, autores como Rodolfo Mondolfo (1961) afirman que la investigación surge cuando se tiene conciencia de un problema, que el ser humano tiende a buscar la solución del mismo, la indagación realizada para alcanzar esa solución representa la investigación propiamente dicha.

Se denomina investigación científica a aquella investigación realizada con el propósito de contribuir a la ciencia mediante la recolección, interpretación y evaluación sistemática de datos de manera planificada (Capalar & Dönmez, 2016)

El punto de partida de la investigación es la existencia de un problema, que debe ser definido, examinado, valorado y analizado críticamente, para el surgimiento de su solución.

1.2.1 ***Clasificación de la investigación científica***

Las investigaciones pueden definirse desde distintos puntos de vista, las más usuales que se considera en cualquier área se detallan en la figura 1-4.

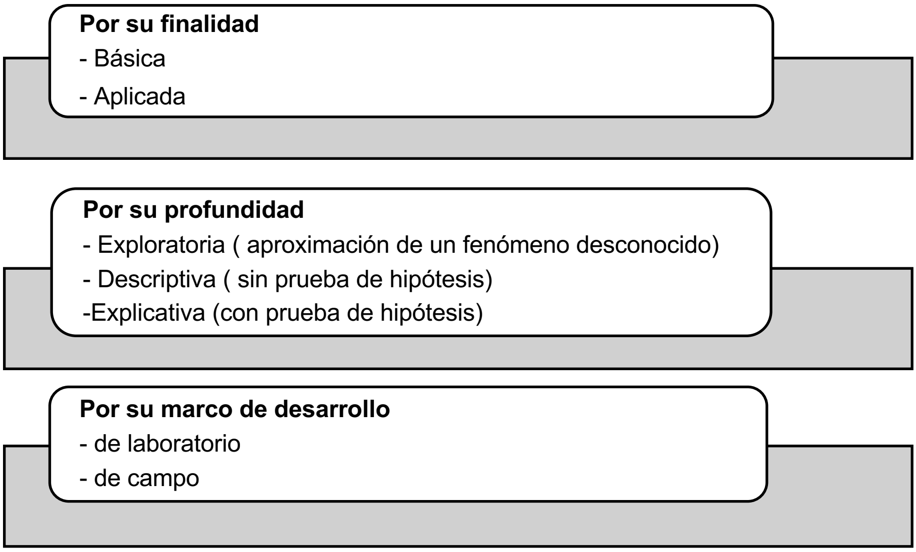


Figura 1-4. Tipos de Investigación.

Investigación Básica: también denominada pura o teórica, se caracteriza porque es parte del marco teórico, su finalidad es formular nuevas teorías o modificar las existentes, incrementar conocimientos científicos.

Investigación Aplicada: conocida como empírica, su característica radica en la búsqueda de la aplicación o utilización de los conocimientos que se adquiere, se vincula con la investigación básica, porque depende de los resultados y avances de esta.

Investigación Exploratoria: se realiza cuando el objeto a examinar es un tema o problema de investigación poco estudiado, del cual existen muchas dudas, sirva para familiarizarse con fenómenos relativamente desconocidos.

Investigación Descriptiva: este tipo de investigación indaga como especificar características, propiedades y perfiles sea de personas, grupos, comunidades, objetos, procesos o cualquier fenómeno que se someta a un análisis, en esta investigación se mide o recolecta información de manera independiente o conjunta de las variables, además ofrece la posibilidad de hacer predicciones.

Investigación explicativa: su principal objetivo es contribuir con los investigadores a estudiar un problema con mayor profundidad y entender el fenómeno a detalle, busca definir las causas que han generado el problema, no solo busca describir sino determinar las causas del problema, este tipo de investigación va más allá de la descripción del fenómeno, se enfoca en responder por las causas de los eventos, hechos o fenómenos.

Investigación de Laboratorio: es una investigación experimental, que recurre al razonamiento, a través de experimentos busca dar respuesta a una hipótesis.

Investigación de Campo: es la recopilación de datos de fuentes primarias para un fin específico, este tipo de investigaciones utiliza instrumentos como ficheros o representaciones estadísticas que permiten recopilar y analizar los datos que se van a estudiar. Se ejecuta en el lugar de los hechos, implica tomar información de fuente directa, sin manipular ni controlar las variables, este tipo de investigación permite observar un fenómeno en condiciones reales (Monroy & Nava, 2018).

*1.2.2 **Método Científico***

El método científico es un proceso sistemático que se sujeta a nociones y reglas para alcanzar un fin predeterminado, procura establecer los procedimientos que deben seguirse, de manera general se pude decir que el método científico se aplica a un ciclo completo de una investigación, en búsqueda de soluciones a cada problema del conocimiento, autores como Monroy (2008) establecen que el método científico es un procedimiento que exige sistematización del pensamiento para desarrollar una investigación reflexiva.

Las ramas de la estadística utilizan el método científico, el cual se adapta a ella en cinco pasos:

- Definir el problema, estableciendo el objeto de estudio.
- Formular un plan para recopilar datos.
- Recopilar información.
- Analizar e interpretar los datos.
- Anotar las conclusiones y otros hallazgos, de tal manera que sean de fácil comprensión y puedan ser utilizados al tomar decisiones.

Es importante mencionar que la estadística al ser un conjunto de herramientas y métodos aplicados en la solución de problemas, no se considera netamente matemático.

1.2.3 ***Enfoques de la investigación***

La investigación al ser un conjunto de procesos sistemáticos, críticos y empíricos aplicables a un problema, existen dos enfoques tradicionales de la investigación: enfoque cuantitativo y enfoque cualitativo, ambos enfoques son utilizados para la generación del conocimiento, sin embargo; en la última década los investigadores han tendido al uso combinado de ambos enfoques generando el enfoque mixto, argumentando que al probar una teoría con ambos enfoques se generan resultados más confiables, en la figura 1-5 se observa una breve descripción de los enfoques.

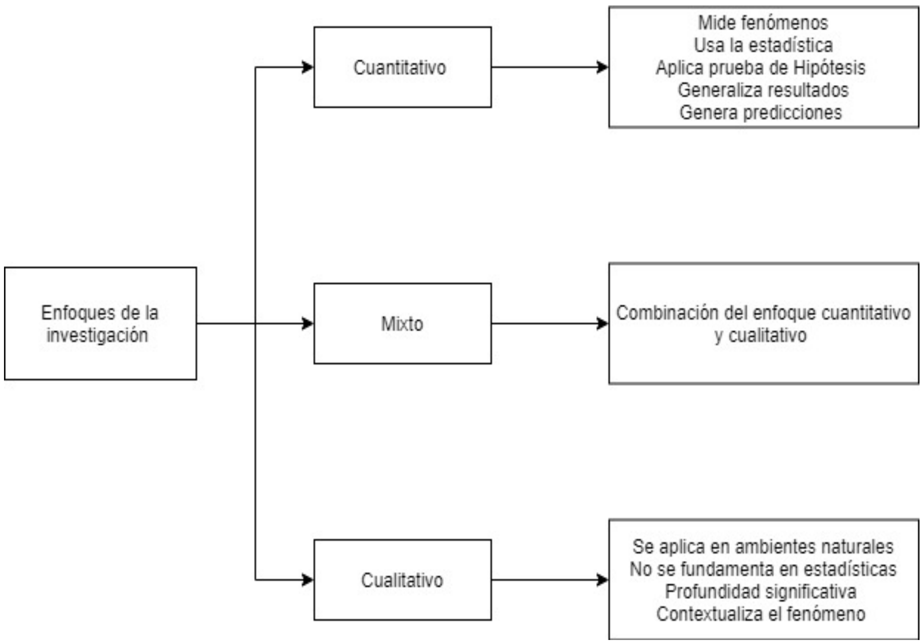


Figura 1-5. Enfoques de la Investigación.

1.2.3.1 ***Enfoque Cuantitativo***

El enfoque cuantitativo es secuencial probatorio, la investigación cuantitativa considera que el conocimiento debe ser objetivo, se genera a partir de un proceso deductivo, por medio de medición numérica y análisis estadístico inferencial, prueba hipótesis formuladas con anterioridad, este enfoque está asociado con normas y prácticas del positivismo (Hernández et al., 2017).

En este contexto el enfoque investigativo cuantitativo utiliza la recolección de datos con fundamento en la cálculo numérico e investigación estadística para establecer pautas de comportamiento y probar teorías.

a. Características del Enfoque Cuantitativo

Las principales características del enfoque cuantitativo son:

- Muestra la necesidad de evaluar y calcular magnitudes de los fenómenos a investigar.
- El planteamiento del problema a investigar se fundamenta en cuestiones específicas.
- Con el planteamiento del problema, quien investiga (investigador) toma en consideración los antecedentes investigativos, construye la revisión de literatura, se fundamenta en las teorías existentes como una guía, de la cual se deriva hipótesis que son sometidas a prueba.
- Las hipótesis se generan antes de recolectar y analizar los datos.
- Los datos recolectados tienen su base en la medición de variables contenidas en la hipótesis, se utilizan procesos estandarizados y aceptados en la comunidad científica para la recolección de datos.
- El análisis de datos se realiza con métodos estadísticos.
- Confianza en la experimentación o las pruebas de causalidad.
- Las predicciones iniciales permiten interpretar los análisis cuantitativos, como una explicación de los resultados del conocimiento existente.
- Debe ser objetiva, evitar sesgos que influyan en el resultado.
- Este tipo de investigación sigue un patrón estructurado.
- Trata de generalizar los resultados en una muestra hacia el universo o población.
- Pretende confirmar y predecir los fenómenos analizados, identificando relaciones causales entre elementos.
- Identifica leyes universales y causales.
- Explica cómo se concibe la realidad con una aproximación a la investigación.

b. Proceso del enfoque cuantitativo

El proceso del enfoque cuantitativo consta de 10 fases que se enlistan a continuación:

- Idea.
- Planteamiento del problema.

- Revisión de literatura y desarrollo del marco teórico.
- Visualización del alcance del estudio.
- Elaboración de hipótesis y definición de variables.
- Desarrollo del diseño de investigación.
- Definición y selección de la muestra.
- Recolección de datos.
- Análisis de datos.
- Elaboración del reporte de resultados.

1.2.3.2 *Enfoque Cualitativo*

El enfoque cualitativo está basado en la recolección de datos sin medición numérica para descubrir o afinar preguntas de investigación en el proceso de interpretación, tiene como objetivo la descripción de las cualidades de un fenómeno, a través de este enfoque se genera teorías e hipótesis.

A este enfoque también se le denomina investigación naturalista, fenomenológica, interpretativa o etnográfica, en el cual se incluyen una variedad de técnicas no cuantitativas, dentro de sus principales características se destacan:

- El investigador plantea un problema, pero no sigue un proceso definido claramente.
- La investigación se inicia examinando hechos, se fundamenta en la lógica y el proceso inductivo.
- Los estudios cualitativos no prueban hipótesis.
- Se basa en métodos de recolección no estandarizados, ni predeterminados, la recolección consiste en obtener los puntos de vista de los participantes.
- Las principales técnicas de recolección en este enfoque son la observación no estructurada, entrevistas abiertas, revisión de documentos, evaluación de experiencias individuales, reflexión de grupos.
- La indagación es flexible, su propósito es reconstruir la realidad.
- Valora el progreso natural de los procesos, no hay manipulación de la realidad.
- Interpreta centradamente el significado de las acciones.

1.2.3.3 *Diferencias entre la investigación cuantitativa y cualitativa*

La utilización de ambos enfoques en una investigación, permiten corregir sesgos de cada uno de los métodos aplicados, en la tabla 1-2 se identifican las principales diferencias entre estos enfoques.

Tabla 1-2. Diferencia entre enfoque cuantitativo y cualitativo.

Enfoque Cuantitativo	Enfoque Cualitativo
- Basada en la inducción probabilística	- Centrado en la fenomenología y comprensión
- Medición penetrante y controlada	- Observación naturista sin control
- Objetivo	- Subjetiva
- Inferencia más allá de los datos	- Inferencia de sus datos
- Confirmatoria, inferencia, deductiva	- Exploratoria, inductiva y descriptiva
- Orientada a resultados	- Orientada a procesos
- Datos sólidos y repetibles	- Datos profundos y ricos
- Generalizable	- No generalizable
- Particularista	- Holística
- Realidad estática	- Realidad estática

En la aproximación del enfoque cuantitativo los planteamientos a investigar desde el inicio del estudio son delimitados y específicos, las hipótesis son establecidas previo a la recolección y análisis de datos, mientras que en el enfoque cualitativo se enfoca en reconstruir la realidad tal como observan los actores sociales previamente definidos. La investigación cuantitativa es objetiva y busca generalizar los resultados encontrados en una muestra a una población, la cualitativa es subjetiva no busca generalizar los resultados hacia una población.

1.3 *Lineamientos éticos para la práctica de la estadística en la investigación*

El componente ético es fundamental en todo el accionar del investigador, los problemas éticos surgen en la estadística debido a la importancia que tiene esta en la recopilación, análisis, presentación e interpretación de datos; en un estudio estadístico el comportamiento poco ético provoca un muestreo no apropiado, como consecuencia se genera un desarrollo erróneo de resultados estadísticos.

A medida que se avance en el trabajo investigativo estadístico es recomendable que sea justo, meticuloso, objetivo y neutral en la recolección de datos, análisis, presentación de informes, de acuerdo a autores como Seltzer (2005) los desafíos éticos sobre la estadística se pueden expresar como: utilizar metodología adecuada, proteger la confidencialidad.

1.3.1 *Guía básica de la ética aplicada a la estadística investigativa*

La ética se aplica en la estadística porque se importante tener valores morales para no engañar a las personas con datos falsos que en algunos casos se generan en el contexto estadístico, por esta razón todo proceso estadístico debe presentar:

- Respeto a las personas,
- Tener como fin hacer el bien.
- Fomentar la justicia.
- Promover el desarrollo pleno.
- Ser transparentes.
- Evitar la selección y exclusión de datos a conveniencia del investigador.
- Evitar realizar solo aquellos análisis estadísticos que parecen favorecer alguna hipótesis en particular, examinar las aplicaciones alternativas a lo observado.
- La narrativa debe ser acorde con los resultados esperados, es decir las explicaciones e interpretaciones de los resultados deben responder a la evidencia.

Una pauta en los valores compartidos del informe de la American Stadistical Association (2016) establece que los profesionales de la estadística deben evitar cualquier tendencia a sesgar su trabajo hacia resultados predeterminados, este tipo de práctica poco ética se observa con frecuencia cuando utiliza muestra poco representativas.

La disciplina de la estadística vincula la capacidad de observar con la capacidad de recopilar evidencia y tomar decisiones, proporcionando una base para construir una sociedad más informada. Debido a que la sociedad depende de juicios informados respaldados por métodos estadísticos, todos los profesionales de la estadística, independientemente de su capacitación, ocupación o cargo, tienen la obligación de trabajar de manera profesional, competente y ética y de desalentar cualquier tipo de mala conducta profesional y científica.

La buena práctica estadística se basa fundamentalmente en supuestos transparentes, resultados reproducibles e interpretaciones válidas. En algunas situaciones, los principios de las directrices pueden entrar en conflicto, lo que requiere que las personas prioricen los principios de acuerdo con el contexto.

1.3.2 *Profesional estadístico ético*

El estadístico ético utiliza una metodología y datos que son relevantes y apropiados, sin favoritismos ni prejuicios, y de una manera destinada a producir resultados válidos, interpretables y reproducibles.

El estadístico ético no acepta a sabiendas trabajos para los que no está suficientemente calificado, es honesto con el cliente sobre cualquier limitación de experiencia y consulta a otros estadísticos cuando es necesario o tiene dudas, más allá de la muestra a una población relevante para los objetivos con un error mínimo bajo supuestos razonables.

Respeta y reconoce las contribuciones y la propiedad intelectual de otros, al establecer el orden de autoría para carteles, artículos y otras becas, se esfuerza por dejar en claro las bases de este orden, si se determina por motivos distintos a la contribución intelectual.

Identifica y mitiga cualquier preferencia por parte de los investigadores o proveedores de datos que pueda predeterminar o influir en los análisis / resultados.

Emplea métodos de selección o muestreo y enfoques analíticos apropiados y válidos para la pregunta específica a ser abordada, de manera que los resultados se extiendan.

Revela conflictos de interés, financieros y de otro tipo, y los maneja o resuelve de acuerdo a lo establecido (institucional / regional / local) reglas y leyes.

Acepta la plena responsabilidad de su desempeño profesional. Proporciona solo testimonios de expertos, trabajos escritos y presentaciones orales que él / ella estaría dispuesto a recibir revisión de pares.

1.3.3 *Integridad de datos y métodos*

El estadístico ético es franco sobre cualquier restricción, deterioro o sesgo conocido o sospechado en los datos que pueda aquejar la integridad o confidencialidad del análisis estadístico. La interpretación objetiva y válida de los resultados demanda que el análisis profundo registre y reconozca el nivel de seguridad e probidad de los datos.

Desde un punto de vista ético el investigador estadístico debe:

Reconocer los supuestos estadísticos y sustantivos hechos en la ejecución e interpretación de cualquier análisis. Al informar sobre la validez de los datos utilizados, reconoce los procedimientos de edición de datos, incluidos los mecanismos de imputación y datos faltantes.

Informa las limitaciones de la inferencia estadística y las posibles fuentes de error, en publicaciones, informes o testimonios, identifica quién es el responsable del trabajo estadístico si de otra manera no sería evidente.

Informa las fuentes y la idoneidad evaluada de los datos; explica todos los datos considerados en un estudio y explica las muestras realmente utilizadas.

Informa de manera clara y completa los pasos tomados para preservar la integridad de los datos y los resultados válidos, cuando sea apropiado, aborda las posibles variables de confusión no incluidas en el estudio.

En publicaciones e informes, transmite los hallazgos de manera honesta y significativa para el usuario / lector, esto incluye tablas, modelos y gráficos, identifica al patrocinador financiero final del estudio, el propósito declarado y el uso previsto de los resultados del estudio.

Al informar análisis de datos de voluntarios u otros datos que pueden no ser representativo de una población definida, incluye las exenciones de responsabilidad adecuadas y, si se utiliza, la ponderación adecuada Para ayudar a la revisión por pares y la replicación, comparte los datos utilizados en los análisis siempre que sea posible / permitido, y ejerce la debida precaución para proteger los datos patentados y confidenciales, incluidos todos los datos que podrían revelar de manera inapropiada las identidades de los encuestados.

Se esfuerza por corregir rápidamente cualquier error descubierto durante la elaboración del informe final o después de la publicación. Según corresponda, difunde la corrección públicamente o entre otras personas dependiendo de los resultados.

1.3.4 Responsabilidades con la ciencia, público, financiador, cliente

El estadístico ético apoya inferencias válidas, transparencia y buena ciencia en general, teniendo en cuenta los intereses del público, el financiador, el cliente o el cliente (así como los colegas profesionales, los pacientes, el público y la comunidad científica).

En la medida de lo posible, presenta a un cliente o empleador opciones entre enfoques estadísticos alternativos válidos que pueden variar en alcance, costo o precisión, se esfuerza por explicar cualquier consecuencia adversa esperada de no cumplir con un plan de muestreo o análisis acordado.

Aplica procedimientos de muestreo y análisis estadístico de manera científica, sin predeterminar el resultado, se esfuerza por hacer que los nuevos conocimientos estadísticos estén ampliamente disponibles para proporcionar beneficios a la sociedad en general y más allá de su propio ámbito de aplicación. Entiende y cumple con los requisitos de confidencialidad de la recopilación, divulgación y difusión de datos y cualquier restricción sobre su uso establecida por el proveedor de datos (en la medida en que lo exija la ley), y protege el uso y la divulgación de datos en consecuencia. Protege la información privilegiada del empleador, cliente o financiador.

1.3.5 ***Responsabilidades con los sujetos de investigación***

El estadístico ético protege y respeta los derechos e intereses de los sujetos humanos y animales en todas las etapas de su participación en un proyecto. Esto incluye a quienes respondieron al censo o encuestas, aquellos cuyos datos están contenidos en registros administrativos y sujetos de investigación física o psicológicamente invasiva.

El investigador estadístico desde una perspectiva ética se mantiene informado y se adhiere a las reglas, aprobaciones y pautas aplicables para la protección y el bienestar de los sujetos humanos y animales. Se esfuerza por evitar el uso de un número excesivo o inadecuado de sujetos de investigación y el riesgo excesivo para los sujetos de investigación (en términos de salud, bienestar, privacidad y propiedad de sus propios datos), haciendo recomendaciones informadas sobre el tamaño del estudio.

Protege la privacidad y la confidencialidad de los sujetos de investigación y los datos que les conciernen, ya sean obtenidos de los sujetos directamente, de otras personas o de registros existentes. Anticipa y solicita aprobación para usos secundarios e indirectos de los datos, incluida la vinculación con otros conjuntos de datos, al obtener las aprobaciones de los sujetos de investigación, y obtiene las aprobaciones apropiadas para permitir la revisión por pares y la replicación independiente de análisis.

Conoce las limitaciones legales sobre las garantías de privacidad y confidencialidad y no promete en exceso ni asume protecciones legales de privacidad y confidencialidad donde es posible que no se apliquen. Considera si se obtuvieron las aprobaciones apropiadas de los sujetos de investigación antes de participar en un estudio en el que participaron seres humanos u organizaciones, antes de analizar los datos de dicho estudio y al revisar los manuscritos para su publicación o uso interno.

El estadístico considera el tratamiento de los sujetos de la investigación (por ejemplo, acuerdos de confidencialidad, expectativas de privacidad, notificación, consentimiento, etc.) al evaluar la idoneidad de las fuentes de datos. Al contemplar la posibilidad de participar en un análisis de datos de una fuente en particular, se niega a hacerlo si la participación en el análisis podría ser interpretada razonablemente por personas que proporcionaron información como sancionando una violación de sus derechos.

1.3.6 ***Responsabilidades con los colegas del equipo de investigación***

La práctica científica y estadística a menudo se lleva a cabo en equipos formados por profesionales con diferentes estándares profesionales, el estadístico debe saber trabajar éticamente en este entorno. En este contexto el estadístico reconoce que otras profesiones tienen estándares y obligaciones, que las prácticas y los estándares de investigación pueden diferir entre disciplinas y que el estadístico no tiene obligaciones con los estándares de otras profesiones que entren en conflicto con estas directrices, asegura que todas las discusiones y reportes del diseño y análisis estadístico sean consistentes, evita comprometer la validez científica por conveniencia, además lucha por comenzar la transparencia en el diseño, ejecución y presentación de informes o presentación de todos los análisis.

1.3.7 ***Responsabilidades con respecto a las acusaciones de mala conducta***

El estadístico ético comprende la diferencia entre las prácticas científicas cuestionables y las prácticas que constituyen una mala conducta, evita ambas, pero sabe cómo deben manejarse cada una de ellas.

Evita tolerar o aparentar aprobar prácticas incompetentes o poco éticas en el análisis estadístico, reconoce que las diferencias de opinión y los errores honestos no constituyen mala conducta; merecen discusión, pero no acusación, conoce las definiciones y las operaciones afines con la mala conducta. Si está involucrado en una investigación de mala conducta, siga los procedimientos prescritos.

Mantiene la confidencialidad durante una investigación, pero revela los resultados de la investigación con honestidad a las partes y partes interesadas correspondientes una vez que están disponibles. Después de una investigación de mala conducta, apoya los esfuerzos apropiados de todos los involucrados, incluidos aquellos que informan del posible error científico o mala conducta, para reanudar sus carreras de la manera más normal posible. Evita y actúa para desalentar las represalias o el daño a la empleabilidad de quienes llaman la atención de manera responsable sobre posibles malas conductas.

CAPÍTULO 2
Consideraciones de la estadística descriptiva aplicada

2. Consideraciones de la estadística descriptiva aplicada

2.1 Estadística Descriptiva

Una vez recolecto los datos de las variables de estudio, y siguiendo los objetivos de investigación el siguiente paso es el análisis estadístico con el fin de describir la muestra, variables estudiadas, es decir; se procede al estudio de la relación entre variables dependientes e independientes (Mias, 2018)

La estadística descriptiva está enfocada a describir, resumir, visualizar la distribución de los datos, así como la organización, dispersión en relación a medidas de dispersión central; estima estadísticos descriptivos como la media, mediana, moda, rango, desviación típica (estándar), varianza, recuento de casos, porcentajes y percentiles. Luego los datos pueden ser representados por histogramas de frecuencia, diagramas de barra, gráficos circulares o piramidales entre otros.

La estadística descriptica es empleada para un análisis exploratorio del comportamiento de variables, datos, los principales estadísticos se resumen en:

- **Medidas de tendencia central:** son medidas de centralización, es un número ubicado hacia el centro de la distribución de valores de una serie de datos, ente estas medidas se destacan: moda (para todas las escalas), mediana (ordinal e intervalar), media (ordinal e intervalar).

- **Medidas de dispersión:** denominadas de variabilidad, representan el grado en que una distribución se comprime o estira, las principales medidas de dispersión son el rango o recorrido (incluye máximo y mínimo), desviación estándar y varianza.

- **Razones, proporciones y tasas:** en estadística la razón se utiliza como índices, la tasa es una medida de comparación de datos entre diferentes tiempos y poblaciones, las proporciones representan las frecuencias relativas, que estiman las probabilidades de ocurrencia de un evento.

- **Riesgo relativo:** es una medida de efecto que establece en términos relativos la relación entre la probabilidad de ocurrencia de un evento en el grupo expuesto y la probabilidad de que el mismo ocurra en el grupo no expuesto.

*2.2 **Descripción y mediciones de datos univariados***

Los datos son la representación de atributos o variables que representan hechos, al analizarlos y procesarlos estos se transforman en información (Martínez, 2020). En este contexto es importante conocer algunos elementos como las propiedades del operador sigma, interpretación de medidas de tendencia central y de dispersión para series simples y compuestas, representaciones gráficas.

*2.2.1 **Notación y utilización de la Sigma***

En el contexto estadístico generalmente es necesario calcular la suma de un conjunto de número, la letra griega sigma, representada por el símbolo Σ, el cual se utiliza en estadística para denotar un conjunto de elementos a sumarse.

Ejemplo 2.1

Una variable estudiada puede tomar distintos valores, y se la puede representar como x_i y a cada uno de sus valores como X_1, X_2, X_3, si a estas letras se les asigna un número, como, por ejemplo: 15, 10, 20; podrían ser una cantidad grande, al querer sumar estos valores que toma la variable se representaría de la siguiente manera:

$$X_1 + X_2 + X_3 \rightarrow 15 + 10 + 20$$

Utilizando la Sigma:

$$\sum_{i=1}^{4} X_{i= X_1+X_2+X_3=15+10+20=45}$$

$$\sum_{i=1}^{4} X_{i=45}$$

El símbolo se lee: la suma de los valores de X cuando va de uno hasta 4.

