

# ESTADÍSTICA PARA NO EXPERTOS

UNA GUÍA PASO A PASO PARA COMPRENDER DATOS

**JUAN CARLOS ANAGUANO PÉREZ**  
**CARMEN VERÓNICA VALENZUELA CHICAIZA**



Religación  
**Press**  
Ideas desde el Sur Global

**int** Instituto  
Nelson  
Torres



| Colección Estadística |

# **Estadística para no expertos**

una guía paso a paso para comprender datos

Juan Carlos Anaguano Pérez,  
Carmen Verónica Valenzuela Chicaiza

RELIGACION PRESS  
QUITO · 2023



### **Equipo Editorial**

Eduardo Díaz R. Editor Jefe  
Roberto Simbaña Q. Director Editorial  
Felipe Carrión. Director de Comunicación  
Ana Benalcázar. Coordinadora Editorial  
Ana Wagner. Asistente Editorial

### **Consejo Editorial**

Jean-Arsène Yao | Dilrabo Keldiyorovna Bakhronova | Fabiana Parra |  
Mateus Gamba Torres | Siti Mistima Maat | Nikoleta Zampaki | Silvina  
Sosa



**Religación Press**, es una iniciativa del Centro de Investigaciones CICSHAL-  
RELIGACIÓN.

Diseño, diagramación y portada: Religación Press.  
CP 170515, Quito, Ecuador. América del Sur.  
Correo electrónico: [press@religacion.com](mailto:press@religacion.com)  
[www.religacion.com](http://www.religacion.com)

# Estadística para no expertos: una guía paso a paso para comprender datos

*Statistics for non-experts: a step-by-step guide to understanding data*

*Estatísticas para não especialistas: um guia passo a passo para entender os dados*

---

**Derechos de autor:** Juan Carlos Anaguano Pérez©, Carmen Verónica Valenzuela Chicaiza©, Religación Press©

Primera Edición: 2023

Editorial: Religación Press

Materia Dewey: 310 - Colecciones de estadística general

Clasificación Thema: PBT - Probabilidad y estadística

BISCA: MAT029000 MATHEMATICS / Probability & Statistics / General

Público objetivo: Profesional/Académico

Colección: Estadística

Soporte: Digital

Formato: Epub (.epub)/PDF (.pdf)

Publicado: 2023-12-07

ISBN: 978-9942-642-39-4

Este título se publica bajo una licencia de Atribución 4.0 Internacional (CC BY 4.0)



## Citar como (APA 7)

Anaguano Pérez, J.C., y Valenzuela Chicaiza, C.V. (2023). *Estadística para no expertos: una guía paso a paso para comprender datos*. Religación Press. <https://doi.org/10.46652/ReligacionPress.124>

## Patrocinio:

Instituto Superior Tecnológico Nelson Torres.

Cayambe, Ecuador. Avenida Luis Cordero, Vía a Ayora.

ISBN: 978-9942-642-39-4



<https://press.religacion.com>



## **Revisión por pares / Peer Review**

Este libro fue sometido a un proceso de dictaminación por académicos externos. Por lo tanto, la investigación contenida en este libro cuenta con el aval de expertos en el tema, quienes han emitido un juicio objetivo del mismo, siguiendo criterios de índole científica para valorar la solidez académica del trabajo.

This book was reviewed by an independent external reviewers. Therefore, the research contained in this book has the endorsement of experts on the subject, who have issued an objective judgment of it, following scientific criteria to assess the academic soundness of the work.

## Sobre los autores



### **Juan Carlos Anaguano Pérez**

Instituto de Educación Superior Nelson Torres | Cayambe | Ecuador

<https://orcid.org/0009-0003-5883-1720>

[juan.anaguano@intsuperior.edu.ec](mailto:juan.anaguano@intsuperior.edu.ec)

Ingeniero Comercial, con un MBA en Gerencia Empresarial y una especialización en Gestión de Proyectos, ha consolidado una carrera de 25 años en roles directivos en empresas líderes a nivel nacional. Ha desempeñado funciones de jefatura y gerencia, liderando equipos y desarrollando estrategias en áreas críticas como Administración, Gestión de Personal y el Desarrollo de Mercados. Cuenta con 5 años de experiencia como profesor de Estadística en instituciones de educación superior.



### **Carmen Verónica Valenzuela Chicaiza**

Instituto de Educación Superior Nelson Torres | Cayambe | Ecuador

<https://orcid.org/0000-0002-3267-7273>

[carmen.valenzuela@intsuperior.edu.ec](mailto:carmen.valenzuela@intsuperior.edu.ec)

Ingeniero en Economía Mención finanzas, posee una maestría en Gestión de la Responsabilidad Social Corporativa, se ha desempeñado como docente de educación superior en varias universidades e institutos del Ecuador, ejerce la actividad de consultor independiente mediante la elaboración de proyectos y asesorías. Ocupa cargos de dirección investigativa, participa como revisor y par académico en varias instituciones de educación superior.

## Resumen

Este libro tiene como objetivo fundamental capacitar tanto a estudiantes como a profesionales en las habilidades y técnicas esenciales de la estadística, que desempeñan un papel vital en la investigación y en la toma de decisiones. Ofrece un enfoque integral que abarca desde la recopilación de datos y conceptos de probabilidad hasta herramientas estadísticas avanzadas. La importancia de esta obra radica en su capacidad para simplificar el proceso de aprendizaje de la estadística, permitiendo a los lectores analizar datos de manera efectiva, tomar decisiones basadas en evidencia y aplicar estos conocimientos en diversos campos profesionales, especialmente en la administración. Es una herramienta valiosa que mejora la comprensión y fomenta el autoaprendizaje.

Palabras claves: estadística, administración, investigación.