Ejemplo 2.2

Al suponer que n números multiplicados cada uno por 3, se busca sumar para obtener el resultado de n productos, se expresaría de la siguiente forma:

$$3X_1 + 3X_2 + 3X_3 + \cdots + 3X_n$$

Aplicando factor común la expresión queda de la siguiente manera:

$$3(X_1 + X_2 + X_3 + \cdots + X_n)$$

Utilizando sigma se reduce a:

$$\sum_{i=1}^{n} 3X_i$$

La constante 3 puede quedar fuera de la sigma y no altera el producto:

$$3\sum_{i=1}^{n} X_i$$

Ejemplo 2.3

Cuando cada uno de los valores que toma la variable hay que sumarle una constante c

$$X_1 + C, X_2 + C, X_3 + C \ldots, X_n + C$$

Para la suma de estos elementos se debería realizar lo siguiente:

$$(X_1 + C) + (X_2 + C) + (X_3 + C) + \cdots + (X_n + C)$$

Se reordena la expresión:

$$(X_1 + X_2 + X_3 + X_n) + (C + C + C + \cdots + C)$$

En este caso al conjunto X_i se puede abreviar y a la C se le expresa como el producto de n

$$\sum_{i=1}^{n}(X_i + C) = \sum_{i=n}^{n} X_i + \sum_{i=1}^{n} C = \sum_{i=1}^{n} X_i + n * C$$

2.2.2 ***Series de Tiempo***

Una serie estadística es un conjunto de número o términos que miden las variaciones de un fenómeno, existen series simples, de frecuencia, de clase y frecuencia.

2.2.2.1 ***Serie Simple***

Una serie simple representa un conjunto de números de las variaciones de un fenómeno particular que conforman un grupo. Ejemplo: las estaturas de 18 personas se muestran a continuación:

1.60	1.61	1.55
1.70	1.55	1.61
1.55	1.60	1.61
1.55	1.55	1.60
1.70	1.70	1.70
1.45	1.70	1.70

El conjunto de los datos es el resultado de las medidas de las estaturas de 15 personas, estos datos reciben el nombre de series simples, en este ejemplo la cantidad de datos es muy pequeña, pero cuando el número de datos es muy grande se puede deducir con la serie simple de características del fenómeno.

En este caso, se puede concluir que 3 personas miden 1.60 m, 5 personas miden 1.55 m, 3 personas miden 1.61 m, 6 personas miden 1.70 m, sólo 1 persona mide 1.45 m.

2.2.2.2 ***Series de frecuencia***

Una serie de frecuencia se refiere al número de veces que un valor o término se repite en una serie simple, es decir; que tan frecuente es el dato dentro de la serie.

Estatura (metros)	**Número de personas (frecuencia)**
1.70	6
1.60	3
1.61	3
1.55	5
1.45	1
Total	**18**

La suma de las frecuencias es igual al número de términos de la serie simple, las frecuencias son solamente la agrupación de los términos iguales de la serie, lo cual no modifica la cantidad original o total.

2.2.2.3 ***Series de clase y frecuencia***

Al tener un número muy grande de observaciones de un hecho o fenómeno, ejemplo 1000 estudiantes de la ESPCOH, en este caso se tendría la serie simple (1000 datos), pero para trabajar la información con mayor celeridad y facilidad se procede a formar una tabla de frecuencia, sin embargo, los datos serían muchos, en este caso se compacta más la información formando grupos o intervalos de clase.

Clase: se define como ciertos grupos o intervalos en los que se concentran las frecuencias observadas de la serie.

Los pasos principales en la elaboración de una distribución de clases y frecuencias para datos de una muestra son los siguientes:

- Establecer los intervalos de clases[2] en los que se agrupan los datos.
- Ordenarlos en clases mediante conteo de marcas.
- Contar el número de frecuencias de cada clase y anotarlo.
- Presentar los resultados en una tabla.

Para establecer los intervalos de clases de la serie es fundamental ejecutar los siguientes pasos:

- Determinar la amplitud de la variación de datos (recorrido de la variable), que es uno más la diferencia entre el mayor y menor de las puntuaciones.

  $$AV = 1 + (puntuación\ mayor - puntuación\ menor) \qquad (2.1)$$

- Decidir el número de clases o intervalos, existen diversos criterios para el número de clases, una aproximación empírica se la obtiene de $\sqrt{n}$, dónde n = número de datos; ejemplo n= 1000; $\sqrt{1000}$ = 31.62 ≈ 32, es importante redondear el resultado a un número entero; otro criterio es la decisión individual de cada investigador en función de su experiencia.

[2] Intervalo de clase: son intervalos en los que se agrupan y ordenan los valores observados, cada intervalo está delimitado por dos valores extremos llamados límites.

$$Nc = \sqrt{n} \qquad (2.2)$$

- Calcular el tamaño de cada clase: consiste en dividir la amplitud de la variación de los datos entre el número de clases, se debe dividir entre el número de clases redondeado.

$$Tc = Av/Nc \qquad (2.3)$$

- Realizar la tabulación de datos o determinar los límites de clases, se debe iniciar por el primer intervalo asegurando que se incluya la mínima observación. Las clases deben ser cercanas entre sí, de tal manera que no existan vacíos considerables. Los números que limitan una clase se denominan límite de clase.

Ejemplo 2.4

Se presentan los siguientes datos de las edades de 24 personas.

15	28	33
20	27	32
18	16	15
17	40	17
16	45	18
32	40	20
42	30	21
33	33	25

Se solicita construir una serie de clases y frecuencias.

Intervalos de clases en los que se agrupan los datos, se ordena los valores de mayor a menor y se construye una serie de frecuencias como se muestra a continuación:

Xi	Fi
15	2
16	2
17	2
18	2
20	2
21	1

25	1
27	1
28	1
30	1
32	2
33	3
40	2
42	1
45	1
Total	**24**

- Determinar la amplitud de variación de datos

$$AV = 1 + (puntuación\ mayor - puntuación\ menor)$$
$$AV = 1 + (45 - 15)$$
$$AV = 31$$

- Decidir el número de clases o intervalos

$$Nc = \sqrt{n}$$
$$Nc = \sqrt{24}$$
$$Nc = 4.8989 \approx 5$$

Al ser un resultado no entero se debe **redondear al inmediato superior, nunca al inferior.**

- Calcular la amplitud o tamaño de la clase

$$Tc = Av/Nc$$
$$Tc = 31/5$$
$$Tc = 6.2 \ \approx 7$$ [3]

El tamaño de la clase será de 7.

[3] Al considerar las reglas del redondeo, el valor correcto del tamaño de la clase sería 6 y no 7, sin embargo, si se dejan 6 se corre el riesgo de generar al final otra clase.

- Efectuar la tabulación o determinación de los límites o fronteras de clases

Clases		Fi
15	22	11
22	29	3
29	36	6
36	43	3
43	50	1

Es importante mencionar la existencia de tablas de clase y frecuencias abiertas en su extremo inferior, superior o ambos.

*2.2.3 **Medidas de tendencia central***

La medida de tendencia central denominada parámetro de tendencia central, es un número que se ubica hacia el centro de la distribución de los valores de una serie de observaciones en la que se encuentra ubicada en el conjunto de datos, en este contexto los estadísticos de tendencia central son: moda, mediana y media; se utilizan para describir características de centralidad de las variables (Puente, 2018).

*2.2.3.1 **Moda***

La moda es el valor o categoría de la variable que se repiten más veces o que tiene una mayor frecuencia, esta moda es considerada como absoluta; puede haber otras modas que se denominan relativas y su característica es el valor de la variable que tiene una frecuencia mayor que los valores anteriores y posteriores.

Este estadístico se puede utilizar con las variables de medida nominal, orinal, intervalo y razón, el cálculo se realiza a partir de las tablas de frecuencia, el resultado puede variar de acuerdo al agrupamiento de los intervalos, en las variables categóricas el valor de la moda se calcula por observación de la tabla de frecuencia.

Es importante mencionar que no todos los conjuntos de puntuaciones tienen moda, a veces no hay moda o puede existir más de una, a continuación, se presentan algunos ejemplos de moda.

Ejemplo 2.5

Se tiene las siguientes puntuaciones 2,6,6,7,8,8,8,9,10; la moda en este caso es 8, porque es la puntuación que más se repite en la serie.

Ejemplo 2.6

En el caso de que todas las puntuaciones o elementos de un grupo tienen la misma frecuencia, es decir; se repiten igual número de veces, en este caso no hay moda. Ejemplo 0.5, 0.5, 1.1, 1.1, 2.5, 2.5, en este grupo no hay moda.

Ejemplo 2.7

En el caso de que dos puntuaciones consecutivas tengan igual frecuencia, la cual es mayor que cualquier otra puntuación, la moda se considera como el promedio de las dos puntuaciones, a continuación, se presenta el siguiente conjunto de puntuaciones 0, 1, 1, 1, 2, 2, 2, 3, 4; en este caso los números que son consecutivos y tienen mayores repeticiones son 1 y 2, sumados son igual a 3; este valor dividido para 2 la moda será 1.5.

Ejemplo 2.8

En el caso de un grupo de datos existen dos datos no consecutivos y tienen la misma frecuencia, y está es mayor que cualquier frecuencia del resto de datos, se evidencia la existencia de dos modas. 9,9,10,11,12,12,13; tanto 9 como 13 se consideran como modas y en este caso el grupo se llama bimodal.

Moda para una serie de clases y frecuencias

En este tipo de series se aplica el mismo criterio para la moda de los casos anteriores, es decir; se busca el valor que se repite el mayor número de veces, solo que en este escenario los valores se encuentran agrupados en clases o intervalos.

Proceso de cálculo de la moda

1. Se ordena la tabla de frecuencia de menor a mayor.
2. La moda está en aquella categoría que tenga mayor número de casos.
3. Se calcula el valor.

$$M_o = L_{i-1} + \frac{D_1}{D_1 + D_2} * a \qquad (2.4)$$

Dónde:

L_{i-1} = Límite inferior del intervalo modal

a = amplitud de los intervalos

D_1 = Diferencia de frecuencia absoluta entre intervalo modal y el anterior

D_2 = Diferencia de frecuencia absoluta entre intervalo modal y el siguiente

Ejemplo 2.9

A continuación, se presenta una tabla de frecuencia del peso de 95 estudiantes de la ESPOCH

Intervalos	Frecuencia
45 – 55 kg	23
55 – 65 kg	21
65 – 75 kg	32
75 – 85 kg	19
Total	**95**

1. Se ordena la tabla de frecuencia de menor a mayor.

Intervalos	Frecuencia
65 – 75 kg	32
45 – 55 kg	23
55 – 65 kg	21
75 – 85 kg	19
Total	**95**

2. La moda está en aquella categoría que tenga mayor número de casos. El intervalo que mayor número de casos contiene es el de 65 - 75 kg.

3. Se calcula el valor.

$$M_o = L_{i-1} + \frac{D_1}{D_1 + D_2} * a$$

Datos

L_{i-1}= 65

D_1 = (32-21)

D_2 = (32-19)

a = 10 kg

$$M_o = 65 + \frac{(32-21)}{(32-21)+(32-19)} * 10$$

$$M_o = 69.75\ Kg$$

*2.2.3.2 **Mediana***

La mediana en un conjunto de datos es el valor que se encuentra a la mitad de los otros valores, es decir que, al ordenar los números de menor a mayor, este valor se encuentra en el medio, algunas de las características de esta medida de tendencia central son:

- Las operaciones para su cálculo son sencillas de realizar.
- La mediana no depende de los valores de las variables, solamente de su orden.
- Generalmente sus valores son enteros.
- Se puede calcular a pesar que los número anteriores y posteriores no tengan límites.

Mediana para una serie simple

Para encontrar la mediana es necesario primero ordenar los valores de menor a mayor, luego se debe separar la mitad de valores para obtener la mediana, el procedimiento para identificar la mediana se resume en los siguiente:

1. Ordenar o clasificar los valores de mayor a menor o de menor a mayor.
2. Contar los valores para determinar si existe un número de datos par o impar.
3. En el caso de casos impar la mediana es el valor central, pero si el número es un valor par, la mediana es el promedio de los valores centrales.
4. La posición de la mediana se determina con la fórmula ${(n+1)}/{2}$, esta fórmula determina la posición de la mediana, pero no su valor; este se halla al encontrar la posición de la mediana.

Ejemplo 2.10
Encontrar la mediana del grupo 2, 3, 4, 5, 6.

Primero se define la posición de la mediana con fórmula $\frac{n+1}{2}$; por tanto para cinco valores, el lugar donde se encuentra la mediana es 3, en este escenario la tercera posición de la serie que corresponde al valor de 4 y es la mediana.

En este ejemplo al ser número impar la mediana es el dato que está en la posición central o del medio.

Ejemplo 2.11

Se presentan los siguientes datos 9, 10, 11, 12, 13,14.

Al aplicar la fórmula $\frac{n+1}{2}$; es 3.5, en esta posición se encuentran dos valores intermedios, en este caso es el 11 y 12; sumando y calculando el promedio de estos valores la mediana es 11.5.

Mediana para una serie de clases y frecuencia

La mediana es el valor de la variable que deja por debajo al 50% de los casos, la mediana se puede utilizar con variables que al menos tengan un nivel de medida ordinal, pero su uso es más adecuado con intervalos, en variables nominales no se pueden utilizar porque no se pueden ordenar los casos.

La fórmula de la mediana es una derivación de la fórmula de los percentiles y se desarrolla con una regla de tres.

Proceso de cálculo

1. Se ordena la tabla de frecuencias de menor a mayor.
2. Se calculan las frecuencias acumuladas
3. Se divide el número de casos por dos y el intervalo que los contiene se denomina intervalo crítico (IC).
4. Entonces se procede a calcular el valor exacto de la mediana

$$M_e = L_{i-1} + \frac{\frac{N}{2} - N_{i-1}}{n_i} * a \qquad (2.5)$$

Dónde:

M_e = mediana
L_{i-1} = Límite inferior del IC
N/2 = mitad de los casos
N_{i-1} = Total de casos bajo el IC
a = amplitud del IC
n_i = Frecuencia del intervalo crítico

Ejemplo 2.12

Intervalos	Frecuencia	Frecuencia acumulada
45 – 55 kg	23	23
55 – 65 kg	21	44
65 – 75 kg	32	76
75 – 85 kg	19	95
Total	**95**	

- Se presenta la tabla de frecuencia ordenada de acuerdo a los intervalos, se determina la frecuencia acumulada.
- Se divide el número de casos para 2; es decir; 95/2= 47.5, el intervalo que corresponde es 65 – 75 kg de acuerdo a las frecuencias acumuladas.
- Se reemplaza los valores en la fórmula para calcular el valor de la mediana

$$M_e = L_{i-1} + \frac{\frac{N}{2} - N_{i-1}}{n_i} * a$$

$$M_e = 65 + \frac{\frac{95}{2} - 32}{44} * 10$$

$$M_e = 66.09\ Kg$$

Ejemplo 2.13

Intervalos Estatura cm	Frecuencia	Frecuencia acumulada
118 – 126	3	3
127 – 135	5	8
136 – 144	9	17
145 – 153	12	29
154 - 162	5	34
163 - 171	4	38
172 - 180	2	40
Total	**40**	

- Se presenta la tabla de frecuencia ordenada de acuerdo a los intervalos, se determina la frecuencia acumulada.

- Se divide el número de casos para 2; es decir; 40/2= 20, la suma de frecuencias de las tres primeras clases es de 17, para llegar al valor de 20, se necesita tres elementos que se tomará de los 12 que se encuentran en la cuarta clase.
- Se reemplaza los valores en la fórmula para calcular el valor de la mediana.

$$M_e = L_{i-1} + \frac{\frac{N}{2} - N_{i-1}}{n_i} * a$$

$$M_e = 145 + \frac{\frac{40}{2} - 17}{12} * 8$$

$$M_e = 149$$

2.2.3.3 ***Media***

La media también conocida como promedio, es el valor que se obtiene al dividir la suma de todos los números que conforman el conglomerado, algunas de las características de esta medida de tendencia central son:

- Considerar todos los elementos.
- El numerador de la fórmula es la cantidad de valores.

La media aritmética es el valor obtenido al sumar todos los datos y dividir el resultado entre el número total de datos. Denotamos la media con el símbolo $\overline{X}$

Media aritmética para una serie simple

La media se calcula sumando los valores de la serie, de los cuales se obtiene el promedio, dividiendo entre el número de datos que se toma en cuenta en la suma, la media de las n medidas se determina de la siguiente manera:

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n} \quad (2.6)$$

Ejemplo 2.14

Se presentan los siguientes datos: 8, 5, 6, 7, 10, 11, 15

$$\bar{X} = \frac{8 + 5 + 6 + 7 + 10 + 11 + 15}{7}$$

$$\bar{X} = \frac{62}{7}$$

$$\bar{X} = 8.86$$

Ejemplo 2.15

Se recolectan las estaturas de 10 estudiantes de un curso la carrera de Química de la ESPOCH, se muestra a continuación: 1.70; 1.60; 1.55; 1.60; 1.75; 1.63; 1.69; 1.59; 1.61; 1.68

$$\bar{X} = \frac{1.70 + 1.60 + 1.55 + 1.60 + 1.75 + 1.63 + 1.69 + 1.59 + 1.61 + 1.68}{10}$$

$$\bar{X} = \frac{16.4}{10}$$

$$\bar{X} = 1.64$$

La media de la estatura de los 10 estudiantes de un curso de la carrera de Química de la ESPOCH es de 1.64 m

Media aritmética para una serie de frecuencias

La media en el caso de una serie de datos organizados por frecuencias se calcula de la siguiente manera:

$$\bar{X} = \frac{\sum_{i=1}^{n} f_i Y_i}{n} \qquad (2.7)$$

Dónde

$\bar{X}$ = Media

Y_1 = marca de clase

n = Total de los casos

Ejemplo 2.16

Intervalos	Frecuencia fi	Marca de Clases (yi)	fi * yi
45 – 55 kg	23	50	1150
55 – 65 kg	21	60	1260
65 – 75 kg	32	70	2240
75 – 85 kg	19	80	1520
Total	**95**		**6170**

$$\bar{X} = \frac{6170}{95} = 64.95\ kg$$

Ejemplo 2.17

De acuerdo a lo siguientes datos determine la media

Intervalos	Frecuencia
50-55	6
55-60	13
60-65	9
65-70	8
70-75	4
n=	**40**

Con los datos obtenidos se procede al cálculo de la marca de clase y la frecuencia por la variable como se muestra a continuación:

Intervalos	Frecuencia fi	Marca de Clases (yi)	fi * yi
50-55	6	52,5	315
55-60	13	57,5	747,5
60-65	9	62,5	562,5
65-70	8	67,5	540
70-75	4	75,5	302
Total	**40**		**2467**

$$\bar{X} = \frac{2467}{40} = 61.68$$

2.2.3.4 ***Comparación moda, mediana y media***

El cálculo de la media, mediana y moda se traduce en un procedimiento mecánico, a continuación, se presenta una comparación entre estas tres medidas donde se determinan sus ventajas y limitaciones.

Tabla 2-1. Comparación de las medidas de tendencia central.

Medida	Concepto	Ventaja	Desventaja
Media	Es el dato que se define como el promedio de todas las puntuaciones	- Refleja el valor - Propiedades matemáticas atractivas	- Puede ser excesivamente influida por los extremos
Mediana	Es el valor que divide a la serie en dos partes iguales	- Menos sensible a los extremos que la media	- Difícil de determinar si hay una gran cantidad de datos
Moda	Es el valor con frecuencia más alta	- Valor típico, más valores reunidos en este punto que en cualquier otros	- No se presta para el análisis matemático. - No en todo conjunto de datos se puede calcular la moda

Nota: recolección de varios autores

2.2.3.5 ***Otras medidas de posición***

En estadística las medidas de posición son indicadores estadísticos que permiten resumir los dato en uno solo, o dividir su distribución en intervalos de igual tamaño, entre las medidas más usuales están:

- **Cuartil:** es un punto en una escala numérica que agrupa una serie de datos dividiéndoles en cuatro partes iguales, existen tres cuartiles (Q1, Q2, Q3).

- **Quintil:** es un punto en una escala numérica que agrupa una serie de observaciones, dividiendo la distribución en cinco partes; por lo tanto, hay cuatro quintiles, es de menor uso que el cuartil.

- **Decil:** es un punto en la escala numérica que divide a los datos en diez partes iguales, existen nueve deciles y se representan de la siguiente manera D1, D2, D3, D4, D5, D6, D7, D8, D9; el D5 corresponde a la mediana.

- **Percentil:** es un estadístico de tendencia central, divide al conjunto de datos y su distribución en cien partes iguales, hay 99 percentiles, tiene a su vez una equivalencia con los deciles y cuartiles.

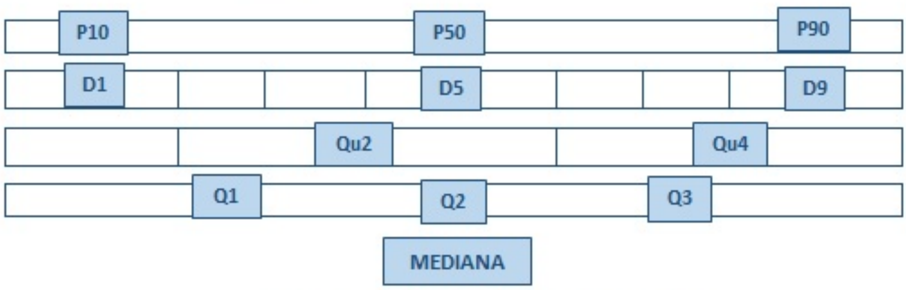


Figura 2-1. Representación gráfica y equivalencias de cuartiles, deciles, percentiles.

Cuartil

$$\mathrm{Q}_1 = \frac{\propto (\mathrm{n}+1)}{4}$$

Decil

$$\mathrm{D}_1 = \frac{\propto (\mathrm{n}+1)}{10}$$

Percentil

$$\mathrm{P}_1 = \frac{\propto (\mathrm{n}+1)}{100}$$

Dónde:
Q_1 = Cuartil.
D_1 = Decil.
P_1 = Decil.
$\propto$ = El Q, D, P que se puede calcular.
n = número de datos.

Ejemplo 2.18
Cuartiles para una serie de datos simple:

Se presentan los siguientes datos 1,4,5,7,10,13,14,16,20,23,25

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

$$Q_1 = \frac{1(11+1)}{4} = \frac{12}{4} = 3$$

$$Q_2 = \frac{2(11+1)}{4} = \frac{12}{2} = 6$$
$$Q_3 = \frac{3(11+1)}{4} = \frac{3*12}{4} = 9$$

Para determinar el valor del Q_1, su posición es el 3, en este caso el cuartil 1 corresponde a 5; Q2 = 13 y el Q3 = 20.

Es importante mencionar que el valor del cuartil, decil o percentil indica el lugar dentro de la serie de datos en el que se encuentra cada uno de ellos; es importante ordenar los datos de forma ascendente o descendente.

Con los siguientes datos: 1,4,5,7,10,13,14,16,20,23,25, calcular el decil 1, 5,9

$$\mathrm{D}_1 = \frac{\propto (\mathrm{n}+1)}{10}$$

$$\mathrm{D}_1 = \frac{1(11+1)}{10} = \frac{12}{10} = 1.2$$

$$\mathrm{D}_5 = \frac{5(11+1)}{10} = \frac{5*12}{10} = 6$$

$$\mathrm{D}_9 = \frac{9(11+1)}{10} = \frac{9*12}{10} = 10.8$$

Para determinar el valor de D1, su posición es 1.2, es un número con decimales, entonces el decil 1 corresponde al valor entre la posición 1 y 2 (1+4)/2 = 2.5

El D2 corresponde a la posición 6, es decir; 13; el D9 corresponde a la posición entre 10 y 11 por ser un número decimal (23+25)/2=24.

Con los siguientes datos: 1,4,5,7,10,13,14,16,20,23,25, calcular el percentil 10, 50, 90

$$\mathrm{P}_{10} = \frac{\propto (\mathrm{n}+1)}{100}$$

$$\mathrm{P}_{10} = \frac{10(11+1)}{100} = \frac{120}{100} = 1.2$$

$$P_{50} = \frac{50(11+1)}{100} = \frac{600}{100} = 6$$

$$P_{90} = \frac{90(11+1)}{100} = \frac{108}{100} = 10.8$$

D1 (decil 1) = P10 (percentil 10)
D5 (decil 5) = P50 (percentil 50) = Q2 (cuartil 2) = Me (mediana)
D8 (decil 8) = P80 (percentil 80)

2.2.4 ***Medidas de dispersión***

Desde el punto de vista estadístico las medidas de dispersión representan el grado en que una distribución se extiende o se comprime (Gamero, 2017); este tipo de medidas tiene como objetivo expresar en un solo valor la dispersión que tiene un conjunto de datos, entre las medidas más utilizadas son: rango de variación, varianza, desviación estándar, coeficiente de variación.

2.2.4.1 ***Rango o recorrido***

El rango (R) o recorrida de una variable X en estadística representa la diferencia entre el valor máximo y el valor mínimo de un conjunto de datos, matemáticamente se representa:

$$R = \max(X_1) - \min(X_1) \tag{2.9}$$

El recorrido viene expresado en igual unidad de medida que la variable de análisis, considerando como ejemplo los siguientes datos: 19, 25, 36; la diferencia entre el número mayor y menor es 36 – 19 = 17.

De manera general el rango se expresa al establecer la diferencia entre el número mayor y el número menor de un conjunto de datos, una de las ventajas de utilizar el rango como una medida de dispersión es el hecho de su fácil cálculo, aun cuando se trate de un conjunto de datos muy amplio, su principal limitación es que considera datos extremos, pudiendo generar una falsa imagen del conjunto, esta limitación genera que se utilice otros recorridos como los que se muestran en la tabla 2-2

Tabla 2-1. Recorridos intercuantílicos.

Recorrido	Descripción
Recorrido intercentílico	Diferencia entre el último y primer percentil, lo que excluye para el cálculo del recorrido del 2% de las observaciones, que admite valores extremos $R_c = C_{99} - C_1$ (2.10)
Recorrido interdecílico	Diferencia entre el último y el primer decil, razón por la cual se elimina para el cálculo del 20% de las observaciones. $R_D = D_9 - D_1$ (2.11)
Recorrido intercuartílico	Diferencia entre el último y el primer cuartil, razón por la cual se elimina para el cálculo del 50% de las observaciones. $R_Q = Q_3 - Q_1$ (2.12)

En este contexto los recorridos intecuantílicos eliminan observaciones atípicas para estudiar la dispersión de la distribución de la zona central, contienen el 98%, 80% y el 50% de los datos observados, por esta razón se crea una ventaja frente al recorrido (R) de prescindir valores extremos, sin embargo; aún está presente el inconveniente general de no considerar toda la información disponible.

*2.2.4.2 **Varianza***

La varianza es una medida de dispersión, representa la variabilidad de un conjunto de datos en relación a la media de estos; es una medida de variación igual al cuadrado de la desviación estándar, en el contexto estadístico existen dos tipos de varianza: varianza poblacional y varianza muestral.