## Abstract

This book aims primarily to train both students and professionals in the essential skills and techniques of statistics, which play a vital role in research and decision making. It offers a comprehensive approach ranging from data collection and probability concepts to advanced statistical tools. The importance of this work lies in its ability to simplify the process of learning statistics, enabling readers to analyze data effectively, make evidence-based decisions, and apply this knowledge in various professional fields, especially in management. It is a valuable tool that enhances understanding and encourages self-learning.

Keywords: statistics, management, research.

## Resumo

Este livro tem como objetivo principal treinar estudantes e profissionais nas habilidades e técnicas essenciais da estatística, que desempenham um papel vital na pesquisa e na tomada de decisões. Ele oferece uma abordagem abrangente que vai desde a coleta de dados e conceitos de probabilidade até ferramentas estatísticas avançadas. A importância deste livro está em sua capacidade de simplificar o processo de aprendizado de estatística, permitindo que os leitores analisem dados de forma eficaz, tomem decisões baseadas em evidências e apliquem esse conhecimento em vários campos profissionais, especialmente na administração. É uma ferramenta valiosa que aumenta a compreensão e incentiva o autoaprendizado.

Palavras-chave: estatística, administração, pesquisa.

## Contenido

|                                  |    |
|----------------------------------|----|
| Revisión por pares / Peer Review | 7  |
| Sobre los autores                | 8  |
| Resumen                          | 9  |
| Abstract                         | 9  |
| Resumo                           | 9  |
| Prólogo                          | 18 |
| Introducción                     | 20 |

### Capítulo 1

|                                                                                                 |           |
|-------------------------------------------------------------------------------------------------|-----------|
| <b>Presentación de datos estadísticos e indicadores de<br/>tendencia central y variabilidad</b> | <b>24</b> |
| 1.1 Conceptos generales de estadística.                                                         | 25        |
| 1.1.1 Definiciones de Estadística.                                                              | 25        |
| 1.1.2 Historia de la Estadística.                                                               | 25        |
| 1.1.3 Finalidad de la Estadística                                                               | 26        |
| 1.1.4 Importancia de la Estadística                                                             | 27        |
| 1.1.5 Tipos de Estadística                                                                      | 28        |
| 1.1.6 Conceptos generales                                                                       | 28        |
| 1.1.7 Conclusión                                                                                | 30        |
| 1.1.8. Autoevaluación                                                                           | 31        |
| 1.2 Variables y gráficos                                                                        | 32        |
| 1.2.1 Tipos de variables                                                                        | 32        |
| 1.2.1.1 Variable cualitativa                                                                    | 32        |
| 1.2.1.2 Variable cuantitativa                                                                   | 32        |
| 1.2.2 Tipos de gráficos                                                                         | 32        |
| 1.2.3 Conclusión                                                                                | 36        |
| 1.2.4 Autoevaluación                                                                            | 37        |
| 1.3. Distribución de frecuencias                                                                | 38        |
| 1.3.1. Conceptos básicos                                                                        | 38        |
| 1.3.2. Tipos de distribuciones de frecuencia                                                    | 40        |
| 1.3.3. Datos agrupados                                                                          | 40        |
| 1.3.4. Propiedades de las frecuencias: absoluta, relativa y acumulada.                          | 41        |
| 1.3.5. Conclusión                                                                               | 41        |
| 1.3.6. Autoevaluación                                                                           | 42        |
| 1.4. Distribución de frecuencias                                                                | 43        |
| 1.4.1. Datos no agrupados                                                                       | 43        |
| 1.4.2. Conclusión                                                                               | 43        |
| 1.4.3. Autoevaluación                                                                           | 44        |

## Capítulo 2

### Presentación de datos estadísticos e indicadores de tendencia central y variabilidad

46

|                                                                                    |    |
|------------------------------------------------------------------------------------|----|
| 2.1. Medidas de tendencia central                                                  | 47 |
| 2.1.1. Conceptos básicos                                                           | 47 |
| 2.1.1.1 Media aritmética de la población                                           | 47 |
| 2.1.1.2 Media aritmética de la muestra                                             | 47 |
| 2.1.1.3. Media aritmética ponderada                                                | 48 |
| 2.1.1.4. Mediana                                                                   | 48 |
| 2.1.1.5. Moda                                                                      | 48 |
| 2.1.1.6. Media geométrica                                                          | 48 |
| 2.1.2. Conclusión                                                                  | 48 |
| 2.1.3. Autoevaluación                                                              | 49 |
| 2.2. Medidas de dispersión: varianza, desviación típica y coeficiente de variación | 50 |
| 2.2.1. Conceptos básicos                                                           | 50 |
| 2.2.1.1. Rango                                                                     | 50 |
| 2.2.1.2. Desviación media                                                          | 50 |
| 2.2.1.3. Desviación estándar o típica                                              | 51 |
| 2.2.1.4. Varianza                                                                  | 51 |
| 2.2.1.5. Coeficiente de variación                                                  | 52 |
| 2.2.2. Conclusión                                                                  | 52 |
| 2.2.3. Autoevaluación                                                              | 53 |
| 2.3. Medidas de posición: cuartiles, deciles y percentiles                         | 54 |
| 2.3.1. Conceptos básicos                                                           | 54 |
| 2.3.1.1. Cuartiles                                                                 | 54 |
| 2.3.1.2. Deciles                                                                   | 54 |
| 2.3.1.3. Percentiles                                                               | 54 |
| 2.3.2. Conclusión                                                                  | 54 |
| 2.3.3. Autoevaluación                                                              | 55 |
| 2.4 Momento, sesgo y curtosis                                                      | 56 |
| 2.4.1 Conceptos básicos                                                            | 56 |
| 2.4.2 Conclusión                                                                   | 58 |
| 2.4.3. Autoevaluación                                                              | 58 |

## Capítulo 3

### Probabilidades, regresión y correlación

60

|                                           |    |
|-------------------------------------------|----|
| 3.1 Teoría elemental de probabilidades    | 61 |
| 3.1.1 Conceptos básicos                   | 61 |
| 3.1.2 Enfoques de probabilidades          | 62 |
| 3.1.2.1 Enfoque objetivo                  | 62 |
| 3.1.2.2 Probabilidad subjetiva            | 63 |
| 3.1.3 Reglas para calcular probabilidades | 64 |

|                                                                    |    |
|--------------------------------------------------------------------|----|
| 3.1.3.1 Reglas de la adición                                       | 64 |
| 3.1.3.2 Reglas de la multiplicación                                | 66 |
| 3.1.4 Conclusión                                                   | 67 |
| 3.1.5 Autoevaluación                                               | 68 |
| 3.2 Principios y técnicas de conteo: permutaciones y combinaciones | 69 |
| 3.2.1. Permutaciones                                               | 69 |
| 3.2.1.1. Ejercicio en clase                                        | 69 |
| 3.2.2 Combinaciones                                                | 69 |
| 3.2.3 Conclusión                                                   | 70 |
| 3.2.4 Autoevaluación                                               | 70 |
| 3.3 Distribuciones probabilísticas                                 | 71 |
| 3.3.1 Distribución de probabilidad                                 | 71 |
| 3.3.2 Características de una distribución de probabilidad          | 71 |
| 3.3.3 Conclusión                                                   | 73 |
| 3.3.4 Autoevaluación                                               | 74 |
| 3.4 Distribución binomial y de Poisson                             | 75 |
| 3.4.1 Conceptos básicos                                            | 75 |
| 3.4.2 Distribución binomial                                        | 75 |
| Ejercicio en clase                                                 | 76 |
| 3.4.3 Distribución de Poisson                                      | 77 |
| Ejercicio en clase                                                 | 77 |
| 3.4.4 Conclusión                                                   | 78 |
| 3.4.5 Autoevaluación                                               | 79 |

## **Capítulo 4**

### **Probabilidades, regresión y correlación**

**81**

|                                                       |     |
|-------------------------------------------------------|-----|
| 4.1. Distribución normal                              | 82  |
| 4.1.2. Características de la distribución normal      | 83  |
| 4.1.3. Distribución de probabilidad normal estándar   | 85  |
| 4.1.4. Distribución de probabilidad normal estándar   | 89  |
| 4.1.5. Conclusión                                     | 90  |
| 4.1.6. Autoevaluación                                 | 91  |
| 4.2. Análisis de regresión                            | 92  |
| 4.2.1. Principio de los mínimos cuadrados             | 92  |
| 4.2.2. Conclusión                                     | 94  |
| 4.2.3. Autoevaluación.                                | 95  |
| 4.3. Análisis de correlación                          | 96  |
| 4.3.1. Coeficiente de correlación                     | 96  |
| 4.3.2. Características del coeficiente de correlación | 98  |
| 4.3.3. Conclusión                                     | 98  |
| 4.3.4. Autoevaluación.                                | 100 |
| 4.4. Pronósticos y series                             | 101 |
| 4.4.1. Introducción                                   | 101 |

|                                                                           |     |
|---------------------------------------------------------------------------|-----|
| 4.4.2. Variación en las series de tiempo                                  | 102 |
| 4.4.2.1. Tendencia secular                                                | 102 |
| 4.4.2.2. Fluctuación cíclica                                              | 103 |
| 4.4.2.3. Variación estacional                                             | 103 |
| 4.4.2.4. Variación irregular                                              | 103 |
| 4.4.3. Análisis de tendencia                                              | 105 |
| 4.4.3.1. Razones para estudiar tendencias                                 | 105 |
| 4.4.3.2. Ajuste de la tendencia lineal con el método de mínimos cuadrados | 107 |
| 4.4.4. Variación Cíclica                                                  | 108 |
| 4.4.4.1. Método de residuos                                               | 109 |
| 4.4.5 Variación estacional                                                | 111 |
| 4.4.6. Método de razón de promedio móvil                                  | 112 |
| 4.4.7. Variación irregular                                                | 113 |
| 4.4.8. Conclusión                                                         | 113 |
| 4.4.9. Autoevaluación                                                     | 114 |

|                    |            |
|--------------------|------------|
| <b>Referencias</b> | <b>116</b> |
|--------------------|------------|

|              |            |
|--------------|------------|
| <b>Anexo</b> | <b>119</b> |
|--------------|------------|

## Tablas

Tabla 1. Serie de datos.

102

## Gráficos

|                                                                                                                   |     |
|-------------------------------------------------------------------------------------------------------------------|-----|
| Gráfica 1. Gráfica de barras.                                                                                     | 33  |
| Gráfica 2. Gráfica de pastel.                                                                                     | 33  |
| Gráfica 3. Histograma.                                                                                            | 34  |
| Gráfica 4. Polígono de frecuencias.                                                                               | 34  |
| Gráfica 5. Ojiva o polígono de frecuencias acumuladas.                                                            | 35  |
| Gráfica 6. Diagrama de puntos.                                                                                    | 36  |
| Gráfica 7. Tipos de curvas con diferentes sesgos.                                                                 | 56  |
| Gráfica 8. Comparación de dos curvas sesgadas y dos curvas con la misma posición central pero diferente curtosis. | 56  |
| Gráfica 9. Tipos de curtosis.                                                                                     | 57  |
| Gráfica 10. Enfoques de la probabilidad                                                                           | 63  |
| Gráfica 11. Curva con forma de campana de la distribución normal.                                                 | 82  |
| Gráfica 12. Distribución normal que tienen la misma desviación estándar.                                          | 83  |
| Gráfica 13. Distribución normal con la misma media y diferentes desviaciones estándar.                            | 84  |
| Gráfica 14. Áreas bajo la curva de cualquier distribución normal.                                                 | 85  |
| Gráfica 15. Distribución normal estándar.                                                                         | 85  |
| Gráfica 16. Áreas bajo la curva.                                                                                  | 86  |
| Gráfica 17. Áreas bajo la curva con intervalos.                                                                   | 88  |
| Gráfica 18. Áreas bajo la curva con intervalos.                                                                   | 89  |
| Gráfica 19. Áreas bajo la curva.                                                                                  | 89  |
| Gráfica 20. Análisis de regresión, diagrama de puntos.                                                            | 93  |
| Gráfica 21. Diagrama de dispersión con correlación negativa perfecta y correlación positiva perfecta.             | 97  |
| Gráfica 22. Diagrama de correlación negativa y correlación positiva.                                              | 97  |
| Gráfica 23. Variación en las series de tiempo.                                                                    | 104 |
| Gráfica 24. Relaciones de tendencia curvilínea.                                                                   | 106 |
| Gráfica 25. Fluctuaciones cíclicas alrededor de la línea de tendencia.                                            | 110 |
| Gráfica 26. Gráfica del porcentaje de tendencia alrededor de la línea de tendencia.                               | 111 |