La varianza de una muestra su cálculo es similar al de la desviación estándar con pequeñas diferencias: 1. Las desviaciones se elevan al cuadrado antes de ser sumadas. 2. Se obtiene el promedio utilizando n-1 en lugar de n, porque proporciona un mejor cálculo de la varianza.

Varianza de una población

$$\sigma^2 = \frac{\sum_{i=1}^{N}(X_i - \mu)^2}{n-1} \qquad (2.13)$$

Dónde
X_i = Valores de X.
μ= Valor de la media poblacional.
n = número de elementos.

Varianza de una muestra para series simples

$$S^2 = \frac{\sum_{i=1}^{N}(X_i - \bar{X})^2}{n-1} \qquad (2.14)$$

Dónde
X_i = Valores de X.
$\bar{X}$= Valor de la media.
n = número de elementos.

Varianza de una muestra para series de frecuencia

$$S^2 = \frac{\sum_{i=1}^{N} f_i(X_i - \bar{X})^2}{n-1} \qquad (2.15)$$

Dónde
X_i = Valores de X.
$\bar{X}$= Valor de la media.
n = Número de elementos.
f_i = Frecuencias.

Varianza de una muestra para series de frecuencia e intervalos

$$S^2 = \frac{\sum_{i=1}^{N} f_i(y_i - \bar{X})^2}{n-1} \qquad (2.16)$$

Dónde
Y_i = Valores de la marca de clase o punto medio de la marca de clase.
$\bar{X}$= Valor de la media.
n = Número de elementos.
f_i = Frecuencias.

Ejemplo 2.19
Varianza para una serie simple:

Se presentan las edades de 5 estudiantes (14, 16, 18, 20, 22), calcule la varianza para la muestra.

1. Calcular la media.

En este caso son 5 datos por lo tanto n = 5

$$\bar{X} = \frac{14 + 16 + 18 + 20 + 22}{5}$$

$$\bar{X} = \frac{90}{5}$$

$$\bar{X} = 18\ años$$

2. Restar la media a cada valor del conjunto de datos, originando las desviaciones respectivas.

Edades (X_i)	$\boldsymbol{X_i - \bar{X}}$
14	14-18 = -4
16	16-18 = -2
18	18-18 = 0
20	20-18 = 2
22	22-18 = 4
Σ 90	

Desviación estándar

3. Elevar al cuadrado cada desviación

Edades (X_i)	$\boldsymbol{X_i - \bar{X}}$	$\boldsymbol{(X_i - \bar{X})^2}$
14	14-18 = -4	16
16	16-18 = -2	4
18	18-18 = 0	0
20	20-18 = 2	4
22	22-18 = 4	16
Σ 90		**Σ 40**

4. Dividir el resultado de la sumatoria entre n-1, si todas las observaciones equivalen a una población.

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

$$S^2 = \frac{40}{5-1}$$

$$S^2 = 10$$

Ejemplo 2.20
Varianza para una serie de frecuencia:

Edades	Frecuencias
14	5
15	2
16	6
17	7
19	2
20	1
Σ	**23**

1. Calcular la media.

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

Edades	Frecuencias	fi *Xi
14	5	70
15	2	30
16	6	96
17	7	119
19	2	38
20	1	20
Σ	**23**	**373**

$$\bar{X} = \frac{373}{23}$$

$$\bar{X} = 16.217$$

2. Restar la media a cada valor del conjunto de datos y con ello se origina las desviaciones con respecto a la media.

Edades	Frecuencias	fi*Xi	$X_i - \overline{X}$
14	5	70	-2,22
15	2	30	-1,22
16	6	96	-0,22
17	7	119	0,78
19	2	38	2,78
20	1	20	3,78
Σ	**23**	**373**	**3,70**

14-16.217

3. Elevar al cuadrado cada desviación.

Edades	Frecuencias	fi*Xi	$X_i - \overline{X}$	$(X_i - \overline{X})^2$
14	5	70	-2,22	4,92
15	2	30	-1,22	1,48
16	6	96	-0,22	0,05
17	7	119	0,78	0,61
19	2	38	2,78	7,74
20	1	20	3,78	14,31
Σ	**23**	**373**	**3,70**	

4. Multiplicar la frecuencia correspondiente a cada una de las desviaciones al cuadrado y sumarlas.

Edades	Frecuencias	fi*Xi	$X_i - \overline{X}$	$(X_i - \overline{X})^2$	$f_i(X_i - \overline{X})^2$
14	5	70	-2,22	4,92	24,58
15	2	30	-1,22	1,48	2,96
16	6	96	-0,22	0,05	0,28
17	7	119	0,78	0,61	4,29
19	2	38	2,78	7,74	15,49
20	1	20	3,78	14,31	14,31
Σ	**23**	**373**	**3,70**		**61,91**

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

$$S^2 = \frac{61.91}{23-1}$$

$$S^2 = 2{,}81$$

Ejemplo 2.21

Varianza para una serie de clase y frecuencia:

Intervalos	Frecuencia fi
45 – 55 kg	23
55 – 65 kg	21
65 – 75 kg	32
75 – 85 kg	19
Total	**95**

1. Agregar una columna en donde se calcule la marca de clase, es decir; los valores de y_i que ocuparan el lugar de X en la fórmula.

Intervalos	Frecuencia fi	Marca de Clases (yi)
45 – 55 kg	23	50
55 – 65 kg	21	60
65 – 75 kg	32	70
75 – 85 kg	19	80
Total	**95**	

2. Agregar una columna donde se coloquen las frecuencias de la serie.

Intervalos	Frecuencia fi	Marca de Clases (yi)	fi * yi
45 – 55 kg	23	50	1150
55 – 65 kg	21	60	1260
65 – 75 kg	32	70	2240
75 – 85 kg	19	80	1520
Total	**95**		**6170**

3. Calcular la media.

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

$$\bar{X} = \frac{6170}{95} = 64.95\ kg$$

4. Restar la media a cada marca de clases, lo cual origina las desviaciones respecto a la media.

Intervalos	**Frecuencia fi**	**Marca de Clases (yi)**	**fi * yi**	$\mathbf{y_i - \bar{X}}$
45 – 55 kg	23	50	1150	-14,95
55 – 65 kg	21	60	1260	60
65 – 75 kg	32	70	2240	70
75 – 85 kg	19	80	1520	80
Total	**95**		**6170**	

5. Elevar al cuadrado cada una de las desviaciones con r.

Intervalos	**Frecuencia fi**	**Marca de Clases (yi)**	**fi * yi**	$\mathbf{y_i - \bar{X}}$	$\mathbf{(y_i - \bar{X})^2}$
45 – 55 kg	23	50	1150	-14,95	223,42
55 – 65 kg	21	60	1260	60	3600
65 – 75 kg	32	70	2240	70	4900
75 – 85 kg	19	80	1520	80	6400
Total	**95**		**6170**		**15123,42**

6. Multiplicar cada desviación al cuadrado por la frecuencia.

Intervalos	**Frecuencia fi**	**Marca de Clases (yi)**	**fi * yi**	$\mathbf{y_i - \bar{X}}$	$\mathbf{(y_i - \bar{X})^2}$	$\mathbf{f_1(y_i - \bar{X})^2}$
45 – 55 kg	23	50	1150	-14,95	223,42	5138,75
55 – 65 kg	21	60	1260	60	3600	75600
65 – 75 kg	32	70	2240	70	4900	156800
75 – 85 kg	19	80	1520	80	6400	121600
Total	**95**		**6170**		**15123,42**	**359138,75**

7. Aplicar la fórmula de la varianza.

$$S^2 = \frac{35913.78}{95 - 1}$$

$$S^2 = 3820.62$$

2.2.4.3 ***Desviación estándar***

La desviación estándar también denominada desviación típica, representa la raíz cuadrada de la varianza, esta medida es utilizada para cuantificar la variación o la dispersión de un conjunto de datos numéricos. Una desviación típica baja representa que la mayor parte de datos están agrupados cerca de su media, si la desviación típica es alta muestra que los datos se extienden en un rango más amplio.

$$S = \sqrt{S^2} \tag{2.17}$$

Dónde
S = desviación típica.
S^2= Varianza.

Ejemplo 2.22
Del ejemplo 2.21, calcule la desviación estándar

$$S = \sqrt{3820.62}$$

$$S = 61.81$$

2.2.4.4 ***Coeficiente de Variación***

El coeficiente de variación es la estandarización de la desviación típica al eliminar la unidad de medida de la variable, este coeficiente también recibe el nombre de coeficiente de variación de Pearson. Su cálculo se basa en dividir la desviación típica para la media de un conjunto de datos.

$$CV = \frac{S}{\bar{X}} \tag{2.18}$$

El coeficiente de variación puede ser expresado como CV o R

Ejemplo 2.23
Tomando los datos del ejemplo 2.21, calcule el coeficiente de variación

$$\bar{X} = \frac{6170}{95} = 64.95\ kg$$

$$S = 61.81$$

$$CV = \frac{61.81}{64.95} = 0.9516 * 100 = 95.16\%$$

El coeficiente de variación establece la relación entre la desviación típica y la media, cuando más se aproxime a cero, menor será la dispersión de la variable, no tiene unidad de medida porque las unidades de medida del numerador y el denominador al ser las mismas se eliminan.

2.2.5 Medidas de asimetría y curtosis

La asimetría y cutosis informan sobre la forma de la distribución de una variable, estas medidas permiten saber la homogeneidad de los datos sin necesidad de una gráfica (Gamero, 2017).

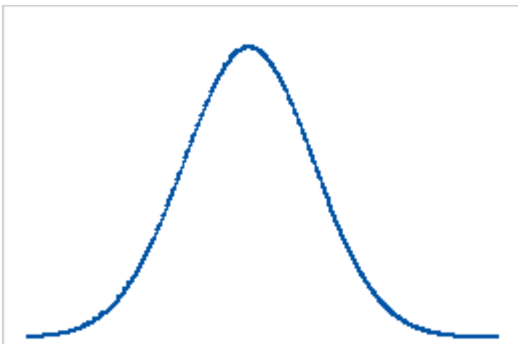


Figura 2-1. Representación Gráfica de la asimetría y curtosis.

2.2.5.1 Medidas de asimetría

La asimetría es un indicador que permite establecer el grado de simetría que presenta una distribución de probabilidades de una variable aleatoria, sin necesidad de representación gráfica; existirá una asimetría perfecta cuando los valores de la media, mediana y moda coincidan.

La asimetría es positiva si la media es mayor a la mediana y es negativa si la mediana es mayor a la media., es un valor relativo y por lo tanto puede utilizarse con fines de comparación.

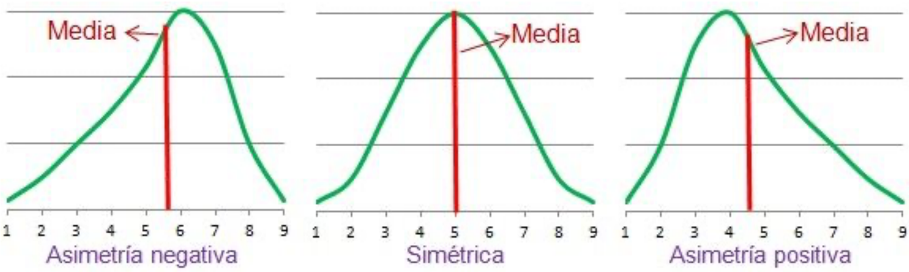


Figura 2-1. Tipos de Asimetría.

$$CA = \frac{\sum_{i=1}^{N}(X_1 - \bar{X})^3}{N * S^3} \quad (2.19)$$

Si el coeficiente de asimetría es = 0 entonces hay una distribución simétrica, si el mayor a cero es una distribución asimétrica positiva y si es menor a 0 es una distribución asimétrica negativa.

En la actualidad con los programas estadísticos se puede calcular esta medida de manera automática en softwares como STATA, SPSS.

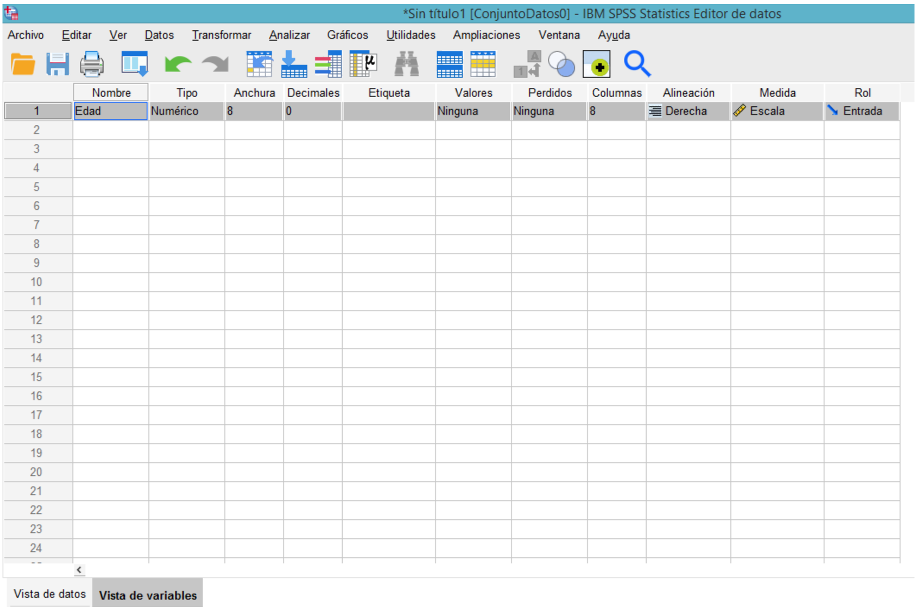
Ejemplo 2.24

Se presentan los siguientes datos, calcule los estadísticos descriptivos utilizando un programa estadístico[4]

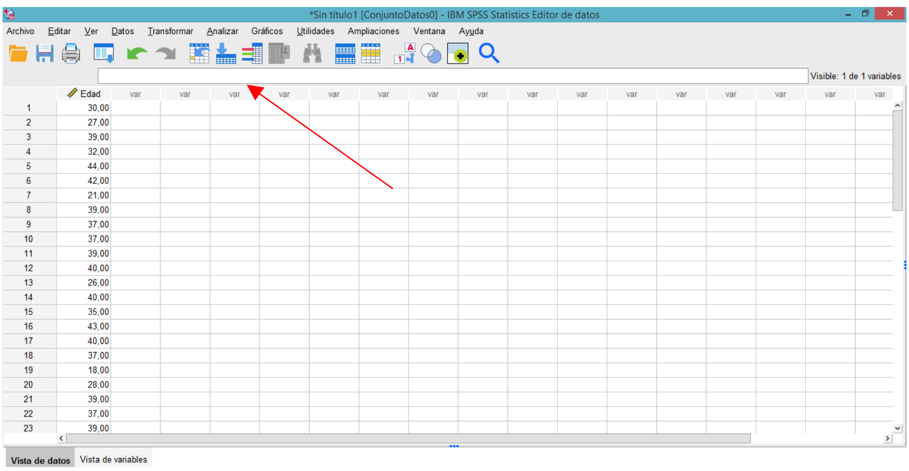
30	32	21	37	26	43	18	37	39	42	40
27	44	39	39	40	40	28	39	25	39	22
39	42	37	40	35	37	39	37	23	35	40

[4] Programa SPSS 25.0, versión gratuita obtenida de: https://www.ibm.com/support/pages/downloading-ibm-spss-statistics-25

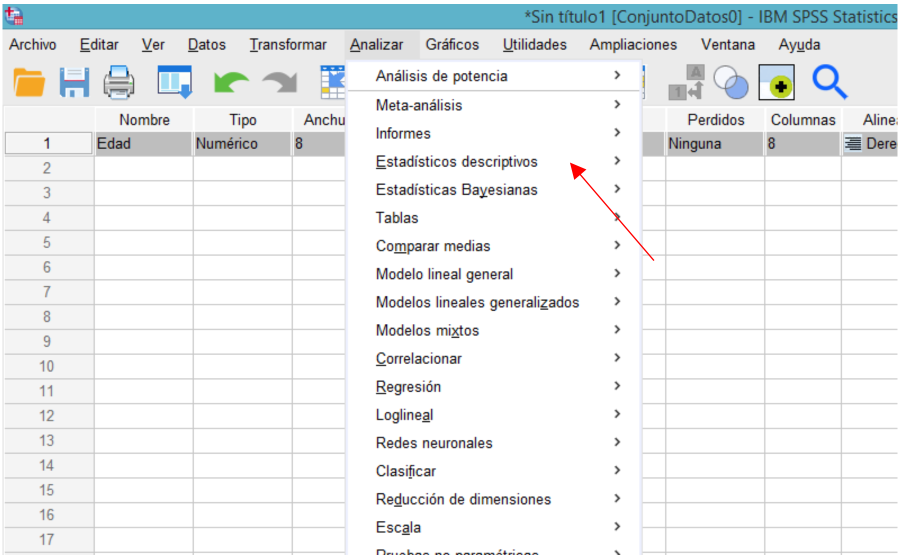
1. Crear la variable edad.



Al crear la variable identificar qué tipo de variable es, en este caso es numérica; una vez creada la variable ingresar los datos de la variable como se muestra a continuación:



Una vez ingresado los datos clic en analizar, se despliega una serie de pestañas como se muestra a continuación:

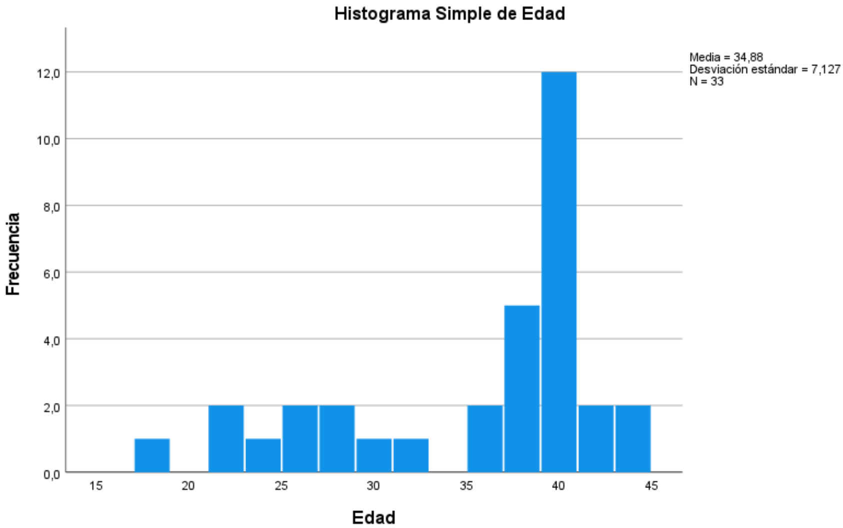


Seleccionar estadísticos descriptivos, incluyendo medida de asimetría.

Estadísticos

Edad		
N	Válido	33
	Perdidos	0
Media		34,8788
Mediana		37,0000
Moda		39,00
Desv. Desviación		7,12723
Varianza		50,797
Asimetría		-,982
Error estándar de asimetría		,409
Rango		26,00

En este caso la asimetría es de -0.982, se puede concluir que al ser menor que cero, se indica el sesgo es negativo, es decir; la distribución presenta una cola larga hacia la izquierda, esto se puede comprobar al graficar los valores de las edades de acuerdo a sus frecuencias como se muestra en la siguiente gráfica.



Como se muestra en la gráfica los datos tienen con un sesgo negativo, es decir; hacia la izquierda.

*2.2.5.2 **Medida de curtosis***

La curtosis de una variable estadística es una característica de la forma de la distribución de frecuencias o probabilidades (Chuhutin & Jespersen, 2017), es una medida estadística que define el grado de concentración que presentan los valores de una variable alrededor de la zona central de la distribución de frecuencias, se le denomina también medida de apuntamiento, en la figura 2-3 se muestran los tipos de curtosis que se pueden presentar.

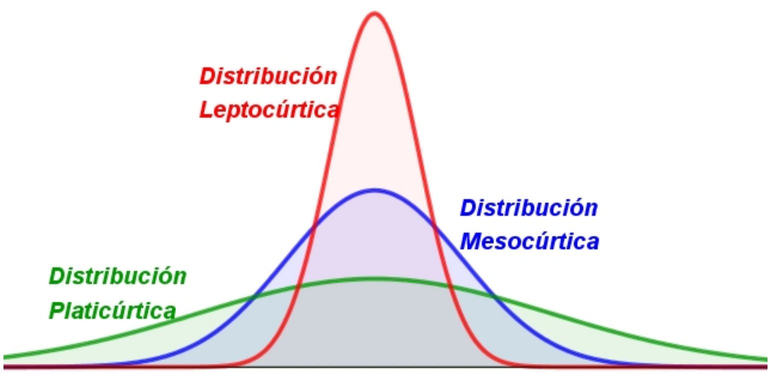


Figura 2-1. Tipo de Curtosis.

- **Leptocúrtica:** cuando existe un gran número de valores concentrado en torno a su media (a> 3).
- **Mesocúrtica:** cuando hay una concentración normal de los valores entorno a su media (a=3).
- **Platicurtica:** se origina con la existencia de una baja concentración de valores en torno a su media (a<3).

Dependiendo del tipo de datos la fórmula de cálculo varía.

$$\propto_4 = \frac{1}{N} * \frac{\sum_{i=1}^{n}(X_1 - \bar{X})^4}{S^4} \qquad (2.20)$$

Dónde:
S^4 = desviación estándar elevada a la cuarta potencia.
$\bar{X}$= media
N = número de elementos del conjunto de datos

Ejemplo 2.25

Danilo trabaja en una compañía de mantenimiento ABC S.A, su trabajo consiste en brindar servicio de cuidados de jardinería, a continuación, se muestra el número de citas que hizo en los últimos tres días de trabajo de 8 horas laborables.

9	5	2	6	2	8	3	7
3	6	3	4	7	5	2	4
5	4	8	3	8	4	2	2

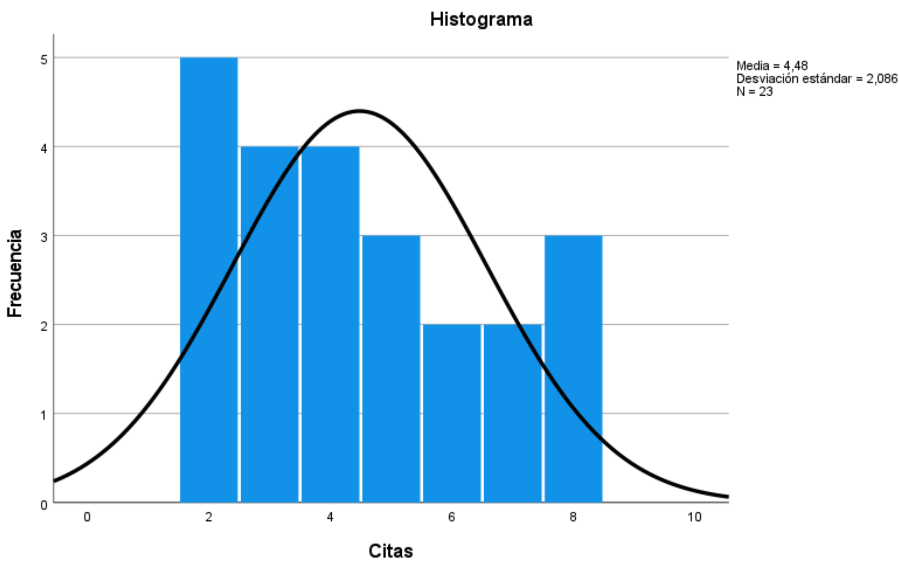
1. Tabla de frecuencia.

Citas					
		Frecuencia	**Porcentaje**	**Porcentaje válido**	**Porcentaje acumulado**
Válido	2	5	21,7	21,7	21,7
	3	4	17,4	17,4	39,1
	4	4	17,4	17,4	56,5
	5	3	13,0	13,0	69,6
	6	2	8,7	8,7	78,3
	7	2	8,7	8,7	87,0
	8	3	13,0	13,0	100,0
	Total	23	100,0	100,0	

2. Estadísticos descriptivos.

Estadísticos		
Citas		
N	Válido	23
	Perdidos	0
Media		4,48
Error estándar de la media		,435
Mediana		4,00
Moda		2
Desv. Desviación		2,086
Varianza		4,352
Curtosis		-1,056
Error estándar de curtosis		,935
Mínimo		2
Máximo		8
Suma		103

3. Gráfico de la distribución.



2.3 Datos bivariados

En estadística los datos bivariados, son datos de cada una de las dos variables de estudio, donde cada valor de las variables se combina entre sí ejemplo: edad y género, peso y altura, es usual que los investigadores se interesen en más de una sola variable que puedan medir en su investigación. Cuando dos variables se miden en una sola unidad experimental, los datos resultantes se denominan datos bivariados.

Cuando los datos bivariados provienen de dos variables cuantitativas, resulta importante analizar la relación entre ellos, la relación puede ser lineal, cuadrática, exponencial, logarítmica, trigonométrica, entre otras, desde una perspectiva estadística la relación que interesa es la lineal, por lo cual se lleva a cabo una regresión lineal y análisis de correlación.

Los datos bivariados se expresan a través de pares de valores; tipo pares ordenados (a,b); cuando se obtiene estos valores se refiere a una distribución bidimensional de la información analizada, cada una de las dos variables puede ser cuantitativa o cualitativa.

2.3.1 Dos variables cualitativas

Cuando existen datos bivariados de dos variables cualitativas, es frecuente que se ordenen en una tabulación cruzada o tabla de contingencia, la cual e un medio particular para representar simultáneamente dos características observadas en una misma población.

Una muestra de 30 docentes que pertenecen a IES privadas y públicas, tienen 3 tipos de dedicación: Tiempo Completo, tiempo parcial y medio tiempo como se muestra en la siguiente tabla.

Tabla 2-1. Datos Bivariados Variables Cualitativas.

	Tipo de Entidad	Tiempo de dedicación
1	Público	Prof_TC
2	Privado	Prof_MT
3	Público	Prof_TP
4	Público	Prof_TP
5	Privado	Prof_TP

6	Privado	Prof_TP
7	Privado	Prof_TP
8	Privado	Prof_TP
9	Público	Prof_TP
10	Privado	Prof_TC
11	Público	Prof_MT
12	Público	Prof_TC
13	Privado	Prof_MT
14	Privado	Prof_TC
15	Privado	Prof_TC
16	Público	Prof_TC
17	Público	Prof_TC
18	Privado	Prof_MT
19	Privado	Prof_TC
20	Privado	Prof_MT

Estos 30 datos se pueden resumir en una tabla cruzada de 2*3, donde las dos filas representan el tipo de entidad, y las tres columnas representan el tiempo de dedicación de los docentes, en la tabla de contingencia se muestra la frecuencia para cada categoría cruzada.

Tabla 2-2. Tabla Cruzada Variables Cualitativas.

Tipo Entidad*Tipo Docente						
Recuento						
		Tipo Docente				
			Prof_MT	Prof_TC	Prof_TP	Total
Tipo Entidad		3	0	0	0	3
	Privado	0	4	4	4	12
	Público	0	1	4	3	8
Total		**3**	**5**	**8**	**7**	**23**

También se puede generar una tabla con los porcentajes de las categorías cruzadas.