| Colección Estadística |

# Estadística para no expertos:

una guía paso a paso para comprender datos



## Prólogo

El presente libro es de naturaleza Teórica-Práctico, con mayor incidencia en la parte práctica, tiene como propósito el desarrollo de aprendizaje que permita al estudiante recolectar, procesar e interpretar datos para tomar decisiones, a partir de la determinación de medidas y pruebas estadísticas.

El propósito, es proporcionar al estudiante los conocimientos y técnicas básicas de la estadística para utilizarlas como herramienta, en el campo de la investigación, presentando los datos en forma ordenada, en tablas y gráficos estadísticos, permitiendo procesar correctamente los estadígrafos de tendencia central, de dispersión y uso la probabilidad existente entre dos variables.

Para cumplir con dichos objetivos, el libro hace una exposición clara y sistemática de los conceptos, métodos, procesos y sistemas de evaluación y valoraciones relevantes; orientados en todo momento a asociarlo a la formación de los tecnólogos en Administración.

Los contenidos de la asignatura están estructurados en dos unidades. la Unidad 1: “Estadística descriptiva, probabilidades y funciones densidad de probabilidad” se abordarán temas relacionados al estudio de la población y muestra, variables y gráficos, distribuciones de frecuencias, medidas de tendencia central, medidas de dispersión, momento, sesgo y curtosis, probabilidades y funciones densidad de probabilidad, teoría elemental de probabilidades y finalmente una evaluación de la unidad.

En la Unidad 2: “Probabilidades y funciones densidad de teoría probabilidad”, en este capítulo se estudiará distribuciones: binomial, normal, de Poisson, teoría elemental de muestreo, teoría de estimación estadística, estadística inferencial, teoría estadística de las decisiones, teoría de pequeñas muestras, test Ji-cua-

drado, ajuste de curvas y el método de mínimos cuadrados, teoría de la correlación y finalmente una evaluación de la unidad.

## Introducción

La asignatura de Métodos Estadísticos es de naturaleza Teórica-Práctica, con mayor incidencia en la parte práctica, tiene como propósito el desarrollo de aprendizaje que permita al estudiante recolectar, procesar e interpretar datos para tomar decisiones, a partir de la determinación de medidas y pruebas estadísticas.

El propósito de la asignatura es proporcionar al estudiante los conocimientos y técnicas básicas de la estadística para utilizarlas como herramienta, en el campo de la investigación, presentando los datos en forma ordenada, en tablas y gráficos estadísticos, permitiendo procesar correctamente los estadígrafos de tendencia central, de dispersión y uso la probabilidad existente entre dos variables.

Para cumplir con dichos objetivos, El texto nos proponemos hacer una exposición clara y sistemática de los conceptos, métodos, procesos y sistemas de evaluación y valoraciones relevantes; orientados en todo momento a asociarlo a la formación de los tecnólogos en Administración.

Los contenidos de la asignatura están estructurados en dos unidades, una en cada bimestre. Dentro del primer bimestre en la Unidad 1: “Estadística descriptiva, probabilidades y funciones densidad de probabilidad” se abordarán temas relacionados al estudio de la población y muestra, variables y gráficos, distribuciones de frecuencias, medidas de tendencia central, medidas de dispersión, momento, sesgo y curtosis, probabilidades y funciones densidad de probabilidad, teoría elemental de probabilidades y finalmente una evaluación de la unidad.

En el segundo bimestre en la Unidad 2: “Probabilidades y funciones densidad de teoría probabilidad”, en este capítulo se

estudiará distribuciones: binomial, normal, de Poisson, teoría elemental de muestreo, teoría de estimación estadística, estadística inferencial, teoría estadística de las decisiones, teoría de pequeñas muestras, test Ji-cuadrado, ajuste de curvas y el método de mínimos cuadrados, teoría de la correlación y finalmente una evaluación de la unidad.

Se espera que la presente guía didáctica apoye en el proceso de enseñanza y aprendizaje de la asignatura y fortalezca el autoaprendizaje.

## **Objetivos**

Proporcionar al estudiante los conocimientos y técnicas básicas de la estadística para utilizarlas como herramienta, en el campo de la investigación, presentando los datos en forma ordenada, en tablas y gráficos estadísticos, permitiendo procesar correctamente los estadígrafos de tendencia central, de dispersión y uso la probabilidad existente entre dos variables. Además, ayudará al estudiante a tener una mayor capacidad de análisis y síntesis, resolución de problemas y razonamiento crítico

Orientaciones básicas para el estudio.

Material didáctico.

- La bibliografía incluida en el texto, le servirá para el abordaje de los temas descritos en el syllabus de la asignatura.
- Este material desarrolla contenidos de las unidades académicas 1 y 2, le brinda indicaciones claras y precisas sobre las actividades que tienen que desarrollar en cada uno de los temas y subtemas.

## **Planificación del estudio**

- Organice sus actividades laborales, personales y familiares de tal manera de que pueda disponer de horas para el desarrollo de las actividades autónomas como las conexiones obligatorias de la asignatura.
- Le sugerimos dedicar dos horas diarias a sus actividades académicas: revisión de materiales, preparación de trabajos autónomos, evaluaciones y conexiones obligatorias con el docente.

## **Técnica de estudio**

- Organice un espacio específico para realizar sus actividades de estudio y organice en esta sus recursos físicos.

## **Competencias de la asignatura de Métodos Estadísticos**

La asignatura de Métodos Estadísticos es de naturaleza Teórica-Práctica, con mayor incidencia en la parte práctica, tiene como propósito el desarrollo de técnicas que permita al estudiante recolectar, procesar e interpretar datos para tomar decisiones en cualquier entorno laboral. Para ello, se desarrollan los principales instrumentos estadísticos que sirven para la descripción, resumen y comprensión de la información disponible. Asignatura que contribuye alcanzar el perfil de egreso establecido. Respecto al aporte de las bases cognitivas con otras asignaturas contribuye con el pensamiento crítico, analítico y reflexivo.

El estudiante desarrolla sus capacidades y logra analizar los resultados de los diferentes temas de métodos estadísticos, permitiéndole comprender la importancia del estudio de estadística y su implicación en el mundo actual.



## **Capítulo 1**

Presentación de datos estadísticos e indicadores de tendencia central y variabilidad

## **1.1 Conceptos generales de estadística.**

### **1.1.1 Definiciones de Estadística.**

Es considerada por algunos autores como:

- Ciencia que recoge, organiza, presenta, analiza e interpreta datos con el fin de propiciar una toma de decisiones más eficaz (Lind et al., 2012, p. 5).
- La Estadística es la ciencia cuyo objetivo es reunir una información cuantitativa concerniente a individuos, grupos, series de hechos, etc.; y deducir de ello, gracias al análisis de estos datos, unos significados precisos o unas previsiones para el futuro (Galindo, 2011, p. 1).

Sin embargo, hay quienes prefieren no encasillarla como una rama o ciencia y la definen como un arte o un método:

- La palabra estadística se refiere a un sistema o método usado en la recolección, organización, análisis y descripción numérica de la información. También se puede decir que la estadística estudia el comportamiento de hecho o fenómenos de grupo (Martínez, 2012, p. 2).

### **1.1.2 Historia de la Estadística.**

El uso de la estadística tiene su origen en la antigua China, donde el emperador Yao (2239 A.C.) ordenó la realización de un censo, y también se menciona en la época del Rey Herodes en relación con un empadronamiento relacionado con el nacimiento de Jesús.

Como disciplina formal, la estadística se originó en Alemania en el siglo XVII, durante el reinado de Godofredo de Achewa-

II (1719-1772), cuando se acuñó la palabra “estadística” y se distinguió de la sociología.

En la actualidad, la estadística ha experimentado un considerable desarrollo gracias al trabajo de numerosos investigadores dedicados a perfeccionar métodos estadísticos y al avance tecnológico. Esto ha permitido la selección de procedimientos que buscan mejorar la precisión de los resultados y reducir los márgenes de error.

El origen de la palabra “estadística” es motivo de debate. Algunos sostienen que proviene de la palabra “status,” que significa “estado,” mientras que otros sugieren que se deriva del término griego “statera,” que significa “balanza.” Una explicación más fiable la vincula con la palabra alemana “Stara,” que significa “estado” y se relaciona con la persona encargada de realizar las estadísticas (Martínez, 2012).

### **1.1.3 Finalidad de la Estadística**

En términos generales, se reconoce que la estadística tiene como objetivo principal proporcionar información, y su efectividad depende en gran medida de los objetivos que se busquen alcanzar y de los métodos utilizados para recopilar datos. Con el aumento de la especialización en las organizaciones empresariales, se ha vuelto crucial el uso constante de enfoques estadísticos, lo cual se puede observar en áreas como la producción en serie y el control de calidad. El propósito fundamental de la estadística es:

- Conocer la realidad de una observación o fenómeno
- Determinar lo típico o normal de esa observación
- Determinar los cambios que presenta el fenómeno

- Relacionar dos o más fenómenos
- Determinar las causas que originan el fenómeno
- Hacer estimativos sobre el comportamiento futuro del fenómeno
- Obtener conclusiones de un grupo menor (muestra), para hacerlas extensivas a un grupo mayor (población) (Martínez, 2012, p. 5).

### **1.1.4 Importancia de la Estadística**

Según Quintero (2019), la estadística es la piedra angular del conocimiento concreto y práctico, desempeñando un papel esencial en la comprensión de cómo se desarrollan diversos eventos. Ha adquirido un papel crucial en la investigación y se utiliza como una herramienta valiosa en una amplia variedad de campos de estudio. Se presenta como un lenguaje que facilita la comunicación de información basada en datos cuantitativos.

La relevancia de la estadística es tan significativa que prácticamente no existe actividad humana que no se vea influenciada por ella. Incluso las decisiones más cruciales de nuestras vidas se toman con base en la aplicación de principios estadísticos. Para ilustrar este punto, consideremos algunos ejemplos. La estadística desempeña un papel de suma importancia en la investigación científica debido a que:

- Permite una descripción más exacta.
- Nos obliga a ser claros y exactos en nuestros procedimientos y en nuestro pensar.
- Permite resumir los resultados de manera significativa y cómoda.

- Nos permite deducir conclusiones generales.
- También conocer sobre el tema con el cual se trabajará a lo largo de los años en cualquier profesión y administración.
- Aplicar apropiadamente los métodos estadísticos en la recolección de información y procesos matemáticos básicos en cálculos estadísticos.
- Adquirir los conocimientos y habilidades sobre el tema, ser capaz de reconocer los elementos habituales de la estadística (p. 2)

### 1.1.5 Tipos de Estadística

En términos generales, el campo de la estadística se clasifica en dos categorías principales: estadística descriptiva y estadística inferencial.