Tabla 2-3. Tabla Cruzada Variables Cualitativas porcentajes.

Tabla cruzada Tipo_Entidad*Tipo_Docente											
		Tipo_Docente									
				Prof_MT		Prof_TC		Prof_TP		Total	
		N	%	N	%	N	%	N	%	N	%
Tipo_Entidad		3	100,0%	0	0,0%	0	0,0%	0	0,0%	3	13,0%
	Privado	0	0,0%	4	80,0%	4	50,0%	4	57,1%	12	52,2%
	Público	0	0,0%	1	20,0%	4	50,0%	3	42,9%	8	34,8%
Total		**3**	**100,0%**	**5**	**100,0%**	**8**	**100,0%**	**7**	**100,0%**	**23**	**100,0%**

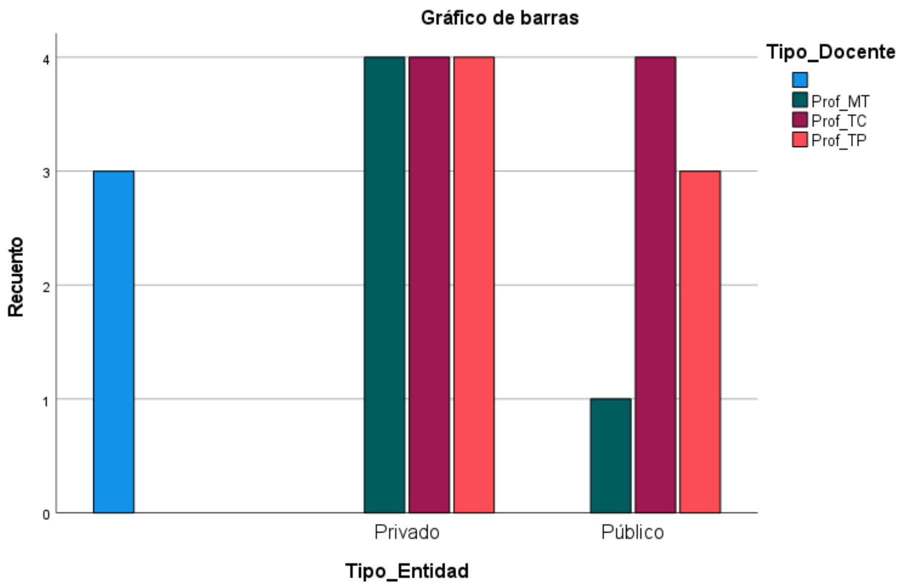


Figura 2-1. Gráfico de Barras datos bivariados cualitativos.

2.3.2 *Dos variables cuantitativas*

Cuando existen datos bivariados de dos variables cuantitativas se acostumbra a expresar matemáticamente los datos como pares ordenados (x, y), donde x es la variable de entrada (variable independiente) y **Y** es la variable de salida (variable dependiente), los datos muestrales se representan en un diagrama de dispersión.

Un diagrama de dispersión es una gráfica de todos los pares ordenados de datos bivariados en un sistema de ejes de coordinadas, la variable X se localiza en el eje horizontal y la variable de salida en el eje vertical.

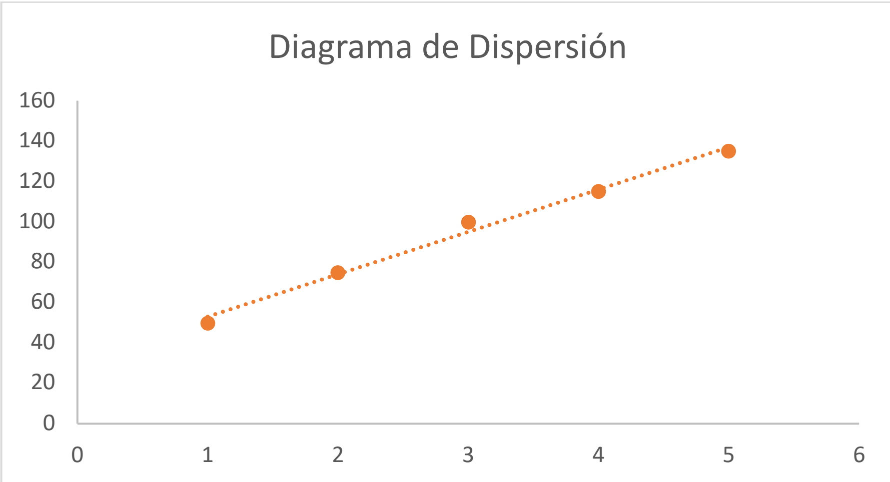


Figura 2-1. Ejemplo de Diagrama de Dispersión.

Ejemplo 2.15

Se mide el número de miembros de una familia (X), así como la cantidad que gastan en una semana en comida.

X	2	2	3	4	1	5
Y	95.75	119.10	113.83	159.20	86.85	160.82

Solución: Marcar los ejes X y Y, graficar los puntos utilizando las coordinadas (x,y) por cada uno de los seis pares ordenados, la gráfica de dispersión es como se muestra en la figura.

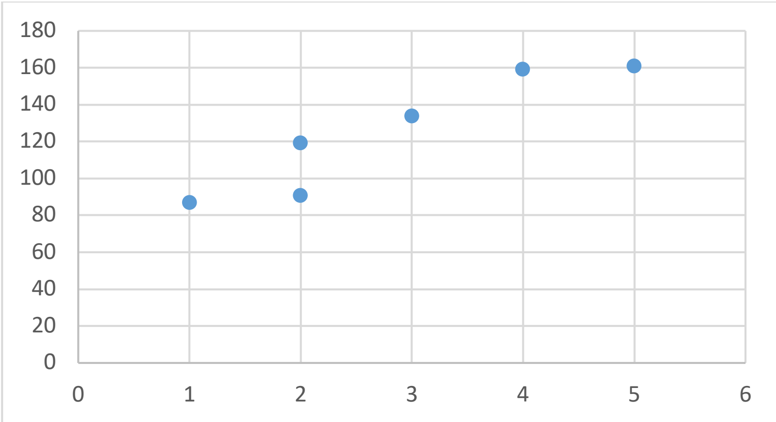


Figura 0-2. Diagrama de dispersión ejemplo 2.15

2.4 Regresión lineal

La regresión lineal es una técnica de modelado estadístico que se emplea para describir una variable de respuesta continua como una función de una o varias variables predictores. Puede ayudar a comprender y predecir el comportamiento de sistemas complejos o a analizar datos experimentales, financieros y biológicos.

Las técnicas de regresión lineal permiten crear un modelo lineal. Este modelo describe la relación entre una variable dependiente **y** (también conocida como la respuesta) como una función de una o varias variables independientes Xi (denominadas predictores) (Puente et al., 2017). La ecuación general correspondiente a un modelo de regresión lineal es:

$$Y = \beta_0 + \sum \beta_1 X_1 + \varepsilon_1 \qquad (2.21)$$

donde β0 representa las estimaciones de parámetros lineales que se deben calcular y ϵϵ representa los términos de error.

Existen varios tipos de regresión lineal tales como regresión lineal simple, regresión lineal múltiple, regresión lineal multivalente.

Ejercicios Propuestos

Ejercicio 2.1
Se presentan las siguientes series simples y se pide calcular la mediana.

a. 2, 3, 3,4
b. 1, 18, 19, 20
c. 5.2, 6.5, 8.1, 9.1, 10.1,15.5
d. 1, 2, 3, 3, 3,4, 7
e. 9, 40, 80, 81, 100

Respuesta: a. 3, b 18.5, c. 8.6, d. 3, e. 80

Ejercicio 2.2
Se presentan las puntuaciones del examen final de matemática de 40 estudiantes de nivelación de la ESPOCH en la siguiente tabla.

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

- A partir de la serie simple, construir la serie o tabla de frecuencias.
- Construir la tabla de clases y frecuencias siguiendo los pasos de determinar amplitud de variación de, número de clases, tamaño de clase y tabulación de datos.
- Mencionar cuales son las fronteras de la segunda clase.

Ejercicio 2.3
En la siguiente tabla se presentan datos de la estatura de 36 estudiantes de la carrera de Ingeniería Electrónica.

150	160	161
151	160	161
153	159	162
156	167	160
159	167	160
160	169	160
160	145	165
165	150	170
165	151	165
170	152	166
171	150	166
172	153	163

- A partir de la serie simple, construir la serie o tabla de frecuencias.
- Construir la tabla de clases y frecuencias siguiendo los pasos de determinar amplitud de variación de, número de clases, tamaño de clase y tabulación de datos.
- Mencionar cuales son las fronteras de la segunda clase.

Ejercicio 2.4

1. Hallar la mediana, media y moda de los siguientes valores: 1.2,1.5, 1.6, 2.1, 2.4, 2.7, 2.8, 3.0, 3.0, 3.0, 3.1, 3.1, 3.1, 3.4

2. Hallar la media, mediana y moda de los siguientes datos: 3, 5, 2, 6, 5, 9, 5, 2, 8

Ejercicio 2.5

Encontrar la moda, mediana y media de las puntuaciones que se presentan en la siguiente distribución de frecuencia agrupadas.

Intervalo	**Frecuencias**
40-48,1	3
48,1-56,1	8
56,1-64,1	11
64,1-72,1	32
72,1-80,1	21
80,1-88,1	18
88,1-96,1	14
96,1-104,0	1

Ejercicio 2.6

Los valores del ph sanguíneo de 32 individuos son:

7.33	7.33	7.37	7.32
7.31	7.32	7.35	7.33
7.26	7.35	7.34	7.32
7.33	7.39	7.32	7.4
7.37	7.33	7.29	7.33
7.27	7.38	7.35	7.32
7.30	7.33	7.38	7.34
7.33	7.31	7.32	7.33

Calcule las medidas de tendencia central, disperción, realice la tabla de frecuencia.

Ejercicio 2.7

Con los siguientes datos calcule las medidas de tendencia central, dispersión, deciles, percentiles.

Intervalos	**Frecuencia fi**
1500-2321	9
2321-3142	1
3142-3963	10
3963-4787	0
4787-5606	4
Total	**24**

Ejercicio 2.8

Cinco niños de 2, 3, 5, 7 y 8 años de edad pesan, respectivamente, 14, 20, 32, 42 y 44 kilos. Hallar la ecuación de la recta de regresión de la edad sobre el peso. ¿Cuál sería el peso aproximado de un niño de seis años?

CAPÍTULO 3

Teoría de probabilidades

3. Teoría de probabilidades

La teoría de probabilidades tiene la finalidad de hacer inferencias o deducciones relacionadas con una población tomando como base los distintos muestreos estadísticos; la utilización de probabilidades se remonta al siglo XVI, donde las primeros usos se relacionaban con los juegos de azar, hoy en día la aplicación y utilidad de las probabilidades es amplia, gobiernos, empresas y entidades profesionales usan la teoría de probabilidades en el proceso diario de la toma de decisiones, porque las probabilidades pueden ser una herramienta útil para el desarrollo de estrategias; el punto central en todos los casos que se utiliza las probabilidades es cuantificar cuan probable es la ocurrencia de un determinado evento.

La probabilidad y la estadística se vinculan entre sí porque la probabilidad es utilizada como una herramienta que evalúa la confiabilidad de las conclusiones sobre una población cuando solamente se tiene una información muestral (Mendenhall, 2017).

3.1 Definiciones básicas de probabilidad

3.1.1 Observaciones y experimentos

La información utilizada en estadística se obtiene al observar sucesos no controlables en un entorno, o bajo condiciones controlables en un laboratorio, en este contexto se define como experimento al proceso por el cual una o varias observaciones son registradas.

Ejemplos de experimentos donde se generan observaciones cotidianas pueden ser:

- Registro mensual de un trabajador durante un período contable.
- Entrevista a un consumidor para definir una marca preferida de producto.
- Registro del valor de las acciones de una empresa en la bolsa de valores, por día, mes año.

Con lo antes expuesto se concluye que un experimento es cualquier suceso que se desea investigar y que al ser analizado se obtiene resultados, los cuales se denominan observaciones.

3.1.2 ***Espacio de resultados o muestral***

Es fundamental definir cuidadosamente los elementos que integran el conjunto[5] de interés para estar en condiciones de decidir si un objeto pertenece o no a un conjunto. Existen dos maneras de describir los elementos de un conjunto:

- Enumeración: Conjunto A= {Oscar, Danilo, Andrés}
- Extensión: Conjunto B={x/x sea <10 }

En este sentido el espacio muestral se define como el conjunto de resultados u observaciones de un experimento o muestra; al conjunto de todos los puntos de la muestra de un experimento se denomina espacio de resultados donde se representa con la letra S.

3.1.3 ***Evento***

Se denomina evento a todo resultado o suceso que genera un experimento, estos pueden ser simples o compuestos.

- Evento simple es el resultado que se observa de una sola repetición del experimento.
- Evento compuesto es aquel que contiene dos o más eventos simples.

3.1.4 ***Definición de probabilidad***

No existe una definición universal de probabilidad, sin embargo; es considerada como un mecanismo que permite la utilización de información parcial de una muestra, para inferir sobre la naturaleza de una población (Llinas, 2017; Puente, 2018, Díaz, 2019); es utilizada para expresar cuan probable es determinar un evento.

---

[5] Conjunto es un grupo de objetos o elementos (observaciones) que tienen características comunes.

La probabilidad se asocia a un suceso o evento aleatorio, es una medida del grado de certidumbre que dicho suceso pueda pasar, suele expresarse como un número entre 0 y 1, en estadística la probabilidad propone modelos para los fenómenos aleatorios, que pueden predecirse con certeza y estudiar las consecuencias lógicas.

3.1.5 ***Enfoque de probabilidad***

Existen tres maneras de estimar la probabilidad como se muestra en la figura 3-1, el seleccionar el enfoque que se usará para calcular la probabilidad de algún evento dependen del contexto, se asigna la letra P para indicar la probabilidad de ocurrencia de un evento, ejemplo P(A) donde se lee probabilidad de ocurrencia de un evento A en un experimento.

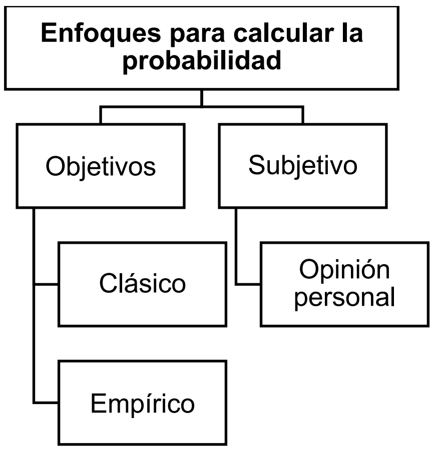


Figura 3-1. Enfoque para calcular la probabilidad.

- **Enfoque Clásico:** se utiliza en situaciones donde los eventos tienen resultados igualmente probables, uno ejemplo tradicional es el tiro de monedas, lanzamiento de dados; en este tipo de eventos la probabilidad de eventos se calcula de la siguiente manera.

$$P\ (cada\ resultado) = \frac{1}{número\ de\ resultados\ posibles} \tag{3.1}$$

Ejemplo 3.1

De un paquete de cartas de 52, la probabilidad de sacar cualquier carta es:

$$P\ (cada\ carta\ ) = \frac{1}{52} = 0.01923$$

Del tiro de una moneda, presenta dos resultados, cara o sello, en este escenario la probabilidad de que caiga cara o sello es:

$$P\ (cara) = \tfrac{1}{2} \qquad P\ (sello\ ) = \tfrac{1}{2}$$

En una rifa donde se han vendido 100 boletos, la probabilidad de tener el boleto ganador es de:

$$P\ (cada\ boleto) = \frac{1}{100}$$

El enfoque clásico se puede aplicar a eventos que comparen dos o más resultados, en este caso la fórmula 3.1 se modifica a la siguiente:

$$P\ (A) = \frac{número\ de\ resultados\ favorables}{número\ total\ de\ resultados} \qquad (3.2)$$

Ejemplo 3.2

¿Cuál es la probabilidad de sacar una de los cuatro Ases de un paquete de 52 cartas?

$$P\ (A) = \frac{4\ Ases}{52\ cartas}$$

La Aseguradora LIFE S.A., ha estudiado las causas de atención médica en el último año, donde se recogieron 230 casos, de los cuales 150 enfermedades gastrointestinales, 50 fracturas, 30 accidentes de tránsito.

Si se escoge aleatoriamente uno de los eventos de salud, calcule la probabilidad que la atención médica haya sido por accidente de tránsito.

$$P\ (accidentes\ de\ tránsito\ ) = \frac{30}{230} = 0.13$$

Interpretación: existe la probabilidad de 0.13 o del 13% de seleccionar una persona que haya recibido atención médica por accidente de tránsito.

- **Enfoque Empírico** (datos históricos o frecuencias relativas): al estimar probabilidades con este enfoque se obtiene una aproximación y no un valor exacto, a medida que se incrementan el total de observaciones, las aproximaciones correspondientes tienden a acercarse a la realidad.

Ejemplo al realizar una encuesta de opinión a una decena de personas, es posible que los resultados no sean representativos para toda la población, pero si la misma encuesta es aplicada a miles de personas seleccionadas al azar, los resultados de la muestra están más cercados a los valores verdaderos de la población.

$$P\ (A) = \frac{número\ de\ veces\ que\ ha\ ocurrido\ A\ (datos\ empíricos)}{número\ total\ de\ elementos\ del\ experimento} \qquad (3.3)$$

- **Enfoque subjetivo:** la probabilidad generada en este enfoque se realiza con una estimación subjetiva de la probabilidad, por ejemplo, como calcular la probabilidad de: ¿Qué calificación tendrá en la prueba de inglés el día lunes?, ¿Las flores plantadas ayer crecerán frondosas? En este tipo de situaciones alguien debe establecer la probabilidad de que ocurra el evento en dichas situaciones. Una probabilidad estimada se fundamenta en el conocimiento de las circunstancias pertinentes y se denomina probabilidad subjetiva.

  Este enfoque presenta algunas desventajas entre las que se mencionan:

  - Las estimaciones subjetivas suelen ser complicadas de comprobar si son cuestionadas.

  - Los prejuicios pueden influir, las ideas preconcebidas respecto a lo que debería suceder pueden afectar la subjetividad, así como los sentimientos de lo que uno quiera que suceda.

3.2 *Principio fundamental de conteo: permutaciones y combinaciones*

Para utilizar el enfoque clásico de probabilidades es básico saber el número total de posibles resultados de un experimento o muestra, en este escenario las técnicas de conteo son el medio para determinar el número total de resultados posibles de un experimento.

3.2.1 *Diagrama de árbol*

El diagrama de árbol se utiliza en problemas de conteo y probabilidad, para su construcción se parte colocando una rama por cada probabilidad, cada una de estas ramas se conocen como rama de primera generación.

Por ejemplo, si un estudiante de la ESPOCH se presenta a un examen de 3 preguntas de verdadero y falso, suponiendo que el estudiante está adivinando las respuestas a las preguntas. ¿Cuál es la probabilidad que resuelva bien el examen? Para esto es necesario determinar el número total de resultados posibles, puede contestar todas verdaderas, todas faltas o alternar o mezclar aleatoriamente las respuestas.

Si el examen fuera de una sola pregunta las posibilidades serían V o F, si fueran dos preguntas las posibilidades serían VV, VF, FV, FF, pero a medida que se incrementan las preguntas las posibilidades también lo hacer.

Al ampliar el diagrama de árbol es posible enumerar los resultados con más preguntas de verdadero y falso, pero no es práctico, debido al número de posibilidades que aumenta considerablemente, a continuación, se presenta el diagrama de árbol con 3 preguntas.

Preguntas			
Núm. 1	**Núm. 2**	**Núm. 3**	**Resultado**
		V	VVV
	V		
		F	VVF
V		V	VFV
	F		
		F	VFF
		V	FVV
	V		
		F	FVF
F		V	FFV
	F		
		F	FFF
2	**2**	**2**	**8**

3.2.2 ***Regla de conteo mn***

Considera un experimento que se realiza en dos etapas, si la primera etapa se efectúa en m formas y para cada una de estas, la segunda etapa se logra en n formas, entonces hay mn formas de efectuar el experimento.

Ejemplo 3.3

Una taza contiene un chocolate con envoltura amarilla y dos chocolates con envoltura roja, se seleccionan dos chocolates de la taza, registrando sus colores.

¿Cuántos eventos simples hay en el espacio muestral S?

Solución: El primer chocolate se elige en m = 3 formas, como uno de los chocolates ya ha sido seleccionado, el segundo chocolate se elige en n = 2 formas, el número total de eventos simples es:

mn = 3*2 = 6

3.2.3 ***Regla de mn extendida***

Si un experimento se realiza en k etapas, con n_i formas para ejecutar la primera etapa, n_2 formas para efectuar la segunda etapa y n_k formas para efectuar la k-esima etapa, entonces el número de formas para efectuar el experimento es: $n_1 n_2 \dots n_k$, donde n los posibles resultados y K el número de lanzamientos o veces que se repite el experimento

$$P = n^k \tag{3.5}$$

Ejemplo 3.4

¿Cuántos eventos simples hay en un espacio muestral de 3 monedas?

Solución: cada moneda puede caer de dos formas n = 2 K=3

$$P = n^k$$

$$P = 2^3 = 8$$

3.2.4 ***Regla de conteo por permutaciones***

La permutación es una técnica de conteo que permite calcular las posibles ordenaciones de elementos de un conjunto o número de elementos de un espacio muestra de un experimento aleatorio utilizando la siguiente fórmula:

$$P_x^n = \frac{n!}{(n-x)!} \tag{3.6}$$

Dónde

n! = n(n-1)(n-2)….(n-x+1)

Ejemplo 3.5

Si existen 6 libros, pero solo 4 espacios para colocar en una repisa. ¿De cuantas maneras se puede colocar los libros en la repisa?

n = 6 X= 4

$$P_x^n = \frac{6!}{(6-4)!} = \frac{6!}{2!} = \frac{6*5*4*3*2*1}{2*1} = 360$$

Ejemplo 3.6

En un grupo de 10 personas, ¿Cuántas ordenaciones de dos de las diez personas son posibles?

El número de formas en las que se pueden ordenar dos personas de un grupo de diez es:

$$P_2^{10} = \frac{10!}{(10-2)!} = \frac{10!}{8!} = \frac{10*9*8!}{8!} = 90$$

3.2.5 ***Regla de conteo para combinaciones***

Las combinaciones es una técnica de conteo que permite calcular el número de arreglos que se pueden realizarse con todos o con una parte de los elementos de un conjunto, en donde no interesa el orden de los elementos.

$$C_x^n \frac{n!}{x!(n-x)!} \tag{3.7}$$

Ejemplo 3.7

Se compra una impresora de planos, elegida entre cinco proveedores ¿En cuántas formas se puede elegir tres de los cinco proveedores?

n = 5 X= 3

$$C_3^5 \frac{5!}{3!\,(5-3)!} = \frac{5*4*3!}{3!*2!} = 10$$

Ejemplo 3.8

¿Cuántas combinaciones de cuatro libros seleccionados entre seis libros se puede realizar?

n = 6 X= 4

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

Para aplicar la regla de conteo para combinaciones se deben dar las siguientes condiciones:

a. Debe usarse un total n cosas distintas disponibles.
b. Debe seleccionarse x de las n cosas (sin repetición).
c. Debe considerarse que reacomodos de las mismas cosas son considerados como el mismo agrupamiento, es decir; iguales.

Ejercicio 3.9

Combinaciones y permutaciones

Si existen 7 libros, pero solo 4 espacios para colocar en una repisa. ¿De cuantas maneras se puede colocar los libros en la repisa y cuantas combinaciones se puede realizar?

n = 7 X= 4

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

$$P_4^7 = \frac{7!}{(7-4)!} = \frac{7!}{3!} = \frac{7*6*5*4*3*2*1}{3*2*1} = 840$$

$$C_x^n \frac{n!}{x!\,(n-x)!}$$

$$C_4^7 \frac{7!}{4!\,(7-4)!} = \frac{7*6*5*4!}{4!*3*2*1} = 140$$

En este caso se pueden colocar de 840 maneras diferentes los libros en la repisa y las combinaciones de los 7 libros agrupados de 4 en cuatro pueden ser 140 combinaciones diferentes.

3.3 Axiomas de probabilidad

Los axiomas de probabilidad constituyen una base para deducir un amplio número de resultados, son condiciones mínimas que deben verificarse para que una función definida sobre un conjunto de sucesos determine consistentemente sus probabilidades (Hernández et al., 2019, Cárdenas, 2014).

Los axiomas fueron establecidos por el matemático ruso Andrés Kolmogorov en 1933.

3.3.1 Axioma de no negatividad

La probabilidad de ocurrencia de cualquier evento A siempre es positiva o cero, P(A)≥ 0, cuando la probabilidad de un evento es 0, se llama suceso imposible.

3.3.2 Axioma de certidumbre

Siempre que algún evento que pertenece a E, su probabilidad de ocurrencia es 1, lo cual se puede expresar como P(E)=1, es lo que se conoce como evento seguro ya que, al realizar un experimento, con toda certeza hay un resultado.

3.3.3 Axioma de adición

En el caso de dos o más eventos incompatibles, denominados A_1, A_2, A_3..., la probabilidad de ocurrencia del suceso A_1 más la del suceso A_2 más A_3 y así sucesivamente, es la suma de las probabilidades de ocurrencia de cada uno por separado, lo que se expresa de la siguiente manera:

$$P(A_1) + P(A_1) + \cdots . + P(A_n) \qquad (3.8)$$

3.3.4 Formas de asignar probabilidades

Los axiomas de probabilidad no forman un método de asignación de valor de probabilidad, sin embargo, para ello hay tres opciones que son compatibles con los axiomas.

- **Regla Laplace:** a cada evento se le asigna la misma probabilidad de ocurrencia.

$$P(A) = \frac{número\ de\ casos\ favorables\ del\ evento}{número\ de\ casos\ posibles} \qquad (3.9)$$

- El experimento tiene que ser repartible, porque el método se fundamenta en realizar un gran número de repeticiones, entonces la probabilidad del evento se representa con la siguiente fórmula:

$$P(A) = \lim_{n\to\infty} \left(\frac{n}{i}\right) \qquad (3.10)$$

Dónde n/i = frecuencia relativa de un suceso.

Definir p(A) de esta manera satisface los axiomas de Kolmogorov, pero tiene un inconveniente, pues hay que realizar muchas pruebas para que la probabilidad sea apropiada.

3.4 Eventos mutuamente excluyentes y no excluyentes

Un evento o suceso es un subconjunto de un espacio muestral, representa una serie de resultados posibles, en este contexto se pueden clasificar los eventos en mutuamente excluyentes y no excluyentes.