La estadística descriptiva se enfoca en técnicas y métodos utilizados para organizar, resumir y presentar datos de una manera que sea informativa y comprensible.

La estadística inferencial se refiere a los métodos empleados para deducir o estimar características o propiedades de toda una población basándose en la información obtenida de una muestra representativa de esa población (Lind et al., 2012).

### 1.1.6 Conceptos generales

- **Población:** Se refiere al conjunto de individuos u objetos que son de interés o las medidas que se obtienen a partir de todos esos individuos u objetos.
- **Población finita:** Se da cuando la población tiene un número limitado.

- **Población infinita:** Cuando la población tiene un número ilimitado de elementos.
- **Muestra:** Representa una porción o parte de la población que es de interés.
- **Marco:** Se llama marco, marco muestral o marco de referencia a una lista, mapa u otro material que contiene todas las unidades o elementos claramente identificados y actualizados, a partir del cual se selecciona la muestra.
- **Características (caracteres):** Corresponden a los rasgos, cualidades o propiedades que poseen los elementos que componen la población o la muestra.
- **Dato:** Hace referencia a las mediciones o valores obtenidos en un estudio estadístico.
- **Variable:** En estadística, una variable es una característica o cualidad de un individuo que puede tomar diferentes valores, y estos valores pueden medirse. Ejemplos de variables son el color de cabello, las calificaciones en un examen, el género o la altura de una persona.
- **Cualitativos (Descriptivos o categóricos):** Estos son tipos de variables que describen características que no se pueden medir, como el sexo de un animal, el color de los zapatos o la ocupación de una persona.
- **Cuantitativos (Numéricos):** Estas son variables que describen características que se pueden medir, como la temperatura ambiente, el número de hijos de una pareja o el salario de una persona.

### **1.1.7 Conclusión**

La estadística descriptiva, se enfocan en la recolección, presentación, agrupación, análisis, interpretación, proyección e inferencia de datos, desempeña un papel fundamental en la formación académica de los estudiantes y en diversas áreas de aplicación. Los estudiantes obtienen herramientas esenciales para comprender y comunicar patrones, tendencias y características de conjuntos de datos, lo que les permite tomar decisiones informadas y respaldadas por evidencia.

Los beneficios que brinda el estudio de la estadística descriptiva, es muy importante en los estudiantes, ya que desarrollan habilidades críticas para la interpretación de datos y la toma de decisiones basadas en evidencia sólida. Aprenden a identificar patrones, tendencias y valores atípicos en los datos, lo que les permite tomar medidas correctivas o aprovechar oportunidades.

Los conocimientos adquiridos en estadística descriptiva son aplicables en una variedad de disciplinas, incluyendo las ciencias naturales, sociales, la economía, la medicina, la psicología, la ingeniería y muchas otras áreas. En la investigación científica, la estadística descriptiva es esencial para resumir y presentar resultados, así como para determinar la significancia estadística de hallazgos. En el ámbito empresarial, ayuda en la toma de decisiones informadas, el análisis de tendencias del mercado y la evaluación de la eficacia de estrategias comerciales. En la medicina y la salud, permite analizar datos clínicos y epidemiológicos para entender la prevalencia de enfermedades y evaluar la efectividad de tratamientos.

### **1.1.8. Autoevaluación**

1. ¿Cuál es la finalidad de la estadística?
2. ¿Cuántos tipos de estadística existen?
3. ¿Cuál es la diferencia entre la estadística descriptiva y estadística inferencial?

## **1.2 Variables y gráficos**

### **1.2.1 Tipos de variables**

Existen dos tipos básicos de variables: 1) cualitativas y 2) cuantitativas

#### *1.2.1.1 Variable cualitativa*

Cuando se enfoca en una característica que no se expresa en términos numéricos, se le denomina variable cualitativa o atributo. Algunos ejemplos de variables cualitativas incluyen el género, la afiliación religiosa, el tipo de automóvil que una persona posee.

#### *1.2.1.2 Variable cuantitativa*

Cuando la variable que se está analizando se presenta en términos de números, se la denomina variable cuantitativa. Las edades de los estudiantes de un paralelo, el número de empleados el departamento de producción.

(Lind et al., 2012).

### **1.2.2 Tipos de gráficos**

**Gráfica de barras.** – En este tipo de representación, las categorías se muestran en el eje horizontal, mientras que en el eje vertical se representan las frecuencias de cada categoría. La altura de las barras es proporcional a las frecuencias de las categorías.

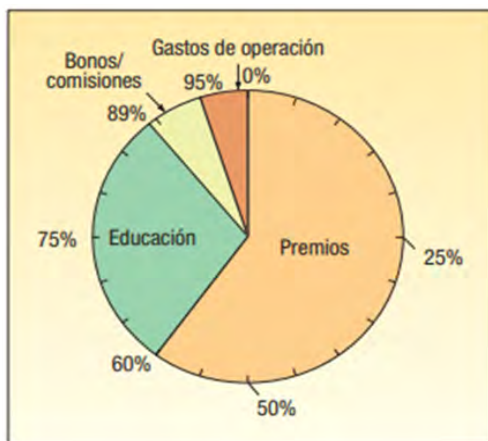
Gráfica 1. Gráfica de barras.



Fuente: Lind et al., 2012.

**Gráfica de pastel:** Representa la proporción o el porcentaje que cada categoría tiene con respecto al total de la frecuencia.

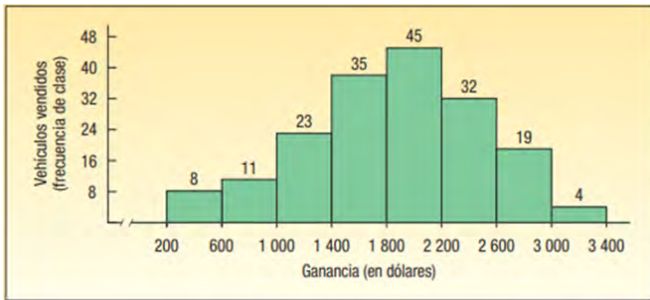
Gráfica 2. Gráfica de pastel.



Fuente: Lind et al., 2012.

- **Histograma:** En esta representación gráfica, las categorías se indican en el eje horizontal, mientras que en el eje vertical se muestran las frecuencias de cada categoría. Las alturas de las barras, que se sitúan contiguas entre sí, representan las frecuencias de las categorías.

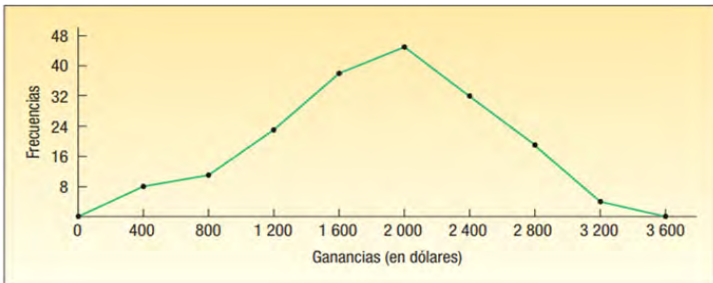
Gráfica 3. Histograma.



Fuente: Lind et al., 2012.

**Polígono de frecuencias:** Este tipo de representación gráfica, similar a un histograma, proporciona información sobre la forma de una distribución. Se compone de segmentos de línea que conectan los puntos en los cuales se cruzan los valores medios de las categorías y las frecuencias correspondientes de esas categorías.

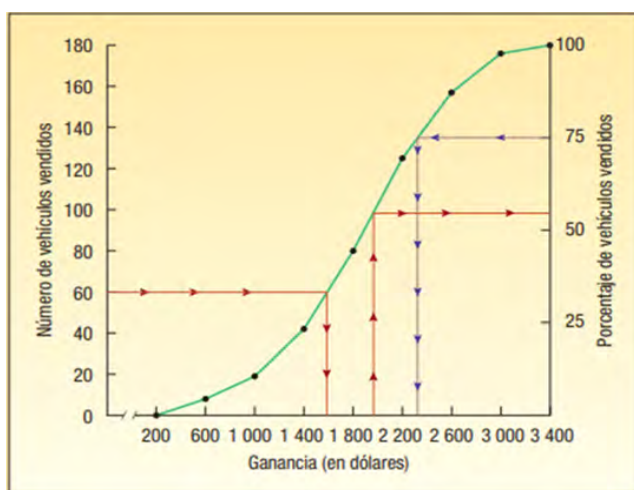
Gráfica 4. Polígono de frecuencias.



Fuente: Lind et al., 2012.

**Ojiva o polígono de frecuencias acumulativas:** Este gráfico es comúnmente empleado en el caso de variables continuas. Se construye seleccionando puntos en función de las marcas de clase y sus frecuencias correspondientes, y luego se conectan mediante una línea quebrada. Si tomamos los puntos medios de la parte superior de cada rectángulo en un histograma de frecuencias y los unimos, obtenemos el polígono de frecuencias

Gráfica 5. Ojiva o polígono de frecuencias acumuladas.

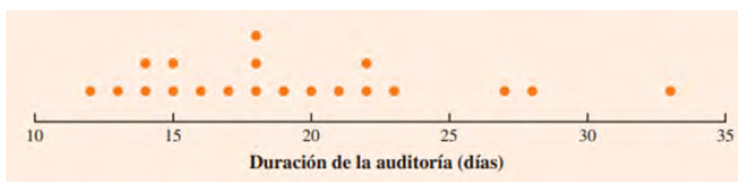


Fuente: Lind et al., 2012.

**Diagrama de puntos:** Una de las representaciones gráficas más simples para resumir datos es el diagrama de puntos. En este gráfico, el eje horizontal refleja el rango de los datos, y cada valor se representa mediante un punto ubicado en dicho eje. Los diagramas de puntos permiten observar con detalle los datos y son eficaces para comparar la distribución de datos de dos o más variables.



Gráfica 6. Diagrama de puntos.



Fuente: Anderson et al., 2012.

### 1.2.3 Conclusión

Comprender las variables continuas y discretas, así como las diferentes escalas de medición (nominal, ordinal, de intervalo y de razón), proporciona a los estudiantes la base necesaria para clasificar y categorizar datos de manera adecuada. Al reconocer cómo se pueden medir y cuantificar distintas características, los estudiantes aprenden a seleccionar las técnicas estadísticas apropiadas para analizar estas variables y a interpretar los resultados de manera precisa. Además, esto les permite distinguir entre diferentes niveles de información y precisión en los datos, lo que influye directamente en la calidad de sus análisis y conclusiones.

Los diversos tipos de gráficos, como el gráfico de pastel, el gráfico de barras, el diagrama de puntos, el diagrama de frecuencias, el polígono de frecuencias y el diagrama de Pareto, ofrecen a los estudiantes una gama de herramientas visuales para representar datos de manera clara y accesible. La elección adecuada de un tipo de gráfico depende de la naturaleza de los datos y los objetivos de la presentación. A través del estudio y la aplicación de estos gráficos, los estudiantes aprenden a resumir información compleja, identificar patrones y tendencias, y comunicar de manera efectiva los resultados de sus análisis.

### **1.2.4 Autoevaluación**

- 1.- ¿Cuáles son los conceptos de variables cualitativas y cuantitativas?
- 2.- ¿Cuál es la clasificación de las variables?
- 3.- ¿Cuántos tipos de gráficos conoce?