3.4.1 Eventos mutuamente excluyentes

Se consideran eventos mutuamente excluyentes si no pueden ocurrir simultáneamente en un mismo experimento, cuando los eventos son mutuamente excluyentes la probabilidad de que uno u otro ocurra equivale a la suma de sus probabilidades, la fórmula que se aplica al cálculo de estas probabilidades es:

$$P(AoB) = P(A) + P(B) \qquad (3.11)$$

Ejemplo 3.10

La probabilidad de tener sello o cara en la misma tirada de una moneda es:

$$P(SoC) = P(S) + P(C) = \frac{1}{2} + \frac{1}{2} = 1$$

Ejemplo 3.11

La probabilidad de tirar un dado y se obtenga un resultado de 5 o 6 es:

$$P(5o6) = P(5) + P(6) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$$

Ejemplo 3.12

La probabilidad de sacar al primer intento una carta de rubíes o una de tréboles de un paquete de cartas de 52 es:

$$P(rubies\ o\ tréboles) = P(rubies) + P(tréboles) = \frac{13}{52} + \frac{13}{52} = \frac{26}{52} = \frac{1}{2}$$

3.4.2 ***Eventos no excluyentes***

Cuando dos eventos no son mutuamente excluyentes, ambos pueden suceder al mismo tiempo, en este tipo de situaciones se debe considerar el hecho de que cada uno de ellos o ambos pueden ocurrir, su fórmula de cálculo es:

$$P(AoB\ o\ ambos\ ) = P(A) + P(B) - P\ (AyB) \qquad (3.12)$$

Ejemplo 3.13

Se quiere determinar la probabilidad de sacar una carta de trébol o un 10 de un conjunto de 52 cartas, en este caso se pueden dar dos eventos, es decir; ser un trébol y un número 10, en esta situación los eventos no son mutuamente excluyentes.

♣	♦	♥	♠
K	K	K	K
Q	Q	Q	Q
J	J	J	J
10	10	10	10
9	9	9	9
8	8	8	8
7	7	7	7
6	6	6	6
5	5	5	5
4	4	4	4
3	3	3	3
2	2	2	2
A	A	A	A

Figura 3-1. Detalle ejemplo 3.13

En la figura 3-2 se muestra el detalle de las probabilidades de cada elemento del conjunto de datos, en este caso simplemente se suman las probabilidades individuales, ya que el 10 de trébol se contará dos veces, una como 10 y otra como trébol como se evidencia en la figura, en consecuencia, se debe restar la probabilidad de sobreestimar o duplicar.

Evento A = tréboles = {A, 2, 3, 4, 5, 6, 7, 8, 9, 10, J, O, K}
Evento B = cartas número 10 = {10 trébol, 10 diamante, 10 corazones rojos, 10 corazones negros}

Por lo tanto:

P(tréboles) = 13/52
P (10) = 4/52
P (10 de tréboles) = 1/52

Entonces la probabilidad buscada

$$P(trébol\ o\ diez\ o\ ambos\ ) = P(trébol) + P(diez) - P\ (diez\ de\ trebol)$$

$$P(trébol\ o\ diez\ o\ ambos\ ) = \frac{13}{52} + \frac{4}{52} - \frac{1}{52} = \frac{16}{52} = 0.30769 = 30.769\% \approx 31\%$$

Ejemplo 3.14

Un directivo de una IES quiere aplicar una encuesta sobre el desempeño del docente en lo que va del período académico, selecciona a una parte de los estudiantes y se decide seleccionar a los participantes aleatoriamente de acuerdo al último dígito de cédula ¿Cuál es la probabilidad de que el número seleccionado será 0 o 1?

$$P(0\ o\ 1\ ) = \frac{1}{10} + \frac{1}{10} = \frac{1}{5}$$

Si se modifica que se considere la probabilidad de obtener un resultado que sea impar o mayor que 6, los posibles resultados cinco impartes (1,3,5,7,9) y tres son mayores a 6 (7,8,9)

Evento A = impares = {1, 3, 5, 7, 9}
Evento B = mayores a 6 = {7, 8, 9}
Evento C = impar y mayor a 6 = {7, 9}

$$P(Impar\ o\ mayor\ a\ 6) = P(>6) + P(impar) - P\ (ambos)$$

$$P(Impar\ o\ mayor\ a\ 6) = \frac{3}{10} + \frac{5}{10} - \frac{2}{10} = \frac{6}{10} = 0.6$$

3.5 Eventos dependientes e independientes

Dos eventos son independientes entre sí cuando la ocurrencia no está relacionada con la ocurrencia del otro, por otra parte, si los eventos no son independientes, significa que son dependientes entonces la ocurrencia del otro.

3.5.1 Evento independiente

Cuando dos eventos son independientes (con reemplazo), la probabilidad de que ambos ocurren es igual al producto de sus probabilidades individuales como se muestra en la siguiente fórmula:

$$P(A\ y\ B) = P(A) * P(B) \qquad (3.13)$$

Ejemplo 3.15

El 10% de personas son zurdas ¿Qué probabilidad hay de seleccionar aleatoriamente a dos personas zurdas?

La probabilidad de que un individuo sea zurdo es de 0.10, pero nada influye para que la otra persona sea zurda, son eventos independientes, por lo tanto, la probabilidad es:

$$P(zurda\ y\ zurda) = P(zurda) * P(zurda) = 0.10 * 0.10 = 0.01$$

3.5.2 ***Eventos dependientes***

Cuando dos eventos son dependientes (sin reemplazo), quiere decir que la probabilidad que ocurra uno u otro se ven afectados por la probabilidad del suceso que ocurra primero, su fórmula de cálculo es la siguiente:

$$P(A\ y\ B) = P(A) * P\left(\frac{B}{A}\right) \tag{3.14}$$

Dónde:

P(B/A) = probabilidad que el suceso B ocurra después de suponer que ocurrió el suceso A.

Ejemplo 3.16

Una empresa fabrica motos modelo QRZ, y debe cubrir al mercado los filtros de gasolina para las motos, en la fábrica de repuestos se ha verificado un lote de 40 filtros de gasolina, seis de ellos con defectos; los filtros salen en cajas de 40 piezas, una casa comercial de repuestos que compra estos filtros recibe una caja y vende dos de ellos. ¿Cuál es la probabilidad de que los dos primeros salgan en buen estado si los filtros se seleccionan sin reemplazo?

Si los filtros se seleccionan sin reemplazo, las dos selecciones son dependientes, porque el primer resultado afecta al segundo, por lo tanto, la probabilidad se calcula de la siguiente manera.

$$P(1Bueno\ y\ 2\ Bueno) = P(1Bueno) * P\left(\frac{2Bueno}{1Bueno}\right)$$

$$P(1Bueno\ y\ 2\ Bueno) = P\left(\frac{34}{50}\right) * P\left(\frac{33}{39}\right)$$

$$P(1Bueno\ y\ 2\ Bueno) = P\left(\frac{34}{50}\right) * P\left(\frac{33}{39}\right)$$

$$P(1Bueno\ y\ 2\ Bueno) = 0.719$$

En este caso se ajusta la segunda probabilidad para tomar en consideración la elección de un filtro bueno en la primera selección, después de escoger el filtro bueno la primera vez, quedaban 33 filtros buenos entre los 39 restantes.

Ejemplo 3.17

En un cartón se encuentran tres bolas rojas y dos bolas verdes, Sea A el suceso "la primera bola extraída es verde" y B el suceso "la segunda bola extraída es verde". Considere que las bolas no son devueltas al cartón; entonces, A y B son sucesos dependientes.

La probabilidad que la primera bola sea verde se representa:

$$P(A) = \frac{2\ bolas\ verdes}{5\ (bolas\ verdes\ y\ rojas)}$$

La probabilidad de que la segunda sea una bola verde, dado que lo ha sido la primera, es:

$$P(B/A) = \frac{1 bola\ verde}{(1\ bola\ verde + 3\ bolas\ rojas)}$$

Luego la probabilidad que ambas sean verdes se representa de la siguiente manera:

$$P(A\ y\ B) = P(A) * P\left(\frac{B}{A}\right)$$

$$P(A\ y\ B) = \frac{2}{5} * \frac{1}{4}$$

$$P(A\ y\ B) = 0.10 \approx 10\%$$

3.6 Teoremas de las probabilidades

3.6.1 Regla de la adición

La regla de la suma se usa en el cálculo de probabilidades de eventos excluyentes y no excluyentes; esta regla define que para obtener P (AoB), se calcula la suma del número de formas que puede ocurrir un evento A y el número de formas que puede ocurrir un evento B; sumado de tal manera que cada resultado se registre una vez.

Ejemplo 3.18

La probabilidad de que se lance un dado y el resultado sea 3 o 4 se representa de la siguiente manera:

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

$$P(3\ o\ 4) = 1/6 + 1/6$$

$$P(3\ o\ 4) = \frac{2}{6} = \frac{1}{3}$$

3.6.2 ***Regla de la multiplicación***

La regla de la multiplicación es utilizada en el cálculo de probabilidades en eventos dependientes e independientes. Esta regla expresa que para determinar P(AyB), se calcula la probabilidad de que el suceso A ocurra primero y el suceso B ocurra segundo.

3.6.3 ***Probabilidad condicional***

La probabilidad condicionada es un tema clave en la teoría de probabilidades, este tipo de probabilidad tiene una clara interpretación en espacios muéstrales finitos, en los cuales se puede aplicar la regla Laplace.

La probabilidad condicionada de B dado A, es la probabilidad de que el suceso B ocurra dado que el suceso A ya ha ocurrido, la siguiente fórmula indica como calcular la probabilidad condicional.

$$P(B/A) = \frac{P\ (A\ y\ B)}{P\ (A)} \qquad (3.15)$$

Ejemplo 3.19

Se requiere determinar la probabilidad de que un estudiante seleccionado al azar sea de Química y tenga un promedio de 8 o mayor, se presenta la siguiente información:

	Química	Otras asignaturas	Total
Promedio menor a 8	20	37	57
Promedio mayor o igual a 8	15	22	37
Total	**35**	**59**	**94**

En la información proporcionada la condición es que el estudiante tenga un **promedio de 8 o mayor**, el espacio muestral se reduce a los estudiantes que tienen dicho promedio, en este caso es 37 estudiantes, en este contexto la probabilidad de que el estudiante sea de Química, dado que tiene un promedio de 8 o mayor, se calcula de la siguiente manera:

$$P(Química\ 8\ o\ mayor) = \frac{\#\ de\ estudiantes\ química\ con\ promedio\ de\ 8\ o\ mayor}{\#\ de\ estudaintes\ con\ promedio\ de\ 8\ o\ mayor}$$

$$P(Química\ 8\ o\ mayor) = \frac{15}{37} = 0.41\ \approx 41\%$$

3.6.4 ***Teorema de Bayes***

La probabilidad condicionada considera la información sobre la ocurrencia de un evento para hallar la probabilidad de otro evento; el concepto puede extenderse para revisar probabilidades cuya base sea información nueva, y que para determinar la probabilidad de un efecto particular sea el resultado de una causa específica, este procedimiento se denomina Teorema de Bayes.

El teorema de Bayes representa una técnica que se usa para la verificación de estimaciones iniciales de la probabilidad basada en datos de la muestra, es decir; es un método de verificación de las probabilidades anteriores con base a la información obtenida por muestreo (Mendenhall, 2017; Puente, 2018). Es una propuesta planteada por el matemático Thomas Bayes, publicada en 1763, que expresa la probabilidad condicional de un evento aleatorio en términos de distribución de la probabilidad.

$$P(A/B) = \frac{P(A)P(B/A)}{P\ (B)} \qquad (3.16)$$

Dónde:
A y B son eventos y además P(B) ≠ 0.
P(A/B) = probabilidad de que ocurra A, dado que ha ocurrido B.
P(B/A) = probabilidad de que ocurra B, dado que ha ocurrido A.
P(A) = probabilidad de que ocurra A.
P(B) = probabilidad de que ocurra B.

Ejemplo 3.20

En la ESPOCH la probabilidad de que un alumno seleccionado al azar le guste el chocolate es del 60%, mientras que la probabilidad que a un alumno le guste el café es del 36%. Además, se conoce que la probabilidad de que un alumno le guste tanto el chocolate como el café es del 40%. Calcular la probabilidad de que a un alumno le guste el chocolate, dado que le gusta café.

Evento A = le gusta el chocolate.
Evento B = le gusta el café.
P (A) = 0.6
P (B) = 0.36
P (B/A) = 0.40
P (A/B) =?

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

$$P(A/B) = \frac{0.6 * 0.40}{0.36}$$

$$P\left(\frac{A}{B}\right) = 0.6667 \approx 66.67\%$$

En este contexto la probabilidad que al alumno le guste el chocolate dado que le gusta el café es del 66.67%.

3.7 Distribución de probabilidades

Una distribución de probabilidades muestra todos los resultados probables de un experimento, así como la probabilidad de ocurrencia de dichos resultados; la forma en la que se distribuyen las diferentes probabilidades de cada valor asumido por la variable aleatoria.

3.7.1 Distribución de frecuencias y frecuencias acumuladas de una distribución de probabilidad

Generalmente se utilizan tablas o gráficos para evidenciar la probabilidad total de un espacio muestral, las probabilidades indican el porcentaje de veces respecto a un gran número de observaciones que se espera de los diferentes resultados de una variable aleatoria.

Una manera fácil de comprender este tema es considerando una variable aleatoria X, que representa el número de caras al tirar dos veces una moneda, el espacio muestral de este experimento se visualiza utilizando el diagrama de árbol.

Primero	Segundo	Total
	CARA	CARA-CARA
CARA		
	SELLO	CARA-SELLO
	CARA	SELLO-CARA
SELLO		
	SELLO	SELLO-SELLO
2	**2**	**4**

Al ordenar los valores obtenidos en el espacio muestral del experimento se tendría lo siguiente:

Resultado	Probabilidad	Número de Sellos	P(X)
CARA-CARA	1/4	0	0.25
CARA-SELLO	1/4	1	0.25
SELLO-CARA	1/4	1	0.25
SELLO-SELLO	1/4	2	0.25

De esta forma la distribución de probabilidad para el número de sellos al tirar dos veces una moneda común es:

Número de Sellos	Probabilidad
0	0.25
1	0.50
2	0.25

La probabilidad estadística debe satisfacer dos requisitos:

a. $\sum P(X) = 1$
b. $0 \leq P(X) \leq 1$

Dónde X asume todos los valores posibles para todo valor de x.

Es importante mencionar que una probabilidad se puede describir mediante una tabla, gráficamente como se muestra en la figura 3-3 o como una función.

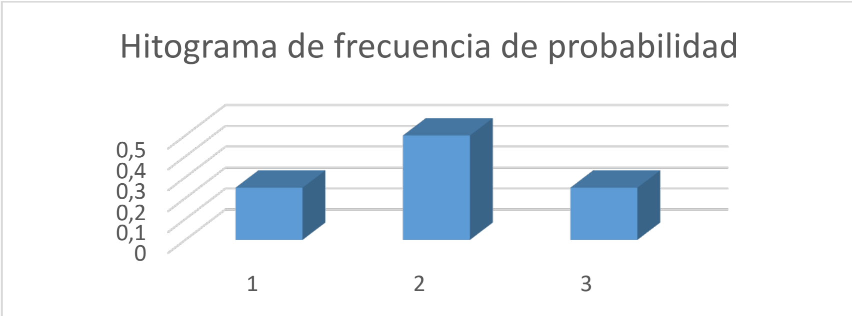


Figura 3-1. Ejemplo Histograma Frecuencias de probabilidades.

*3.7.1.1 **Histogramas***

Un histograma es la representación gráfica en forma de barras, que simboliza la distribución de un conjunto de datos, son utilizados para conseguir una "primera vista" general, o panorama, de la distribución de la población, o de la muestra, respecto a una característica, cuantitativa y continua.

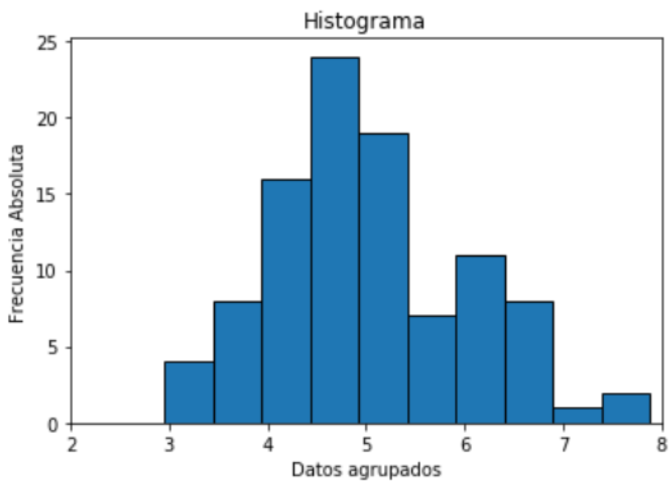


Figura 3-4. Ejemplo de histograma.

En un histograma el eje de las x (o abscisas) consiste del rango en el cual se encuentran los datos. Ahora, las bases de los rectángulos consisten de los intervalos en los cuales agrupamos dichos datos.

Por otro lado, en el eje de las y (u ordenadas) tenemos más opciones, dependiendo estas opciones es el tipo de histograma que tenemos. Los dos tipos principales de histogramas son los siguientes:

- **Histograma de frecuencias absolutas:** Representa la frecuencia absoluta mediante la altura de las barras.

- **Histograma de frecuencias relativas:** Representa la frecuencia relativa mediante la altura de las barras.

Un histograma es una gráfica adecuada para representar variables continuas, aunque también se puede usar para variables discretas. Es decir, mediante un histograma se puede mostrar gráficamente la distribución de una variable cuantitativa o numérica.

*3.7.1.2 **Función de distribución***

La función de distribución describe el comportamiento probabilístico de una variable aleatoria X asociada a un experimento aleatorio y se representa como: fx.

En una distribución de variable continua se induce probabilidad sobre todos los infinitos intervalos que integran el campo de definición de la variable. En consecuencia, ante cualquier incremento de la variable (por pequeño que sea) le corresponderá un incremento de la probabilidad de que se va acumulando, lo que hará que la función de probabilidad acumulada, la función de distribución tenga que ser continua en todos los puntos del campo de definición de la variable. Es esta la razón de que se llamen distribuciones continuas, ya que acumulan de forma continua su probabilidad.

*3.7.1.3 **Función de densidad***

La función de densidad de probabilidad de una variable aleatoria continua describe la probabilidad relativa según la cual dicha variable toma un determinado valor.

Una función de densidad de probabilidad caracteriza el comportamiento probable de una población, en tanto especifica la posibilidad relativa de que una variable aleatoria continua tome un valor cercano a x.

Una variable aleatoria X tiene la función de densidad fx una función no negativa integrable de Lebesgue si:

$$P|a \leq X \leq b| = \int_a^b f_x(x)dx \tag{3.17}$$

La definición formal de densidad requiere conceptos de la teoría de la media, una variable aleatoria continua X con valores en un espacio medible (X,A) tiene una distribución de la media X,P en (X,A), la densidad de X con respecto a la media referencia μ sobre (X,A) es la derivada de Radon- Nikodym.

$$f = \frac{dxP}{d\mu} \tag{3.18}$$

3.7.2 **Distribución de Probabilidad de Poison**

Otra distribución discreta que debe considerarse debido a su utilidad es la distribución de Poisson, En teoría de probabilidad y estadística, la distribución de Poisson es una distribución de probabilidad discreta que expresa, a partir de una frecuencia de ocurrencia media, la probabilidad de que ocurra un determinado número de eventos durante cierto período de tiempo. Concretamente, se especializa en la probabilidad de ocurrencia de sucesos con probabilidades muy pequeñas, o sucesos «raros».

Fue propuesta por Siméon-Denis Poisson, que la dio a conocer en 1838 en su trabajo Recherches sur la probabilité des jugements en matières criminelles et matière civile (Investigación sobre la probabilidad de los juicios en materias criminales y civiles).

La fórmula que se utiliza para el cálculo de probabilidades en este tipo de distribución es:

$$P(x) = \frac{\mu^x e^{-\mu}}{x!} \tag{3.19}$$

Dónde:

μ = es el número promedio de ocurrencias (éxitos) durante un intervalo específico de tiempo, también se puede calcular np

e = es la contante 2.71828

x = es el número de ocurrencias (éxitos)

Es importante mencionar que la distribución de Poisson tiene un solo parámetro que la describe, es decir; la media y la varianza de un modelo de Poisson se consideran iguales.

$$\mu = E(x) = np \tag{3.20}$$

$$\sigma = \sqrt{Var\ X} = \sqrt{np} \tag{3.21}$$

En este contexto, sabiendo que una variable aleatoria está distribuida mediante el método de Poisson y conociendo el número promedio de ocurrencias por unidad, se determina la probabilidad para cualquiera o para todos los resultados posibles.

Ejemplo 3.21

La empresa ABC de aviación pierde el equipaje de un pasajero, la mayoría de vuelos no tiene problemas en el manejo de equipaje; varios se reportan una maleta perdida, en pocos vuelos se registran la pérdida de dos maletas y es muy raro que en un vuelo se pierdan tres o más maletas. Suponiendo una muestra aleatoria de 1000 vuelos, se indica que se han perdido un total de 200; por lo tanto, como enfrentar el problema en donde n es muy grande y la probabilidad, en este caso particular se puede mencionar que la distribución es de tipo Poisson, donde la media aritmética de maletas perdidas por vuelo es de 0.2.

Si el número de maletas perdidas por vuelo sigue una distribución de Poisson con $\mu = 0.20$, es posible calcular varias probabilidades.

Se solicita calcula la probabilidad de no perder ninguna maleta, una maleta, dos maletas.

Solución:

Utilizando la fórmula, se calcula la probabilidad de que no se pierda ninguna maleta

P(x=0)

$$P(x) = \frac{\mu^x e^{-\mu}}{x!} = P(0) = \frac{0.2^0 e^{-0.2}}{0!} = 0.8187$$

Resolviendo la fórmula, se obtiene una probabilidad de 0.8187 equivalente al 81.87% de que los vuelos no tengan problema con equipajes perdidos.

Utilizando la fórmula, se calcula la probabilidad de que se pierda una maleta

P(x=1)

$$P(x) = \frac{\mu^x e^{-\mu}}{x!} = P(1) = \frac{0.2^1 e^{-0.2}}{1!} = 0.1637$$

Resolviendo la fórmula, se obtiene una probabilidad de 0.1637 equivalente al 16.3% de que los vuelos pierdan una maleta.

Utilizando la fórmula, se calcula la probabilidad de que se pierda una maleta P(x=2)

$$P(x) = \frac{\mu^x e^{-\mu}}{x!} = P(2) = \frac{0.2^2 e^{-0.2}}{2!} = 0.005458$$

Resolviendo la fórmula, se obtiene una probabilidad de 0.005458 equivalente al 0.548% de que los vuelos pierdan dos maletas.

*3.7.3 **Distribución binomial***

La distribución binomial también conocida como distribución binomial acumulativa. Una distribución binomial es una distribución de probabilidad discreta que describe el número de éxitos al realizar n experimentos independientes entre sí, acerca de una variable aleatoria.

Existen una gran diversidad de experimentos o sucesos que pueden ser caracterizados bajo esta distribución de probabilidad. Imaginemos el lanzamiento de una moneda en el que definimos el suceso "sacar cara" como el éxito. Si lanzamos 5 veces la moneda y contamos los éxitos (sacar cara) que obtenemos, nuestra distribución de probabilidades se ajustaría a una distribución binomial. Por lo tanto, la distribución binomial se entiende como una serie de pruebas o ensayos en la que solo podemos tener 2 resultados (éxito o fracaso), siendo el éxito nuestra variable aleatoria.

La fórmula para calcular este tipo de distribución es:

$$P(x) = \binom{n}{x} p^x q^{n-x} \tag{3.22}$$

$$P(x) = \frac{n!}{x!(n-x)!} p^x q^{n-x} \tag{3.23}$$

Se debe considerar que $\binom{n}{x}$ corresponde a lo que se muestra en la fórmula (3.23)

Donde:
n = Número de ensayos/experimentos
x = Número de éxitos

p = Probabilidad de éxito
q = Probabilidad de fracaso (1-p)

En el estudio de las distribuciones binomiales también se calculan las medidas de tendencia central y de dispersión, con el fin de conocer el comportamiento de los valores, en este caso la media de una distribución binomial es el promedio a largo plazo, o el valor esperado de una variable aleatoria binomial, la desviación estándar de una distribución binomial indica el nivel en que los valores muéstrales tenderán a variar a partir de la media de la distribución, a continuación se muestran la fórmula correspondientes:

$$Media\ distribución\ binimial: X = n * p$$

Ejemplo 3.22

Suponga que un 80% de personas en el mundo ha visto el partido de la final del último mundial de fútbol. Tras el evento, 4 amigos se reúnen a conversar, ¿Cuál es la probabilidad de que 3 de ellos hayan visto el partido?

Solución:

n= 4 (total de la muestra)
x = 3 (número de éxitos)
p = 0.8 (probabilidad de éxito)
q = 0.2 (probabilidad de fracaso, se obtiene restando 1-p)

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

$$P(3) = \frac{4!}{3!\,(1)!}0.8^3 0.2^1 = 0.4096$$

En este caso la probabilidad de que tres de los cuatro amigos haya visto el partido es de 40.96%.

3.7.4 **Distribución de probabilidad hiperogeométrica**

La distribución hipergeométrica es una distribución discreta que modela el número de eventos en una muestra de tamaño fijo cuando usted conoce el número total de elementos en la población de la cual proviene la muestra. Cada elemento de la muestra tiene dos resultados posibles (es un evento o un no evento).

Las condiciones que se deben cumplir para aplicar este tipo de distribución son:

- Que el tamaño de la población sea pequeño.
- Que la selección de la muestra de una población finita sea sin reemplazo, es decir mutuamente excluyente.
- Que el tamaño de la muestra n sea mayor al 5%de la población N.

Si se cumplen estas condiciones se puede aplicar la distribución hiperogeométrica para determinar la probabilidad de éxito o fracaso, la fórmula que se utiliza en esta distribución es:

$$P(x) = \frac{(sCx)(._{N-s}C_{n-x})}{_{N}C_{n}} \qquad (3.24)$$

Dónde:
N = tamaño de la población.
S = número de éxito en la población.
x = número de éxitos en la muestra.
n = tamaño de la muestra.

Ejemplo 3.23
Suponer que en una semana se fabrican 50 sillas giratorias, 40 de ellas funcionan perfectamente, 10 tuvieron algún defecto. Se selecciona al azar una muestra de 5 sin reemplazo. Utilizando la fórmula hipergeométrica determine cuál es la probabilidad de que 4 de 5 sillas funcionen perfectamente.

Solución:
Los datos evidencian un muestreo in reemplazo, el tamaño de la muestra es de 5, por lo cual representa el 10% de la población, cumple con el requerimiento para este tipo de distribución.

Dónde:
N = 50
S = 40
x = 4
n = 5

$$P(x) = \frac{(sCx)(\,_{N-S}\,C_{n-x})}{_{N}C_{n}}$$

$$P(4) = \frac{(40C4)({}_{50-10}\,C_{5-4})}{{}_{50}C_5} = \frac{913900}{2118760} = 0.4313$$

La probabilidad de seleccionar 5 sillas y encontrar que 4 funcionen perfectamente es de 0.4113 o del 41.13%.

*3.7.5 **Esperanza matemática***

En relación a la distribución de probabilidades se puede estimar estadísticos de tendencia central o dispersión, en este caso, para el cálculo de estos estadísticos se utilizan las siguientes fórmulas:

- **Media o esperanza matemática E(x).**

$$E(x) = \sum[xP(x)] \tag{3.25}$$

- **Varianza**

$$S^2 = \sum[(x_i - \bar{x})^2 * P(x)] \tag{3.26}$$

- **Desviación estándar**

$$s = \sqrt{s^2} \tag{3.27}$$

Al calcular la media de una distribución de probabilidad (esperanza matemática), se obtiene el valor promedio que se espera conseguir si los ensayos pudieran repetirse indefinidamente, se traduce en el valor promedio a largo plazo de la variable aleatoria.

Ejemplo 3.24

Leslie Martínez vende vehículos nuevos Hyundai, generalmente los sábados son los días que mayores ventas registra la Srta. Martínez, en base a su experiencia construye una distribución de probabilidad para el número de vehículos que espera vender el sábado.

Vehículos Vendidos X	Probabilidad P(x)
0	0.1
1	0.2
2	0.3
3	0.3
4	0.1
	1

Para calcular el promedio de automóviles vendido se aplica la fórmula de la esperanza matemática:

$$E(x) = \sum[xP(x)] \tag{3.28}$$

$$E(x) = (0)(0.10) + (1)(0.2) + (2)(0.3) + 3(0.3) + (4)(0.10)$$

$$E(x) = 2.1$$

El valor de 2.1 representa el promedio al día de autor que vendería la Srta. Martínez.

*3.7.2.1 **Esperanza matemática de una variable aleatoria***

La esperanza matemática de una variable aleatoria es la generalización de la media aritmética a toda la población, es decir, es la media de la variable aleatoria. También se llama **valor medio**, **valor esperado** o **esperanza matemática**, y se representa por la letra griega μ.

Si X es una variable aleatoria discreta (representada, de manera general, por una tabla de valores xi y probabilidades pi=P(X=xi); la esperanza se calcula como la media aritmética de los valores, es decir; la suma de los valores por sus probabilidades (las probabilidades representan las frecuencias relativas).

$$\mu = E(x) = \sum_{i=1}^{k} x_1 * P_i \tag{3.29}$$

Sea X : (Ω, A, P) $\longrightarrow$ ($(\mathbb{R}, B, Px)$ una variable aleatoria continua con función de densidad fx

- Existe esperanza matemática de X si $\int_{\mathbb{R}} xfx(x)dx \ < +\infty$

- Si X es una variable aleatoria continua, la variable toma infinitos valores. El equivalente continuo de la suma es la integral. La fórmula matemática incluye en este caso a la función de densidad:

$$\mu = E(X) = \int_{-\infty}^{\infty} x \cdot f(x)dx. \tag{3.30}$$

La condición para la existencia de la esperanza es debido a que la integral que la define es, en general, una integral de Lebesgue, que exige la integralidad absoluta de una función para que esta sea integrable. Dicha condición implica la finitud de la esperanza.

Ejemplo 3.25
Calcular la esperanza de una variable X con función de densidad

$fx(x) = \frac{1}{4}, -2 < x < 2$

$fx(x) = 0 \;\; para \; que \; x \;\; \notin (-2{,}2)$, se tiene:

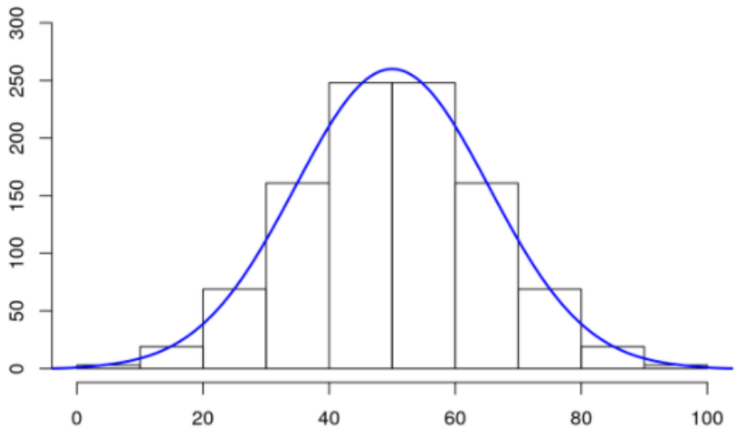
$\int_{\mathbb{R}} xfx(x)dx = \; \int_{-2}^{2} |x| \frac{1}{4} d_x \; = 0$

Entonces, existe la esperanza matemática de la variable X y vale

$E|X| = \int_{-2}^{2} x \frac{1}{4} d_x = 0$

3.7.6 **Distribución Normal**

La distribución normal fue descubierta por primera vez en el siglo XVI, los astrónomos y otros científicos observaron con asombro que las mediciones repetidas de un gran número de veces en un mismo fenómeno tendían a variación y que al reunirlas y agruparlas en distribuciones de frecuencias correspondientes, tienden a un perfil semejante a la siguiente figura.



Desde ese entonces se ha generalizado este criterio, que es aplicable a casi cualquier fenómeno, siempre y cuando cuente con un buen número de observaciones, es importante mencionar que en este tipo de distribución las observaciones tienden a agruparse en el centro.

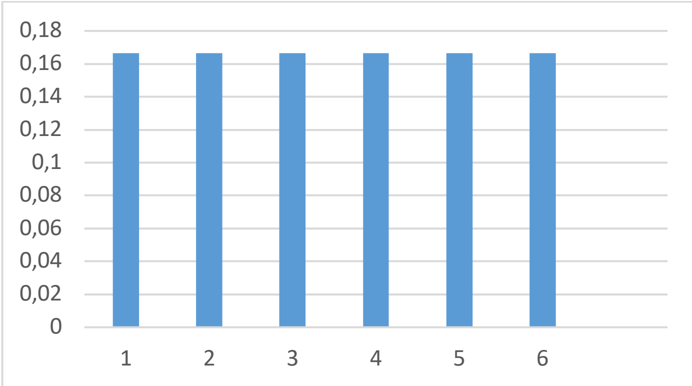
Ejemplo 3.26

Suponga que un dado tira dos veces y registra la suma del número de puntos, de tal modo que si el primer lanzamiento del dado es 3 y el segundo es 4 el total equivale a 7. ¿Cuál es la forma de la población de los números de puntos? ¿Cuáles son los posibles resultados del experimento?

Solución:

La población es una distribución uniforme en la que cada uno de los números enteros del 1 – 6 tiene la misma probabilidad de ocurrencia como se muestra a continuación

Resultado posible	**Probabilidad**
1	1/6= 0.1667
2	1/6= 0.1667
3	1/6= 0.1667
4	1/6= 0.1667
5	1/6= 0.1667
6	1/6= 0.1667



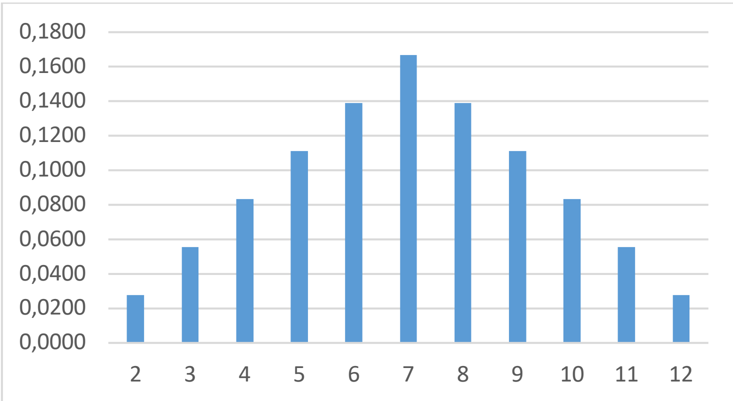
Ahora si se tira el dado dos veces, el número de resultados es 36 y se resume de la siguiente manera.

	Número Total de Puntos						
1era Tirada	6	7	8	9	10	11	12
	5	6	7	8	9	10	11
	4	5	6	7	8	9	10
	3	4	5	6	7	8	9
	2	3	4	5	6	7	8
	1	2	3	4	5	6	7
		1	2	3	4	5	6
		2da Tirada					

El número total de puntos y la probabilidad de que caiga cada uno se resumen en el siguiente diagrama:

Resultados posibles	**# Veces que aparece**	**Probabilidad**
2	1	0,0278
3	2	0,0556
4	3	0,0833
5	4	0,1111
6	5	0,1389
7	6	0,1667
8	5	0,1389
9	4	0,1111
10	3	0,0833
11	2	0,0556
12	1	0,0278

A continuación se muestra el cambio de la distribución de las sumas con pespecto a la de la población, se inició con una población unifmre para los números enteros discretos del 1-6, cuando se lanza dos veces se observa el total de puntos,la distribuciòn cambia a una forma muy parecida a la normal como se muestra a continuación.



Una vez comprendido la distribución normal, la misma puede definirse mediante una distribución matemática:

$$f(y) = \frac{e^{-(x-\mu)^2/2\alpha}}{\sigma\sqrt{2\pi}} \tag{3.31}$$

A esta distribución se le denomina gausiana, la fórmula resulta abrumadora para los estudiantes, pero no es necesario aplicarla realmente, lo que sí es importante es que cualquier distribución normal específica está determinada por dos parámetros: la media y la desviación estándar.

Las principales características de la distribución normal son:

- La curva normal tiene forma de campana.
- Es simétrica respecto a la media de distribución.
- Se extiende desde -∞ a +∞, de forma asintótica.
- Cada distribución normal es completamente definida por su media y desviación estándar.
- El área total bajo la curva entre dos puntos es igual a la probabilidad de que una variable distribuida normalmente asuma un valor de ellos.

Si se considera el área bajo la curva de cualquier distribución normal igual a 100%, en esta circunstancia se puede tomar como referencia la media y la desviación estándar, y con este criterio reordenar cualquier distribución normal para que sea expresada en una forma estandarizada.

Cuando se toma esta estandarización, cualquier valor puede ser transformado en su equivalente medio en términos de desviación estándar y con esto generar una escala denominada z, que se calcula aplicando la siguiente fórmula:

$$Z = \frac{X - \mu}{\sigma}$$

Dónde:
Z = número de desviaciones estándar a las que se encuentra el valor de interés a partir de la media.
X = valor de interés
μ= media de la distribución normal
σ= desviación estándar

Ejemplo 3.27
Convertir los valores de escala real a valor de escala estandarizada (100, 110, 120, 80) considere $\mu = 100 \;\; y \; \sigma = 10$

Solución:

$$Z = \frac{X - \mu}{\sigma}$$

$$Z = \frac{100 - 100}{10} = 0$$

$$Z = \frac{110 - 100}{10} = 1$$

$$Z = \frac{120 - 100}{10} = 2$$

$$Z = \frac{80 - 100}{10} = -2$$

*3.7.6.1 **Tabla normal estándar***

El área bajo la curva de cualquier distribución normal se puede encontrar utilizando una tabla normal estándar y cambiando a unidades reales, la media de la distribución sirve como punto de referencia y la desviación estándar como la unidad que mide distancias relativas a partir de la media.

La tabla normal estándar fue ideada de manera que pueda leer en unidades Z, la tabla muestra el área bajo la curva, es decir; la probabilidad de que un valor quede en ese intervalo, entre la media y los valores seleccionados.

Previo a la utilización de la tabla es importante considerar las siguientes observaciones:

- Los valores de probabilidad bajo la curva, proporcionando en la tabla son valores calculados a partir de la media hasta el valor de z seleccionado.

- La tabla normal se utiliza también para calcular áreas bajo la curva más allá de un valor dado de z, la clave es que la mitad del área 50% y por lo tanto el área de un valor más allá de z es igual a 50% menos el valor de z en las tablas.

- La media de distribución siempre toma el valor de cero, es decir; se encuentra a 0 desviaciones de sí misma.

- Como la distribución normal es simétrica con respecto a su media, el lado izquierdo de la curva es la imagen del lado derecho de la distribución, por esta simetría en la tabla solo se proporciona los valores de la derecha.

- Valores de z mayores a 4 se aproximan a un resultado de 0.5 o 50%.

3.7.7 ***Distribución Uniforme***

La distribución uniforme es útil para describir una variable aleatoria con probabilidad constante sobre el intervalo (a,b) en el que está definida y se denota por U (a,b); también se le denomina como distribución rectangular debido al aspecto de su función de densidad. Este tipo de distribución tiene la peculiaridad de que la probabilidad de un suceso depende exclusivamente de la amplitud del intervalo considerado y no de su posición en el campo de variación de la variable.

Sea la variable aleatoria X que puede asumir valores $x_1, x_2, \ldots x_k$ con idéntica probabilidad. Entonces la distribución uniforme discreta viene dada por:

$$f(x) = p\ (x_n = X) = \frac{1}{n} \qquad (3.32)$$

Ejemplo 3.28

Un reloj de manecillas se detuvo en un punto que no se conoce, determine la probabilidad de que se haya detenido en los primeros 25 minutos luego de señalar la hora en punto.

Solución:

Intervalo: [0 - 60]

$$f(x) = \frac{1}{60-0} = \frac{1}{60}$$

Ejercicios propuestos

Ejercicio 3.1

Utilizando la fórmula de la distribución binomial. La probabilidad de que un vuelo de la aerolínea ABC se retrase es de 0.2, de tal manera que p = 0.2. Hay cinco vuelos (n=5), x es el número de éxitos (en este caso considere que el éxito es que el avión llegue tarde). Considere que lleguen 0 vuelos tarde, 1 vuelo tarde, 2 vuelos tardes, 3 vuelos tarde. Resp. 32.77%, 40.96%, 20.48%, 5.12%

Ejercicio 3.2

Suponga que se desea determinar la probabilidad de obtener dos tréboles en cinco intentos de una baraja normal de 52 cartas barajadas. (utilice la distribución hipergeométrica). Resp. 27.43%

Ejercicios 3.3

¿Cuantos números de 2 cifras diferentes de puede formar con los dígitos 3, 4, 5? Resp. 6

Ejercicio 3.4

Lorena tiene 3 blusas diferentes y dos faldas de distinto color, ¿De cuántos modos diferentes puede vestirse? Resp. 6

Ejercicio 3.5

¿De cuantas maneras se pueden exhibir 7 juguetes diferentes, si el estante tiene sólo tres lugares disponibles? Rep. 42

Ejercicio 3.6

¿De cuantas maneras distintas se pueden sentar 5 personas alrededor de una mesa? Resp. 24

Ejercicio 3.7

¿Cuántas agrupaciones de dos elementos se pueden que se pueden formar con las letras V,W,X,Y si se permite repeticiones? Resp. 10.

Ejercicio 3.8

Calcular la probabilidad de obtener suma 5 al lanzar dos dados. Resp. 0.111

Ejercicio 3.9

Una caja contiene 3 bolas verdes, 5 bolas rojas y 2 bolas azules. Si se extraen al azar dos bolas sin reposición, ¿cuál es la probabilidad de que la primera sea azul y la segunda sea verde? Resp: 0,0667= 6.67%

Ejercicio 3.10

En el consultorio de Oscar, el 40 % de los pacientes fingen tener una enfermedad (para obtener un certificado médico). Además, el 10 % de los pacientes del consultorio son hombres. La probabilidad de que un paciente finja una enfermedad dado que es hombre, es del 50 %. Calcular la probabilidad de que un paciente sea hombre, dado que finge una enfermedad. Resp. 12.5%.

Ejercicio 3.11

En la empresa Envases del Litoral S.A se fabrican latas de dos tamaños 25 ml y 40 ml, si se sabe que hacen la misma cantidad de latas y que el 1% de latas defectuosas corresponden a 25ml y el 4% a las latas de 40 ml ¿Cuál es la probabilidad que, al seleccionar una lata al azar, esta sea de 40ml? Rep. 80%.

Ejercicio 3.12

Los sueldos mensuales en una empresa siguen una distribución normal con media de 1200 um, y desviación estándar de 200 um. ¿Qué porcentaje de trabajadores ganan entre 1000 y 1550 um? Resp. 0.8012

Ejercicio 3.13

Una fábrica de producción de agua embotellada, cuenta con una máquina de envasado automático, la cual vierte en cada botella una cierta cantidad de agua que sigue una distribución normal con media de 500 mililitros y una varianza de 25 mililitros. ¿Qué porcentaje de las botellas se llenan con agua entre 490 y 507 mililitros? Resp. 89.64%

Ejercicio 3.14

Un jugador lanza un dado corriente. Si sea igual a 1 o un número primo, gana tantos cientos de unidades monetarias como marca el dado, pero si no sale número primo, pierde tantos cientos de euros como marca el dado. Determinar la función de probabilidad y la esperanza matemática del juego. Resp 16.67%.

Ejercicio 3.15

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. 5 personas; b. al menos 3 personas y c. exactamente dos personas. Resp. 13.2%; 79.1% y 16.4%

CAPÍTULO 4
Estadística Inferencial

4. Estadística Inferencial

La estadística inferencia es un conjunto de métodos y técnicas que permiten realizar deducciones (Díaz, 2019), es decir; inferir conclusiones, propiedades, tendencias a partir de una muestra de un conjunto, su principal propósito es interpretar, hacer proyecciones y comparaciones.

Usualmente se emplea varios mecanismos que permiten hacer deducciones, entre los cuales se puede mencionar pruebas de estimación puntual, pruebas de hipótesis, pruebas paramétricas y no paramétricas, análisis de correlación y regresión, series cronológicas, análisis de varianza. Algunos ejemplos de la aplicación de la inferencia estadística son: sondeos de tendencia de voto, análisis de mercado, epidemiología médica.

4.1 Teorema del límite central

El teorema del límite central establece que si se extrae m muestra de la población de tamaño n, siendo n en todos los casos mayores a 30, en este contexto El teorema central del límite (TCL) es una teoría estadística que establece que, dada una muestra aleatoria suficientemente grande de la población, la distribución de las medias muestrales seguirá una distribución normal.

El teorema central del límite tiene una serie de propiedades de gran utilidad en el ámbito estadístico y probabilístico. Las principales son:

- Si el tamaño de la muestra es suficientemente grande, la distribución de las medias muestrales seguirá aproximadamente una distribución normal. El TCL considera una muestra como grande cuando el tamaño de la misma es superior a 30. Por tanto, si la muestra es superior a 30, la media muestral tendrá una función de distribución próxima a una normal. Y esto se cumple independientemente de la forma de la distribución con la que estamos trabajando.

- La media poblacional y la media muestral serán iguales. Es decir, la media de la distribución de todas las medias muestrales será igual a la media del total de la población.

- La varianza de la distribución de las medias muestrales será σ^2/n. Que es la varianza de la población dividido entre el tamaño de la muestra.

Este teorema se aplica tanto a suma de variables discretas como de variables continuas, los parámetros de la distribución normal son:

Media: n * m (media de la variable individual multiplicada por el número de variables independientes).

Varianza: n * s2 (varianza de la variable individual multiplicada por el número de variables individuales).

Ejemplo 4.1

Se lanza una moneda 100 veces, cuando sale cara se asigna el valor de 1 y si sale sello el valor de 0. Todo lanzamiento es una variable independiente que se distribuye según el modelo de Bernoullli con una media de 0.5 y varianza 0.25. Se solicita determinar la probabilidad de que en estos 100 lanzamientos salgan más de 70 caras.

Solución:

La variable suma de estas 100 variables independientes se distribuye, por tanto, según una distribución normal.

Media = 100*0.5 = 50
Varianza = 100*0.25 = 25

Para determinar la probabilidad de que salgan más de 70 caras se calcula la variable normal tipificada equivalente.

$$Z = \frac{X - \mu}{\sigma}$$

$$Z = \frac{70 - 50}{5} = 4$$

Z = 4, al verificar en la tabla de la distribución normal toma el valor de 0.99997, por lo tanto:

P (X > 70) = P (Z > 2,0) = 1- P (Z< 2,0) = 1 - 0,99997 = 0,00003, es decir; que la probabilidad que salga más de 70 caras es sólo de 0.003%.

4.2 ***Pruebas puntuales de estimación***

La estimación puntual se refiere al cálculo de valores que apuntan hacia el valor verdadero poblacional, el objetivo de la estimación puntual es aproximar el valor del parámetro desconocido, para ello se utiliza la información de la muestra a través de un estimados. Un estadístico es un valor obtenido de la muestra que representa a un parámetro poblacional, un buen estimador.

La estimación estadística es un proceso a través del cual se intenta determinar el valor de un parámetro de la población sin hacer un censo y a partir de la información de una muestra; una estimación es el valor numérico asignado a un parámetro y el estimados es el estadístico de la muestra utilizada para la estimación.

4.2.1 ***Propiedades de los estimadores***

Un buen estimador es una aproximación de un parámetro teórico o desconocido de una población, en este sentido las principales características se detallan a continuación:

a. **Insegado:** el valor del parámetro coincide con el valor promedio del parámetro a estimar.

b. **Consistente:** el valor de la muestra se acerque al valor del parámetro al incrementar el tamaño de la muestra.

c. **Suficiente:** el estimador usa toda la información que la muestra tiene en relación al parámetro estudiado.

d. **Eficiente:** el estimador tenga menor variabilidad que otro.

4.2.2 ***Métodos de estimación puntual***

El problema que surge en la vida real es que al buscar estimador puntual adecuado para un parámetro poblacional, no es tan sencillo de resolver, involucra una compleja matemática, los métodos más utilizados para la estimación puntual de parámetros son: método de verosimilitud y mínimos cuadrados.

a. **Método de máxima verosimilutid (ML):** este método es aplicable si cumple las siguientes condiciones:

- La distribución de la variable (parámetro de interés) debe ser conocida.
- Las variables muestrales deben ser independientes.

b. **Método de los mínimos cuadrados:** método utilizado para calcular la recta de regresión lineal que minimice los residuos, este método es aplicable bajo lo siguiente:

- Las variables muestrales Y_i, i=1, …, n, independientes deben ser de la siguiente manera:

$$Y_i = \mu_i(\theta) + e_i$$

- En la aplicación de este método no es necesario conocer la distribución de la variable asociada al parámetro.

- Su expresión general se basa en la ecuación de la recta y= mx+b, dónde m es la pendiente y b el punto de corte, viene expresada por la siguiente ecuación:

$$m = \frac{n\sum(x*y) - \sum x \sum y}{n\sum x^2 - |\sum x|^2} \qquad (4.1)$$

$$b = \frac{\sum y \sum x^2 - \sum x \sum(x*y)}{n\sum x^2 - |\sum x|^2} \qquad (4.2)$$

4.2.3 **Estimadores puntuales para parámetros**

- *Estimador puntual para la media poblacional:* según las propiedades presentadas para los estimadores de los parámetros, el estimador se obtiene utilizando el método de máxima verisimilitud, es decir, la media muestral, por lo que un valor estimado par a el parámetro μ es el valor de la variable $\bar{X}$.

- *Estimador puntual para la diferencia de dos medias poblacionales:* en este cao el estimador puntal es $\Delta\bar{X}$, luego un valor estimado para $\Delta\mu\ es\ \Delta\bar{x}$.

- *Estimador puntual para la proporción poblacional:* para la proporción Π es un estimador P € R.

4.3 *Métodos de estimación por intervalos*

La estimación por intervalos consiste en establecer el intervalo de valores donde es más probable que se encuentre el parámetro, en la figura 4-1 se muestran varios elementos de la estimación por intervalos.

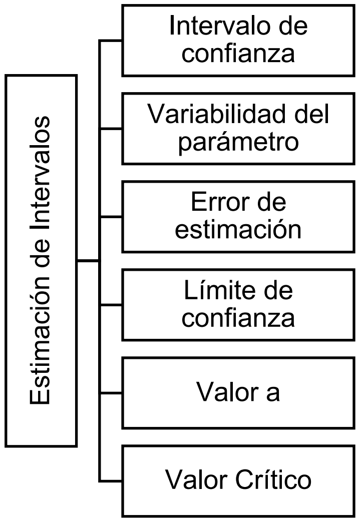


Figura 4-1. Estimación de intervalos.

La estimación por intervalo consiste en determinar un par de valores a y b, de tal manera que constituyan el intervalo [a, b]; y para una probabilidad 1-α prefijada (nivel de confianza).

4.3.1 *Construcción de intervalos*

La construcción de intervalos específicos depende de las características de la población, de los parámetros o combinaciones de parámetros a los que se les construye (media, varianza, proporción, coeficiente de correlación, entre otros).

- **Intervalo de confianza para la media de una población cualquiera, conocida la varianza.**

Las circunstancias específicas para la construcción de este tipo de intervalo son las siguientes:

$$\mu \epsilon \left[\bar{X} - \frac{\sigma}{\sqrt{n * \alpha}}; \bar{X} + \frac{\sigma}{\sqrt{n * \alpha}}\right] \quad (4.3)$$

Ejemplo 4.2

En una población cuya distribución no es conocida, se obtiene una muestra de 2000 valores, de los cuales resulta una media de 225 y una desviación típica de 10; suponiendo que la varianza muestral coincide con la poblacional, estimar un intervalo para la media con un nivel de confianza del 95%.

Datos:

1-α = 0.95

$\alpha = 0.05$

S = 10 = σ

n = 2000

$\bar{X}$ = 225

$$\mu \epsilon \left[\bar{X} - \frac{\sigma}{\sqrt{n * \alpha}}; \bar{X} + \frac{\sigma}{\sqrt{n * \alpha}}\right]$$

$$\mu \epsilon \left[225 - \frac{10}{\sqrt{2000 * 0.05}}; 225 + \frac{10}{\sqrt{2000 * 0.05}}\right]$$

$$\mu \epsilon [224; 226]$$

Con un nivel de confianza de 0.95

- **Intervalo de confianza para la media de una población normal con varianza conocida.**

Las circunstancias específicas para la construcción de este tipo de intervalo son las siguientes:

$$P\left[\bar{X} - \lambda_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + \lambda_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right] = 1 - \alpha \qquad (4.4)$$

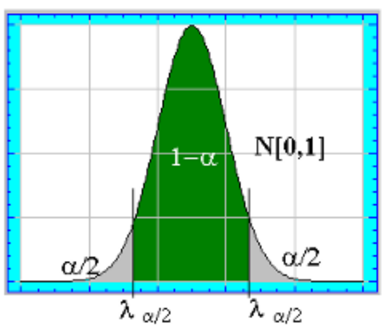


Figura 4-1. Intervalo de confianza para la media de una población normal con varianza conocida.

Siendo N el tamaño de la población[6].

Ejemplo 4.3

En una población cuya distribución no es conocida, se obtiene una muestra de 2000 valores, de los cuales resulta una media de 225 y una desviación típica de 10; suponiendo que la varianza muestral coincide con la poblacional, estimar un intervalo para la media con un nivel de confianza del 95%. Realizar la estimación de μ considerando que la población es normal.

Datos:

1-α = 0.95

$\alpha = 0.05$

S = 10 = σ (muestra grande mayor a 30)

n = 2000

$\bar{X}$ = 225

Población normal

$$P\left[\bar{X} - \lambda_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + \lambda_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right] = 1 - \alpha$$

$$P\left[225 - 0.05/2\frac{10}{\sqrt{2000}} \leq \mu \leq 225 + 0.05/2\frac{10}{\sqrt{2000}}\right]$$

$\mu\epsilon[224; 225]$ con un intervalo de confianza del 95%.

- **Intervalo de confianza para la media de una población normal de varianza desconocida (muestra pequeña).**

Las circunstancias específicas para la construcción de este tipo de intervalo son las siguientes:

$$P\left[\bar{X} - t_{\frac{\alpha}{2}}\frac{S}{\sqrt{n-1}} \leq \mu \leq \bar{X} + t_{\frac{\alpha}{2}}\frac{S}{\sqrt{n-1}}\right] = 1 - \alpha \qquad (4.5)$$

$t_{\frac{\alpha}{2}}$ =valor de t para n-1 grados de libertad y nivel de significancia.

[6] A pesar de no conocer σ (desviación poblacional), si la muestra es bastante grande n> 30, es habitual considerar la desviación típica muestral (S), como si fuera la poblacional y aplicar la estimación del intervalo.

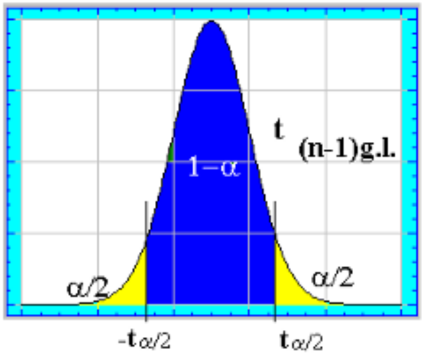


Figura 4-2. Intervalo de confianza para la media de una población normal de varianza desconocida (muestra pequeña).

Ejemplo 4.4

En las ventas diarias de una tienda comercial se evidencia una distribución normal; para estimar el volumen medio de ventas diarias se realiza una muestra de diez días seleccionados al azar, de este experimento resulta una media de ventas de 100 USD con una desviación estándar de 4 USD. Determinar un intervalo de estimación para el volumen medio de ventas diario si el nivel de confianza es de 95%.

Datos:

n = 10 (muestra pequeña)

1-α = 0.95

$\alpha = 0.05$

S = 4

σ (desconocida)

Media muestral 100

$$P\left[\bar{X} - t_{\frac{\alpha}{2}}\frac{S}{\sqrt{n-1}} \leq \mu \leq \bar{X} + t_{\frac{\alpha}{2}}\frac{S}{\sqrt{n-1}}\right] = 1 - \alpha$$

$$P\left[100 - 0.05/2\frac{4}{\sqrt{9}} \leq \mu \leq 100 + 0.05/2\frac{4}{\sqrt{9}}\right]$$

$\mu \epsilon [99.96; 100.03]$ con un intervalo de confianza del 95%.

- **Intervalo de confianza para la diferencia de medias con poblaciones normales y varianzas conocidas.**

Las circunstancias específicas para la construcción de este tipo de intervalo son las siguientes:

Intervalo para $\mu_x - \mu_y$
Conocidas σ_x, σ_y
Distribución poblacional normal.

$$P\left[(\bar{X}-\bar{Y})-\lambda_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_x^2}{n_x}+\frac{\sigma_y^2}{n_y}}\leq\left(\mu_x-\mu_y\right)\leq(\bar{X}-\bar{Y})+\lambda_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_x^2}{n_x}+\frac{\sigma_y^2}{n_y}}\right]=1-\alpha \qquad (4.6)$$

Si la varianza no fuera conocida pero la muestra fuera grande (mayor a treinta cada uno), se tomaría como varianza poblacional su homónima muestral.

- **Intervalo de confianza para la porción de una característica**

Si se desea estimar la proporción con la que se da una característica en una determinada población, esta característica se llama dicotómica, por lo que bien puede ser sí o no; el intervalo se plantea como todos con un nivel de confianza $1-\alpha$ establecido, realizando un muestreo de tamaño n aleatorio simple.

$$P\left[\hat{p}-\lambda_{\frac{\alpha}{2}}\frac{\sqrt{pq}}{\sqrt{n}}\leq p\leq\hat{p}+\lambda_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right]=1-\alpha \qquad (4.7)$$

dónde $\hat{p}$=porción de la muestra

El problema puede resolverse de varias formas de acuerdo al caso como se muestra a continuación:

- El tamaño muestral es muy grande y se podría utilizar como proporciones poblacionales las obtenidas en la muestra, es decir; $\hat{p}\hat{q}$ en lugar de pq, evidentemente si el intervalo se construye previo a la ejecución de la muestra.

- En el caso de tener información fiable de la proporción de la población, sea por estimaciones anteriores o recientes o en su defecto por un muestreo piloto, donde se puede utilizar las estimaciones como valores reales de p, y por lo tanto q=1-p.

- En la resolución de este tipo de problemas, se basa en colocarse en la situación más desfavorable posible en relación a los valores de la proporción poblacional, p=q=0.5, al aplicar esta medida en la construcción de intervalos se consigue una amplitud mayor pero menos precisa.

Ejemplo 4.5

En una investigación de carácter comercial se toma una muestra de 100 personas, resultando que 25 de estas han adquirido el producto. De un intervalo para la proporción de penetración de mercado con un nivel de confianza del 95%.

Datos:

$$1-\alpha = 0.95$$

$$n = 100\ (población\ grande)$$

$$\hat{p} = 0.25$$

El intervalo se calcula con la siguiente fórmula: $P\left[\hat{p} - \lambda_{\frac{\alpha}{2}} \frac{\sqrt{pq}}{\sqrt{n}} \leq p \leq \hat{p} + \lambda_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right] = 1 - \alpha$, donde el valor de $\lambda_{\frac{\alpha}{2}}$ = 1.96 según la tabla de normalidad (0.1) y 0.95 nivel de confianza.

Se desconoce la proporción de la población p; existen dos opciones de solución:

1. p = $\hat{p}$ = 0.25 dado que es una muestra grande, al aplicar el intervalo la proporción de mercado sería entre 16.51% y 33.48% con un nivel de confianza del 95%.

2. p=q=0.5 tomando en consideración en el caso de la varianza máxima, en este sentido la penetración de mercado sería de 15.2% y 34.8% con un nivel de confianza del 95%.

4.3.2 *Variabilidad del parámetro*

Si no se conoce, puede obtenerse una aproximación en los datos de literatura científica o en un estudio piloto, existen métodos para calcular el tamaño de la muestra que prescinde este aspecto; habitualmente se usa como medida de variabilidad la desviación típica poblacional y se denota como σ.

4.3.3 **Error de estimación**

El error de estimación es una medida de precisión que corresponde a la amplitud del intervalo de confianza, cuanto más preciso se requiera el parámetro en su estimación, más estrecho debe ser el intervalo de confianza, es decir; el error de estimación mide las posibles variaciones de la media muestral respecto a su valor real de la media poblacional, su fórmula de cálculo es:

$$E = (\sigma_2 - \sigma_1)/2 \tag{4.8}$$

4.3.4 **Límite de confianza**

El límite de confianza representa la probabilidad de que el valor verdadero del parámetro estimado en la población se ubique en el intervalo de confianza obtenido. El intervalo de confianza se denota por $1-\alpha$, usualmente se lo expresa como porcentaje, y toma un nivel de confianza ente el 95% y 99% con valores α de 0.05 y 0.01 respectivamente.

4.3.5 **Valor α**

El valor α también denominado nivel de significancia, representa la probabilidad (en tanto por uno) de fallar en la estimación, es la diferencia entre la certeza y el nivel de confianza, por ejemplo una estimación con un nivel de confianza del 95%, el valor α es (100-95)/100 = 0.05.

4.3.6 **Valor crítico**

Se representa por $Z_{\frac{\alpha}{2}}$, es el valor de la abscisa en una determinada población que deja a su derecha un área igual a $\alpha/2$, siendo $1-\alpha$ el nivel o intervalo de confianza, usualmente los valores críticos están tabulados o pueden calcularse en función de la distribución normal de la población.

4.4 **Prueba de Hipótesis**

Una prueba de hipótesis es una pauta que detalla si se puede aceptar o rechazar una afirmación a cerca de una población, en concordancia con la evidencia generada por la muestra de datos, la prueba de hipótesis evalúa la probabilidad asociada a la hipótesis nula (H_o) de que no hay efectos o diferencia.

Las pruebas de hipótesis se crearon como respuesta a la necesidad de manejar formalmente el procesamiento de datos (Díaz, 2019), en este contexto la hipótesis estadística es una explicación de los parámetros de una distribución en una población, en el caso paramétrico o sobre la forma de distribución en una población en el caso no paramétrico.

4.4.1 ***Componentes de una prueba de hipótesis***

Una hipótesis es una aseveración respecto a un problema de interés, en el caso de la estadística representa una conjetura vinculada a una población, en este contexto una prueba de hipótesis comprende cuatro componentes principales:

- **Hipótesis Nula:** denotada como H_o, especifica siempre un solo valor del parámetro de la población si la hipótesis es simple o un conjunto de valores si es compuesta.

$$H_0: \mu \ = \mu_0 \ \ H_0: \mu \ \leq \mu_0 \quad H_0: \mu \ \geq \mu_0$$

- **Hipótesis Alternativa:** se presenta por H_1, es aquella hipótesis que responde a la pregunta de investigación, se establece en función de la evidencia que se tiene; pueden existir cuatro formas:

$$H_1: \mu \ = \mu_0 \qquad H_1: \mu \ \leq \mu_0 \qquad H_1: \mu \ \geq \mu_0 \qquad H_1: \mu \ \neq \mu_0$$

Las conclusiones a las que se puede llegar en base de una muestra, general la posibilidad de equivocación, dos decisiones correctas son posibles:

Rechazar H_0 cuando es falsa o no rechazar H_0 cuando es verdadera.

Dos decisiones incorrectas son posibles:

Rechazar H_0 cuando es verdadera o no rechazar H_0 cuando es falsa, en la tabla 1 se evidencia el tipo de errores al tomar una decisión incorrecta en una prueba de hipótesis.

Tabla 4-1. Tamaño de los errores al tomar una decisión incorrecta en una prueba de hipótesis.

	H_0 Verdadera	**H_0 Falsa**
Rechazar H_0	Error Tipo 1 P(error tip I) = α	**Decisión Correcta**
No rechazar H_0	**Decisión Correcta**	Error Tipo 1 P(error tip I) = β

Nota: la probabilidad de cometer un error de tipo I se conoce como nivel de significancia y se lo representa como α y es el tamaño de la región de rechazo; el complemento de la región de rechazo es conocido como coeficiente de confianza.

- **Estadística de prueba:** es una estadística derivada del estimador puntual del parámetro que se está probando; en ella se fundamenta la decisión sobre el rechazo o aceptación de la hipótesis nula.

$$Z = \frac{\hat{\mu} - \mu_0}{\sigma_{\bar{x}}} \tag{4.9}$$

- **Región de rechazo:** es el conjunto de valores tales que, si la prueba de estadística cae en el rango de región de rechazo, la decisión es rechazar la hipótesis nula, su localización depende de la forma de la hipótesis alternativa.

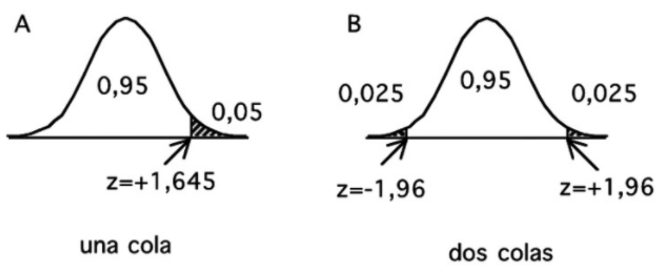


Figura 4-1. Región de rechazo de una cola y dos colas.

4.4.2 Contrataste de hipótesis

Dentro de la inferencia estadística, un contraste de hipótesis denominado test de hipótesis o prueba de significancia es un proceso para calificar si una propiedad que se supone de una población estadísticas es compatible con lo observado en la muestra de la población de estudio. Para autores como Llinas (2017), Puente (2018) & Díaz (2019) el contraste de hipótesis es una técnica estadística para decidir si se acepta o rechaza una hipótesis en términos de probabilidad, en este contexto un contraste de hipótesis se plantea siempre en término de dos hipótesis (nula y alternativa); un contraste no es una demostración de ninguna de las dos hipótesis, más bien indica si con los datos que se dispone es más efectiva una hipótesis que la otra.

Hay dos tipos de contrastes:

1. **Contrastes paramétricos:** las hipótesis se formulan sobre el valor o valores que pueden tomar un parámetro desconocido, los parámetros son la media, varianza y la proporción de p de los individuos que poseen una propiedad determinada.

2. **Contrastes no paramétricos:** las hipótesis se hacen sobre el modelo de distribución que sigue la variable que se está investigando, pueden ser los contrastes de Chi-cuadrado (modelos discretos) y Kolmogorov-Smirnow (k-s test modelos continuos).

*4.4.3 **Prueba de hipótesis para datos paramétricos***

Una hipótesis paramétrica representa una afirmación sobre una o más características de una población, este tipo de pruebas se fundamentan en la ley de distribución de la variable investigada, a pesar que existen muchas leyes de distribución, estas se basan en las normales que tienen dos parámetros: la media y la desviación estándar. Una prueba paramétrica debe tener los siguientes elementos:

- **Normalidad:** el análisis y observaciones que se genera de las muestras deben considerarse normales, para lo cual es necesario realizar pruebas de bondad de ajuste[7] en las que se describe la adaptabilidad de las observaciones y como divergen de los valores esperados.

- **Homocedastisidad:** los grupos deben presentar valore uniformes, es decir; sean homogéneos, es una característica de los modelos de regresión lineal,

[7] La bondad de ajuste de un modelo estadístico describe lo bien que se ajusta un conjunto de observaciones, las medidas de bondad de manera general se resumen en la discrepancia entre valores observados y esperados del modelo de estudio.

que implica que la varianza de los errores es constante a lo largo del tiempo.

- **Errores:** los errores que se presentan deben ser independientes, esto sucede cuando los sujetos se asignan de forma aleatoria y distribuyen de forma normal en el grupo.

*4.4.3.1 **Prueba de hipótesis para la media en población normal con varianza conocida.***

Bajo el supuesto de normalidad en la población, la varianza finita conocida y tomando como estimador de μ a $\bar{X}$ se presentan los siguientes entornos:

1. **Prueba de hipótesis de dos colas:** en este caso las hipótesis nulas y alternativas vienen dado de esta manera:

$H_0: \mu = \mu_0$ vs $H_1: \mu \neq \mu_0$

La varianza de $\bar{X}$ es : $\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$

Si la población es infinita o siendo finita, de tamaño N, es $n/N \leq 0.05$

Se toma el nivel de significancia α, los puntos críticos de acuerdo a la figura 4-4 se representan de la siguiente manera:

$$PC1 = -\frac{Z_\alpha}{2} y\ PC1 = \frac{Z_\alpha}{2}$$

Para determinar si se rechaza o no la hipótesis nula, basta con evaluar el estadístico de prueba en la muestra

$$\frac{\bar{x}-\mu_0}{\sigma_{\bar{x}}} \qquad \text{(4.10) estadístico de prueba}$$

Con la fórmula 4.10 se determina el valor del estadístico de prueba, si este valor cae en el punto crítico 1 (PC1) o a la izquierda del PC1 se rechaza la H_0 en favor de H_1; de igual manera sucede si el valor de la prueba cae en el punto crítico 2 (PC2).

Ejemplo 4.6

En la Escuela Superior Politécnica de Chimborazo suponga que el grado de satisfacción por el sueldo de los docentes auxiliares 1 es igual al grado de satisfacción por el salario de los docentes auxiliares 3. Suponga que además que estudios anteriores demuestran que el grado promedio de satisfacción con el salario de docentes auxiliares tres es de 80 y que para determinar el de los docentes auxiliares 1, se tomó una muestra de los puntajes de satisfacción de 14 docentes auxiliares 1, su media es 61.7 y la desviación estándar es de 15.3. Con un nivel de significancia de 0.05, suponga que los puntajes están normalmente distribuidos. Pruebe la hipótesis que la ESPOCH en relación al grado de satisfacción por salario entre docentes auxiliares 1 y 3.

Datos:
n = 14 docentes auxiliares 1
$\bar{x} = 61.7$ docentes auxiliares 1
σ= 15.3 docentes auxiliares 1
μ = 80 docentes auxiliares 3

La hipótesis nula y alternativa en este caso se plantean de la siguiente manera:

$$H_0 : \mu \ = 80$$

$$H_1 : \mu \ \neq 80 \text{ prueba de hipótesis de dos colas}$$

En este caso la varianza poblacional es conocida y la población de los puntajes del grado de satisfacción con el salario tiene distribución normal, se tiene que el estadístico de prueba Z es igual a:

$$Z = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}}$$

Como α = 0.05 y el problema es un problema de prueba de hipótesis de dos colas, los valores críticos son $-Z_{\alpha/2} = -1.96 \ y \ Z_{\frac{\alpha}{2}} = 1.96$.

La varianza poblacional es conocida, se tiene:

$$\sigma_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}}$$

$$\sigma_{\bar{x}} = \sqrt{\frac{15.3^2}{14}} = 4.09$$

Se procede a reemplazar este valor en el estadístico de prueba, asumiendo como válida H_0

$$Z = \frac{61.3 - 80}{4.09} = -4.47 \; < -1.96 \; PC1$$

Como este valor es menor al valor crítico se rechaza la hipótesis nula en favor de la hipótesis alternativa, es decir; que de acuerdo con los resultados los docentes auxiliares 1 tienen un grado de satisfacción diferente al de los docentes auxiliares 3.

2. **Prueba de hipótesis de cola a la izquierda:** en este caso las hipótesis nulas y alternativas vienen dado de esta manera:

$H_0: \mu \; = \mu_0$ vs $H_1: \mu \; < \mu_0$

Para determinar si se rechaza o no la hipótesis nula, basta con la evaluación del estadístico de prueba:

$$Z = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}}$$

y verificar en que región cae el valor calculado, si este valor cae en el punto crítico de la izquierda se rechaza H_0 en favor de H_1, de lo contrario no se rechaza H_0.

Ejemplo 4.7

Conociendo que el grado de satisfacción del salario de los docentes auxiliares 1 es diferente al de los docentes auxiliares 3, la pregunta es ¿En base a la información del ejemplo 4.6?, es el grado de satisfacción por salario de los docentes auxiliares 1 menor que el del salario de docentes auxiliares 3?

Datos:

n = 14 docentes auxiliares 1

$\bar{x}$ = 61.7 docentes auxiliares 1

σ= 15.3 docentes auxiliares 1

μ = 80 docentes auxiliares 3

Las hipótesis nula y alternativa para este caso se representan así:

$$H_0: \mu \; = 80$$

$$H_1: \mu < 80 \text{ prueba de hipótesis de cola a la izquierda}$$

Como α = 0.05 y el problema es un problema de prueba de hipótesis cola a la izquierda el valore crítico es $-Z_{0.05} = -1.645$.

La varianza es conocida y la población normal, el estadístico de prueba toma la forma de:

$$Z = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}}$$

La varianza poblacional es conocida, se tiene:

$$\sigma_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}}$$

$$\sigma_{\bar{x}} = \sqrt{\frac{15.3^2}{14}} = 4.09$$

Se procede a reemplazar este valor en el estadístico de prueba, asumiendo como válida H_0

$$Z = \frac{61.3 - 80}{4.09} = -4.47 < -1.645\ PC1$$

Como este valor es menor al valor crítico se rechaza la hipótesis nula en favor de la hipótesis alternativa, es decir; que de acuerdo con los resultados los docentes auxiliares 1 tienen un grado de satisfacción menor al de los docentes auxiliares 3 en relación a los salarios.

3. **Prueba de hipótesis de cola a la derecha:** en este caso las hipótesis nulas y alternativas vienen dado de esta manera:

   $H_0: \mu = \mu_0$ vs $H_1: \mu > \mu_0$

   Para determinar si se rechaza o no la hipótesis nula, basta con la evaluación del estadístico de prueba:

$$Z = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}}$$

y verificar en que región cae el valor calculado, si este valor cae en el punto crítico de la derecha se rechaza H_0 en favor de H_1, de lo contrario no se rechaza H_0.

Ejemplo 4.8

Daniela una experta en economía afirma que el grado de satisfacción del salario percibido por los docentes auxiliares 1 de la ESPOCH es mayor a 60, con base a los resultados obtenidos en el estudio de satisfacción donde la muestra generada es de 14 docentes grado 1 con una media de 61.7 con una desviación estándar de 15.3

Datos:

Muestra	**Población**
n= 14	$\mu = 60$
$\bar{X} = 61.7$	$\sigma = 15.3$

Las hipótesis nula y alternativa para este caso se representan así:

$$H_0: \mu = 60$$

$$H_1: \mu > 60 \text{ prueba de hipótesis de cola a la derecha}$$

La varianza poblacional es conocida, se tiene:

$$\sigma_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}}$$

$$\sigma_{\bar{x}} = \sqrt{\frac{15.3^2}{14}} = 4.09$$

Se procede a reemplazar este valor en el estadístico de prueba, asumiendo como válida H_0

$$Z = \frac{61.3 - 60}{4.09} = 0.416 < 1.645\ PC1$$

Como este valor es menor al valor crítico, no se rechaza la hipótesis nula, es decir; que, según los resultados obtenidos, la afirmación de Daniela no es válida, porque al parecer los docentes auxiliares 1 no tienen en promedio un grado de satisfacción mayor a 60.

4.4.3.2 Prueba de hipótesis para la media en población con varianza desconocida, muestra grande

En este contexto la población no obligatoriamente es normal, la varianza puede ser conocida o no, pero la muestra es grande, por lo que se aplica el teorema del límite central sobre la variable $\bar{X}$.

$\frac{\bar{x}-\mu_0}{\sigma_{\bar{x}}}$ con $\sigma_{\bar{X}} = \frac{S}{\sqrt{n}}$

Tiende a una variable con distribución estándar, en el caso de la media poblacional bajo estas condiciones, se presentan las siguientes situaciones:

1. **Prueba de hipótesis de dos colas:** en este caso las hipótesis nulas y alternativas vienen dado de esta manera:

$$H_0: \mu = \mu_0 \quad \text{vs} \quad H_1: \mu \neq \mu_0$$

La varianza de $\bar{X}$ es : $\sigma_{\bar{x}}^2 = \frac{S^2}{n}$

Al ser una muestra grande, entonces la variable:

$$Z = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}}$$

Se determina el valor del estadístico de prueba, si este valor cae en el punto crítico 1 (PC1) o a la izquierda del PC1 se rechaza la H_0 en favor de H_1; de igual manera sucede si el valor de la prueba cae en el punto crítico 2 (PC2).

Ejemplo 4.9

Suponga que la ESPOCH cree que el grado de satisfacción del sueldo de docentes auxiliares 1 es igual al grado de satisfacción por el sueldo de los docentes auxiliares grado 3, suponga que estudios precursores evidencian el grado de satisfacción promedio de los docentes auxiliares 3 es de 80, y para determinar el de los docentes auxiliares grado 1 se tomó una muestra de 50, donde el promedio de la media corresponde a 54.62, la desviación estándar (S) igual a 14.75, con un nivel de significancia de 0.05, pruebe la hipótesis de la ESPOCH en cuanto al grado de satisfacción por salario en docentes auxiliares 1 y 3 son iguales.

Datos

DA 1	**DA3**
n = 50	
$\overline{X}$ =54.62	μ =80
S = 14.75	

Las hipótesis nula y alternativa en este caso son:

$$H_0: \mu \ = 80$$

$H_1: \mu \ \neq 80$ prueba de hipótesis a dos colas.

Como la varianza poblacional es desconocida pero la muestra es mayor a 30, el estadístico Z toma la siguiente forma:

$$Z = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}}$$

Como α = 0.05 y el problema es un problema de prueba de hipótesis de dos colas, los valores críticos son $-Z_{\alpha/2} = -1.96 \ y \ Z_{\frac{\alpha}{2}} = 1.96$.

La varianza poblacional es conocida, se tiene:

$$\sigma_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}}$$

$$\sigma_{\bar{x}} = \sqrt{\frac{14.75^2}{50}} = 2.086$$

Se procede a reemplazar este valor en el estadístico de prueba, asumiendo como válida H_0

$$Z = \frac{54.62 - 80}{2.086} = -12.17 \ < -1.96 \ PC1$$

Como este valor es menor que el valor crítico, se rechaza la hipótesis nula y se acepta la hipótesis alternativa, es decir; que según los resultados los docentes auxiliares 1 parecen tener en promedio un grado de satisfacción diferente a la de los docentes auxiliares grado 3 en relación a los salarios.

2. **Prueba de hipótesis de cola a la izquierda:** en este caso las hipótesis nulas y alternativas vienen dado de esta manera:

$$H_0: \mu = \mu_0 \quad \text{vs} \quad H_1: \mu < \mu_0$$

Para determinar si se rechaza o no la hipótesis nula, basta con la evaluación del estadístico de prueba:

$$Z = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}}$$

y verificar en que región cae el valor calculado, si este valor cae en el punto crítico de la izquierda se rechaza H_0 en favor de H_1, de lo contrario no se rechaza H_0.

Ejemplo 4.10

La ESPOCH considera que el grado de satisfacción por el trabajo de los docentes auxiliares 3 es mayor que en los docentes auxiliares, suponga además que estudios previos evidencian que el grado de satisfacción promedio con el trabajo de docente auxiliar 3 es de 85 y que para la determinación del grado de satisfacción de docentes auxiliares 1 se tomó una muestra de 50 con una media muestral de 79 y una varianza de 68, considere un nivel de significancia de 0.05. pruebe la hipótesis de la ESPOCH en cuanto al grado de satisfacción por el trabajo entre docentes auxiliares 1 y 3.

Datos

DA 1	**DA3**
n = 50	
$\bar{X}$ =79	μ =85
S^2 = 68	

Las hipótesis nula y alternativa en este caso son:

$$H_0: \mu = 85$$

$H_1: \mu < 85$ prueba de hipótesis de cola a la izquierda.

Como la muestra es grande (n>30), por el teorema del límite central, el estadístico Z toma la siguiente forma:

$$Z = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}}$$

Como α = 0.05 y el problema es un problema de prueba de hipótesis de cola a la izquierda, el valor crítico $-Z_{0.05} = -1.645$

La varianza poblacional es desconocida, se tiene:

$$\sigma_{\bar{x}} = \sqrt{\frac{S^2}{n}}$$

$$\sigma_{\bar{x}} = \sqrt{\frac{68^2}{50}} = 1.17$$

Se procede a reemplazar este valor en el estadístico de prueba, asumiendo como válida H_0

$$Z = \frac{79 - 85}{1.17} = -5.13 < -1.645 = -Z_{0.05}$$

Como este valor es menor que el valor crítico, se rechaza la hipótesis nula y se acepta la hipótesis alternativa, es decir; que según los resultados los docentes auxiliares 1 parecen tener en promedio un grado de mayor por el trabajo diferente a la de los docentes auxiliares grado 3 en relación a los salarios.

3. **Prueba de hipótesis de cola a la derecha:** en este caso las hipótesis nulas y alternativas vienen dado de esta manera:

$$H_0\colon \mu = \mu_0 \quad \text{vs} \quad H_1\colon \mu > \mu_0$$

Para determinar si se rechaza o no la hipótesis nula, basta con la evaluación del estadístico de prueba:

$$Z = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}}$$

y verificar en que región cae el valor calculado, si este valor cae en el punto crítico de la derecha se rechaza H_0 en favor de H_1, de lo contrario no se rechaza H_0.

Ejemplo 4.11

Determine si la afirmación que el grado de satisfacción por el trabajo de los docentes auxiliares 1 en la ESPOCH no es tan grande como el que tienen los docentes auxiliares 3, pero que por lo menos es mayor a 75. Si la muestra es aleatoria de 50 y da como resultado, una media muestral de 79 y una varianza de 68, considere un nivel de significancia de 0.05.

Datos

DA 1	**DA3**
n = 50	
$\bar{X}$ =79	μ >
S^2 = 68	75

Las hipótesis nula y alternativa en este caso son:

$$H_0: \mu \ = 85$$

$H_1: \mu \ > 75$ prueba de hipótesis de cola a la derecha.

Como la muestra es grande (n>30), por el teorema del límite central, el estadístico Z toma la siguiente forma:

$$Z = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}}$$

Como α = 0.05 y el problema es un problema de prueba de hipótesis de cola a la derecha, el valor crítico $Z_{0.05} = 1.645$

La varianza poblacional es desconocida, se tiene:

$$\sigma_{\bar{x}} = \sqrt{\frac{S^2}{n}}$$

$$\sigma_{\bar{x}} = \sqrt{\frac{68^2}{50}} = 1.17$$

Se procede a reemplazar este valor en el estadístico de prueba, asumiendo como válida H_0

$$Z = \frac{79 - 75}{1.17} = 3.42 \ > 1.645 = Z_{0.05}$$

Como este valor es mayor que el valor crítico, se rechaza la hipótesis nula y se acepta la hipótesis alternativa, es decir; que según los resultados los docentes auxiliares 1 parecen tener en promedio un grado de satisfacción por el trabajo que realizan superior a 75.

4.4.3.3 ***Prueba de hipótesis para la diferencia de dos medias en población normalmente distribuidas con varianzas conocidas***

Bajo el supuesto de poblaciones normales distribuidas con varianzas finitas conocidas, se presentan las siguientes situaciones en las pruebas de hipótesis para la diferencia de dos medias poblacionales:

H_0: $\Delta\mu = \Delta\mu_0$ con $\Delta\mu_0$ fijo
H_1: $\Delta\mu \neq \Delta\mu_0$ prueba de hipótesis de dos colas.
H_1: $\Delta\mu < \Delta\mu_0$ prueba de hipótesis de cola a la izquierda.
H_1: $\Delta\mu > \Delta\mu_0$ prueba de hipótesis de cola a la derecha.
Siendo : $\Delta\mu = \mu_1 - \mu_2$, cuando $\Delta\bar{X} = \bar{X}_1 - \bar{X}_2$

En estos casos el estadístico de prueba es:

$$Z = \frac{\bar{X} - \Delta\mu_0}{\sigma_{\Delta\bar{X}}} \qquad (4.11)$$

Ejemplo 4.12

Un docente de Química tiene dos grupos, el uno química orgánica (QO) y el otro grupo de química inorgánica (QI), asevera que el presente período académico los estudiantes de QO tendrán un promedio con nota definitiva para el curso, una nota mayor que la nota promedio de los estudiantes de QI; si la varianza en el primer grupo es de 1.2 y el segundo grupo su varianza es de 0.9, el primer grupo tiene una muestra de 30 con una media de 3.9 y el segundo grupo con una muestra de 34 con una media equivalente a 3.7. Considere α=0.06.

Datos

QO	**QI**
n1 = 30	n2=34
$\bar{X}$ =3.9	$\bar{X}$ = 3.7
σ^2 = 1.2	σ^2 = 0.9

Para este caso la hipótesis nula y alternativa se representan así:

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 > \mu_2$$

Con su equivalencia:

$$H_0: \Delta\mu = \mu_1 - \mu_2 = 0$$

$H_1: \Delta\mu > 0$ prueba de hipótesis de cola a la derecha

Como las muestras son grandes, el estadístico de prueba es Z, que en este caso toma la forma:

$$Z = \frac{\overline{\Delta X} - \Delta\mu_0}{\sigma_{\Delta\bar{X}}}$$

Con:

$$\sigma_{\Delta\bar{X}} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

En las muestras:

$$\sigma_{\Delta\bar{X}} = \sqrt{\frac{1.2}{30} + \frac{0.9}{34}} = 0.26$$

Se reemplaza este valor en el estadístico Z, asumiendo como válida H_0:

$$Z = \frac{(3.9 - 3.7) - 0}{0.26} = 0.77$$

En este caso no se ha dado el nivel de significancia, se utiliza el valor p[8] para verificar la validez de H_0. En este caso el p valor viene dado por:

[8] P-valor se define como la probabilidad de que un valor estadístico calculado sea posible dada una hipótesis nula, oscila entre 0 y 1, muestra la posibilidad de haber obtenido un resultado suponiendo que la hipótesis nula sea cierta. Se suela decir que los valores de p altos no permiten rechazar la H0

$$p\ valor = P(\Delta\bar{X} \geq 0.39) = P(Z \geq 0.77) = 0.2206$$

Como el p valor es mayor a 0.05 no se rechaza la hipótesis nula, por lo que no parece válida la afirmación del profesor en relación a los puntajes de los dos grupos.

4.4.3.4 ***Prueba de Chi cuadrado***

Otro procedimiento útil para tomar decisiones y determinar la validez de una hipótesis en parámetros de una o más poblaciones es la denominada prueba de chi cuadro que se aplica a variables discretas, variables aleatorias continuas; en el caso de estas última antes de la aplicación de este estadístico se convierte la variable continua en una categoría mediante intervalos de clase, considerando cada clase como una categoría y se aplica luego el mismo procedimiento en el caso de variable discreta (Díaz, 2019. Pág 199).

- Si una variable categórica tiene el mismo comportamiento en cada una de las poblaciones estudiadas (prueba de homogeneidad)

- Si dos o más variables categóricas están o no relacionadas (prueba de independencia)

- Si un conjunto de datos se ajusta a una tal distribución (bondad de ajuste)

En cada caso, se busca determinar si la frecuencia observada (O) en cada categoría es igual a la frecuencia esperada (E). la prueba de chi cuadrado no es aplicable en el caso que la frecuencia esperada en una categoría de la variable aleatoria sea menor a 5.

$$x^2 = \sum_{i=1}^{rc} \frac{(Oi - Ei)}{Ei} \qquad (4.12)$$

Dónde:
X^2 = Chi cuadrado
O = eventos observados
E = eventos esperados.

Ejemplo 4.13

En un estudio de mercado se requiere conocer la capacidad operativa para diversificar su cartera de productos (H1).

Tabla 4-1. Datos ejemplo 4.13 prueba de Chi Cuadrado.

	¿Qué tipo de alimento consume su mascota?			
	MERCADO ACTUAL		**MERCADO POTENCIAL**	
ALTERNATIVAS	**FRECUENCIA**	**PORCENTAJE**	**FRECUENCIA**	**PORCENTAJE**
Comida procesada	208	76.20	260	95.25
Comida Casera	65	23.80	13	4.75
TOTAL	**273**	**100**	**13**	**100**

Nota: Datos tomados de Proyecto de Investigación Elaboración e Implementación de un plan de negocios en la empresa Agrotécnica para la diversificación de la cartera de productos en la ciudad de Riobamba, en el período Febrero – Agosto 2013 (Puente, 2014)

Tabla 4-2. Chi Cuadrado Calculado Ejemplo 4.13

ALTERNATIVAS	¿Qué tipo de alimento consume su mascota?					
	M. ACTUAL	**M. POTENCIAL**	**TOTAL**	**M. ACTUAL**	**M. POTENCIAL**	**TOTALES**
Comida procesada	208	260	468	O=208 E=234	O=260 E=234	468
Comida Casera	65	13	78	O=65 E=39	O=13 E=39	78
TOTAL	**273**	**273**	**546**	**273**	**273**	**546**

1. $273 * \frac{468}{546}$=234
2. $273 * \frac{78}{546}$=39
3. $273 * \frac{468}{546}$=234
4. $273 * \frac{468}{546}$=39

$$x^2 = \sum_{i=1}^{rc} \frac{(Oi - Ei)}{Ei}$$

$$x^2 = \frac{(208-234)^2}{234} + \frac{(65-39)^2}{39} + \frac{(260-234)^2}{234} + \frac{(13-39)^2}{39}$$

$$x^2 = \frac{676}{234} + \frac{676}{39} + \frac{676}{234} + \frac{676}{39}$$

x^2= 2.88 + 17.33 + 2.88 + 17.33

x^2= 40.42

x2 tabulado = (f-1)(c-1)= (2-1)(2-1) = 1 (gl)

a. Grado de libertad

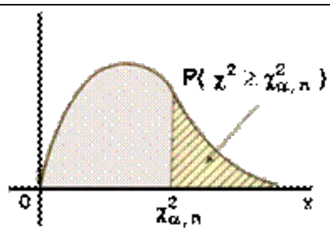
Para calcular el grado de libertad se realiza x^2 tabulado = (f-1) x (c-1)

b. Nivel de significancia

Es el error que se puede cometer al rechazar la hipótesis siendo verdadera.

Para este caso el chi-cuadrado se utilizó con un nivel de significancia del 1%. Entonces se tiene un nivel de significancia del 0.01.

Tabla 4-3. Valores de Chi- Cuadrado crítico.

DISTRIBUCIÓN DE CHI-CUADRADO					
					
	Probabilidad de un valor superior				
Grados de libertad	**0,1**	**0,05**	**0,025**	**0,01**	**0,005**
1	2,71	3,84	5,02	6,63	7,88
2	4,61	5,99	7,38	9,21	10,60
3	6,25	7,81	9,35	11,34	12,84
4	7,78	9,49	11,14	13,28	14,86
5	9,24	11,07	12,83	15,09	16,75
6	10,64	12,59	14,45	16,81	18,55
7	12,02	14,07	16,01	18,48	20,28
8	13,36	15,51	17,53	20,09	21,95
9	14,68	16,92	19,02	21,67	23,59

c. Decisión estadística

El valor de Chi cuadrado calculado es de 40.42 con 1 grado de libertad por lo tanto es mayor que el Chi cuadrado tabulado cuyo valor es de 6.63 con $\alpha = 0.01$ de significación.

Por lo tanto, queda aceptada la hipótesis de que se necesita conocer la capacidad operativa de la empresa para la diversificación de productos.

*4.4.4 **Prueba de hipótesis datos no paramétricos***

Una prueba no paramétrica es una prueba estadística que no requiere que la distribución de la población sea caracterizada por ciertos parámetros, muchas pruebas de hipótesis parten del supuesto que la población sigue una distribución normal con los parámetros μ y σ. Las pruebas no paramétricas no partes de este supuesto, de modo que son útiles cuando los datos son considerablemente no normales y resisten a transformaciones.

Los métodos no paramétricos son útiles cuando no se cumple el supuesto de normalidad y el tamaño de la muestra es pequeño, sin embargo; las pruebas no paramétricas no se liberan de los supuestos, se pueden utilizar pruebas no paramétricas si los datos responden a las siguientes preguntas:

- ¿Es la mediana de los datos investigados igual a cierto valor? Se utiliza la prueba se signos de una muestra.

- ¿Es la mediana de los datos de un evento mayor a la mediana de otro evento similar? Se utiliza la prueba de Mann-Whitney o la prueba de Kruskal-Wallis.

- ¿Son diferentes las medianas de los datos en diferentes eventos? Se utiliza la prueba de la mediana de Mood.

- ¿Cómo incide una variable en otra variable? Se utiliza la prueba de Friedman.

*4.4.4.1 **Prueba de signos de una muestra***

La prueba de signos de una muestra es utilizada para probar si los valores de una muestra son menores o mayores a los valores de otra muestra. Las variables deben ser continuas u ordinales, además los valores pueden jerarquizarse.

H_0: No existe diferencia entre los grupos.
H_1: Existe diferencia.

Ejemplo 4.14

Se genera un curso de educación financiera para 22 estudiantes de la carrera de finanzas, en la primera clase se toma una prueba de diagnóstico, cuatro semanas después se lleva una segunda prueba. Probar la hipótesis de que el curso recibido ayudó a mejorar los conocimientos de educación financiera.

H_0: El curso no ayudó
H_1: El curso ayudó a mejorar los conocimientos.

$$P_1 = \sum_{i=0}^{c} {}^{n}_{i} (\tfrac{1}{2})^n \tag{4.13}$$

Donde:
C = número de diferencias positivas
n = número de datos menos la cantidad de datos iguales al valor $\overline{\mu_0}$

En la tabla 4-5 se presentan los datos obtenidos del curso en las dos pruebas.

Tabla 4-1. Datos ejemplo 4.14

Estudiante	Prueba 1	Prueba 2
1	72	73
2	72	74
3	70	72
4	68	69
5	67	68
6	73	72
7	71	73
8	72	72
9	70	74
10	69	68
11	70	73
12	68	69
13	72	70
14	69	68
15	66	69
16	73	74

17	71	73
18	70	70
19	72	68
20	70	71
21	72	75
22	73	76

Se utiliza el software spss 25 para realizar la prueba correspondiente, en la tabla 4-6 se muestran los resultados obtenidos.

Tabla 4-2. Resultados de la prueba de signos ejemplo 4.13

Frecuencias

		N
Nota de la segunda prueba sobre 100 - Nota de la primera prueba sobre 100	Diferencias negativas[a]	5
	Diferencias positivas[b]	15
	Empates[c]	2
	Total	22

a. Nota de la segunda prueba sobre 100 < Nota de la primera prueba sobre 100
b. Nota de la segunda prueba sobre 100 > Nota de la primera prueba sobre 100
c. Nota de la segunda prueba sobre 100 = Nota de la primera prueba sobre 100

Estadísticos de prueba[a]

	Nota de la segunda prueba sobre 100 - Nota de la primera prueba sobre 100
Significación exacta (bilateral)	,041[b]

a. Prueba de los signos
b. Distribución binomial utilizada

En este caso el p valor (nivel de significancia) de 0.041 es menor a 0.05, se rechaza la hipótesis nula y se concluye que existe evidencia que el curso de educación financiera mejora el nivel de conocimiento de los estudiantes de la carrera de finanzas con un nivel de significancia del 5%.

*4.4.4.2 **Prueba de rangos con signo de Wilcoxon***

La prueba de rangos con signo de Wilcoxon es una prueba no paramétrica para comparar el rango medio de dos muestras relacionadas y determinar si existen diferencias entre ellas, no requiere de una distribución específica. Es homóloga a la prueba paramétrica t para muestra pareadas.

$$W^{+} = \sum_{Z_i} R_1 \qquad (4.14)$$

H_0: No existe diferencia entre las observaciones pareadas.
H_1: Si hay diferencias entre las observaciones pareadas.

Ejemplo 4.15

Se requiere estudiar la efectividad de una diera, por lo cual se toma una muestra aleatoria de 12 mujeres mayores a 30 años, se toma su peso en libras antes de la prueba y luego de un mes de realizar la dieta, los resultados se muestran a continuación en la tabla 4-7:

Tabla 4-1. Datos ejemplo 4.14

Paciente	Peso antes de la dieta	Peso luego de la dieta
1	186	176
2	147	149
3	128	126
4	176	177
5	212	202
6	158	158
7	204	197
8	157	160
9	189	180
10	149	150
11	191	188
12	200	195

H_0: No hay diferencia entre el peso de las mujeres antes de iniciar la dieta y el peso luego de culminar con la dieta un mes después.
H_1: El peso luego de un mes de dieta es inferior al peso inicial.

Rangos				
		N	Rango promedio	Suma de rangos
peso final - peso inicial	Rangos negativos	7[a]	7,71	54,00
	Rangos positivos	4[b]	3,00	12,00
	Empates	1[c]		
	Total	12		

a. peso final < peso inicial
b. peso final > peso inicial
c. peso final = peso inicial

Estadísticos de prueba[a]	
	peso final - peso inicial
Z	-1,871[b]
Sig. asin. (bilateral)	,041

a. Prueba de rangos con signo de Wilcoxon
b. Se basa en rangos positivos.

El p valor obtenido es de 0.041 es menor a 0.05, razón por la cual se rechaza la hipótesis nula y se concluye que hay suficientes evidencias para plantear que la dieta es efectiva en la reducción de peso con un nivel de significancia del 5%.

4.4.4.3 ***Prueba de Kruskal-Wallis***

La prueba de Kruskal-Wallis conocida como la prueba H, es una prueba no paramétrica que usa rangos de datos muestrales de tres o más poblaciones independientes, se utiliza para probar la hipótesis nula de las muestras independientes provenientes de poblaciones medianas iguales, la hipótesis alternativa tiene medianas que no son iguales.

H_0: Media1 = Media 2 = … Media k (poblaciones idénticas)
H_1: Media1 ≠ Media dj al menos para un par (i,j) (poblaciones diferentes)

$$H = \frac{12\sum_{nj}(\bar{R}_j-\bar{R})^2}{N(N+1)} \qquad (4.15)$$

Dónde:
$\bar{R}_j$=rango promedio para el grupo
$\bar{R}$ = rango promedio para todas las observaciones
N = número de observaciones
nj= número de observaciones para el grupo j esimo.

Ejemplo 4.16

Los efectos de dos medicamentos con respecto al tiempo de reacción a determinado estímulo fueron en tres grupos de ratones experimentales, el grupo III sirvió como control, muestra que el I y II se aplicaron las medicinas A y B respectivamente, a continuación, se detalla los tiempos de reacción.

Grupos	**A**	**B**	**C**
	17	8	2
Tiempo de reacción (segundos)	20	9	5
	40	7	4
	33	8	3
	35		

H_0: las tres muestras provienen de la misma población.

H_1: al menos una de las muestras proviene de una población diferente.

Rangos

	Grupo de pertenencia	N	Promedio	Suma de rangos
Tiempos de reacción (segundos)	Grupo A	5	11,00	54,00
	Grupo B	4	6,50	12,00
	Grupo C	4	6,50	
	Total	**13**		

Estadísticos de contraste a.b

	Tiempos de reacción
Chi-Cuadrado	10,711
Gl	2
Sig. asin. (bilateral)	,005

a. Prueba Krustal-Wallis

b. Variable de agrupación.

Se analizan 13 personas, 5 corresponden al grupo A, 4 al grupo B y C, el rango promedio mayor fue para el grupo A, en tanto el menor fue del grupo C, el estadístico H con 2 grados de libertad (gl) fue de 10,711, un nivel de significancia de 0,005, dicho valor es menor al p valor de 0.05, en este contexto se rechaza la hipótesis nula y se concluye que existen evidencias suficientes para plantear que la mediana de tiempo de reacción es diferente entre los grupos.

Ejercicios propuestos

Ejercicio 4.1

La renta media de los habitantes de un país se distribuye uniformemente entre 4,0. y 10,0 mil dólares. Calcular la probabilidad de que al seleccionar al azar a 100 personas la suma de sus rentas supere los 725 mil dólares. Cada renta personal es una variable independiente que se distribuye según una función uniforme. Por ello, a la suma de las rentas de 100 personas se le puede aplicar el Teorema Central del Límite.

Ejercicio 4.2

Un contador extrae una muestra aleatoria de tamaño 36 de una población de 1000 cuentas por cobrar. La media de la población de las cuentas por cobrar es de 260 USD y la desviación estándar poblacional de 45 USD. ¿Cuál es la probabilidad de que la media muestral sea menor a 250 USD?

Ejercicio 4.3

Se presenta una distribución no conocida, se obtiene una muestra de 300 valores, de los cuales resulta una media de 18.20 y una desviación típica de 5; suponiendo que la varianza muestral coincide con la poblacional, estimar un intervalo para la media con un nivel de confianza del 95%.

Ejercicio 4.4

En una panadería se muestra en sus ventas una distribución normal; para estimar el volumen medio de ventas diarias se realiza una muestra de 7 días seleccionados al azar, de este experimento resulta una media de ventas de 200 USD con una desviación estándar de 6 USD. Determinar un intervalo de estimación para el volumen medio de ventas diario si el nivel de confianza es de 95% y del 90%.

Ejercicio 4.5

Al inicio del curso de química para 12 estudiantes se tomó una prueba diagnóstica, luego de 8 semanas se realiza una prueba sorpresa de química, probar la hipótesis que luego de recibir nuevos conocimientos sobre química los estudiantes tienen bases más sólidas sobre la asignatura.

Estudiante	Prueba diagnóstica	Prueba 7segunda
1	2	7
2	2	8
3	1	8
4	6	9
5	5	10
6	4	10
7	2	10
8	3	8
9	4	7
10	7	7
11	8	9
12	9	9

REFERENCIAS BIBLIOGRÁFICAS

American Statistical Association (2016). Ethical Guidelines for Statistical Practice. Obtenido de: https://higherlogicdownload.s3.amazonaws.com

Anderson, D., Sweeney, D. & Williams, T. (2012). Estadística para los negocios (11ava ed). CENGAGE Learning.

Capalar, C., & Dönmez, A. (2016). What is scientific reseach and how can it be done?. Turkish Journal of anaesthesiology and reanimation, 44(4), 212.

Cardenas, R. (2014). Estadística en la Educación. UNID Editorial Digital

Chuhutin, A., Hansen, B., & Jespersen, S. N. (2017). Precision and accuracy of diffusion kurtosis estimation and the influence of b-value selection. NMR in Biomedicine, 30(11), e3777.

Cruz, C., Oliverez, S & González, M.(2014) Metodología de la Investigación. Grupo editorial Patria.

Díaz, M. (2019). Estadística Inferencial Aplicada. Universidad del Norte

Gamero, C. (2017). Estadística I. Elementos de Estadística Descriptiva y teoría de Probabilidad. Universidad de Málaga.

García, J., Ramos, C & Ruiz, G. Estadística Administrativa. Universidad de Cadiz.

Guerra, C. (2003). Estadística. Editorial Felix Varela

Hernández, C. & Oteyza, E. (2015). Probabilidad y Estadística. Person Edición.

Hernández, M., Hernández, S & Papia, M. (2019) Estadística inferencial 2: Aplicaciones para ingeniería. Editorial Patria

Linas, H. (2017). Estadística Inferencial. Universidad del Norte

Lind, D., Marchal, W., & Wathen, S. (2012). Estadística aplicada a los negocios y la economía (15 va ed). Mc Graw Hill

Martínez, E. (2020). Estadística. Ediciones UAPA.

Matus, R. (2010) Estadística. Instituto Politécnico

Mias, C. (2018). Metodología de la investigación, estadística aplicada: Guía práctica para investigación. Editorial Brujas.

Mondolfo, R. (1961). Problemas y métodos de investigación en la historia de la Filosofía. Edit Universitaria Buenos Aires.

Monroy, M. & Nava, N. (2018). Metodología de la Investigación. Lapislazuli Ediciones.

Monroy, S. (2008). Estadística Descriptiva. Instituto Politécnico nacional

Mora, M. & Sepulveda, P. (1999). Metodología de la Investigación

Murray & Spieguel, Estadística (4ed). Mc Graw Hil

Pérez, H. (2008). Estadística para las ciencias sociales, del comportamiento y la salud, 3ed. Cengage Learning

Puente, C. (2018). Estadística descriptiva e inferencial. Ediciones IDT.

Puente, M., Viñan, J & Aguilar, J. (2017). Planeación Financiera y presupuestaria. Escuela Superior Politécnica de Chimborazo.

Sautu, R., Freidin, B., Najmias, C., Otamendi, M., Paredes, D., Ballesteros, M., … & Trepiana, M. (2014) Metodología de la Investigación. Universidad de Buenos Aires

Tabla de distribución Normal

z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952

2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990
3.1	0.4990	0.4991	0.4991	0.4991	0.4992	0.4992	0.4992	0.4992	0.4993	0.4993
3.2	0.4993	0.4993	0.4994	0.4994	0.4994	0.4994	0.4994	0.4995	0.4995	0.4995
3.3	0.4995	0.4995	0.4995	0.4996	0.4996	0.4996	0.4996	0.4996	0.4996	0.4997
3.4	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4998
3.5	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998
3.6	0.4998	0.4998	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.7	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.8	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.9	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000
4.0	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000

Autores

JOHANNA ENITH AGUILAR REYES

Ingeniera en Estadística Informática (ESPOCH). Magister en Gestión y liderazgo educacional (UTPL). Docente – Investigador en la Escuela Superior Politécnica de Chimborazo (ESPOCH) hasta la presente. Actualmente investigadora del grupo de investigación MOODELING – ESPOCH.

NANCY ELIZABETH CHARIGUAMÁN MAURISACA

Ingeniera en Estadística Informática (ESPOCH). Magister en Matemática Básica (ESPOCH). Docente – Investigador en la Escuela Superior Politécnica de Chimborazo (ESPOCH) hasta la presente. Actualmente investigadora del grupo de investigación ESTADISMATICA – ESPOCH; coordinadora del grupo de investigación ESTADISMATICA – ESPOCH.

MARLON ERNESTO MOSCOSO MARTÍNEZ

Ingeniero en Electrónica y Control por la Escuela Politécnica Nacional (EPN), Ecuador; Magíster en Sistemas de Control y Automatización Industrial por la Escuela Superior Politécnica de Chimborazo (ESPOCH), Ecuador; y, Máster en Investigación Matemática por la Universidad Politécnica de Valencia (UPV) y la Universidad de Valencia (UV), España. En la actualidad, es Investigador Predoctoral y candidato a Doctor en Matemáticas por la UPV y la UV. Destacado en sus estudios, graduado de la secundaria con honores al mejor estudiante, como abanderado del Pabellón Nacional, y graduado de la universidad con honores Cum Laude. Posee varios años de experiencia profesional como ingeniero y profesor universitario.

Actualmente, forma parte del grupo de investigación DAMRES del Instituto de Matemática Multidisciplinar de la UPV y presta sus servicios como profesor a tiempo completo en la Facultad de Ciencias de la ESPOCH.

SEGUNDO HUGO CALDERÓN

Ingeniero Químico Universidad Estatal de Guayaquil, Magister en Gestión Ambiental Universidad Nacional de Chimborazo, Magister Scientiarum en Ingeniería Química, Magister Scientiarum en Ingeniería Química Venezuela, Docente de la Escuela Superior Politécnica del Chimborazo (ESPOCH) hasta la presente fecha, Miembro del Grupo de Energías Alternativas y Ambiente GEA – ESPOCH

2022

La Estadística como una Herramienta en la Metodología Científica

© Johanna Enith Aguilar Reyes
Nancy Elizabeth Chariguamán Maurisaca
Marlon Ernesto Moscoso Martínez
Segundo Hugo Calderón