## 1.3. Distribución de frecuencias

### 1.3.1. Conceptos básicos

**Distribución de frecuencias:** Se trata de la manera en que se organiza un conjunto de datos en grupos mutuamente excluyentes, lo que significa que un dato pertenece a un grupo y no puede estar en otro simultáneamente. Esta distribución implica la organización de observaciones en diversos grupos, generalmente en orden ascendente o descendente. Es importante destacar que la distribución de frecuencias se aplica comúnmente en relación a una muestra estadística, aunque también es posible llevar a cabo esta clasificación en función de una población completa.

#### Intervalos de clase y límites de clase

Según Spiegel & Stephens, se puede entender lo siguiente acerca de los intervalos de clase y los límites de clase: El símbolo que representa una clase, como “60-62” en una tabla, se denomina intervalo de clase. Los números que se encuentran en los extremos, es decir, “60” y “62”, se denominan límites de clase. El número más pequeño, “60”, se llama límite inferior de clase, mientras que el número más grande, “62”, se llama límite superior de clase. Aunque los términos clase e intervalo de clase a menudo se usan de manera intercambiable, en realidad, el intervalo de clase es un símbolo que representa la clase.

Un intervalo de clase que no tiene especificado teóricamente ni el límite superior ni el límite inferior se conoce como intervalo de clase abierto. Por ejemplo, al considerar grupos de edades de personas, un intervalo como “65 años o más” se clasifica como un intervalo de clase abierto (Spiegel & Stephens, 2009, p. 38).

### **Fronteras de clase.**

Fronteras de clase: Si estamos registrando alturas redondeadas a la pulgada más cercana, el intervalo de clase de 60 a 62, teóricamente, incluye todas las mediciones desde 59.5000 hasta 62.5000 pulgadas. Para representar estos valores de manera más compacta, los llamamos 59.5 y 62.5, y se refieren como las fronteras reales del intervalo de clase. La frontera inferior de clase es el valor más bajo, que es 59.5, y la frontera superior de clase es el valor más alto, que es 62.5.

En la realidad, se calculan las fronteras de clase sumando el límite superior del intervalo de clase actual al límite inferior del siguiente intervalo de clase, y luego dividiendo el resultado entre 2 (Spiegel & Stephens, 2009).

### **Tamaño o amplitud de un intervalo de clase.**

Según Spiegel & Stephens (2009), “el tamaño o la amplitud de un intervalo de clase es la diferencia entre sus límites superior e inferior. También se le conoce como amplitud de clase, tamaño de clase o longitud de clase. Cuando todos los intervalos de clase en una distribución de frecuencia tienen la misma amplitud, se denota como “c”. En este caso, “c” es igual a la diferencia entre dos límites inferiores de clases sucesivas o entre dos límites superiores de clases sucesivas. Por ejemplo, en los datos del intervalo de clase es  $c = 62.5 - 59.5 = 65.5 - 62.5 = 3$ ”. (p. 38)

### **La marca de clase**

También se conoce como el punto medio de clase, se determina encontrando el punto central en el rango de un intervalo de clase. Este cálculo se realiza sumando los límites inferior y superior del intervalo y luego dividiendo el resultado entre 2. Por ejemplo, en el intervalo 60-62, la marca de clase se calcularía como  $(60 + 62) / 2 = 61$  (Spiegel & Stephens, 2009).

### 1.3.2. Tipos de distribuciones de frecuencia

Se dividen de la siguiente manera:

**Frecuencia absoluta:** Representa la cantidad de observaciones que caen en cada categoría o grupo. También puede entenderse como la cantidad de veces que se repite un evento.

**Frecuencia relativa:** Se obtiene al dividir la frecuencia absoluta entre el número total de datos. En otras palabras, es la proporción de ocurrencia de cada categoría en relación con el conjunto completo de datos.

**Frecuencia absoluta acumulada:** Se calcula sumando las frecuencias absolutas de un grupo o clase particular de la muestra (o población) con las clases previas.

**Frecuencia relativa acumulada:** Es el resultado de sumar las frecuencias relativas, de manera similar a cómo se realiza con la frecuencia absoluta acumulada.

### 1.3.3. Datos agrupados

Los datos agrupados son aquellos que, como su nombre indica, se toman una cantidad dada de datos y se organizan en categorías o grupos, ya sea en función de sus características cualitativas o cuantitativas, con el propósito de facilitar su análisis.

Por lo general, se recomienda agrupar datos cuando la población consta de alrededor de 20 o más elementos que comparten una característica y pueden ser categorizados (es decir, tienen valores repetidos). Esto permite un manejo más eficiente y un análisis más detallado de los datos. La organización se realiza mediante la creación de clases, donde cada clase representa una categoría en la que se agrupan los datos.

Existen dos formas principales de organizar o clasificar estos datos agrupados: en función de la frecuencia o en intervalos.

Los datos agrupados en función de la frecuencia se organizan en una tabla de frecuencia, lo que facilita la identificación de la cantidad de veces que se repite cada valor en un conjunto de datos. Esto simplifica la identificación de patrones y tendencias en los datos.

### **1.3.4. Propiedades de las frecuencias: absoluta, relativa y acumulada.**

Pueden resumirse de la siguiente manera:

- Las frecuencias absolutas siempre son números enteros.
- La suma de todas las frecuencias absolutas es igual al tamaño total de la muestra ( $n$ ).
- Las frecuencias relativas siempre son valores fraccionarios o decimales.
- La suma de todas las frecuencias relativas es igual a 1, lo que indica que representan la proporción total.
- El último valor en las frecuencias acumuladas absolutas debe ser igual al tamaño total de la muestra ( $n$ ).
- El último valor en las frecuencias acumuladas relativas también debe ser igual a 1, ya que representa la proporción total de la muestra (Martínez, 2012).

### **1.3.5. Conclusión**

El conocimiento de los distintos tipos de distribuciones de frecuencia y su cálculo asociado brinda a los estudiantes una comprensión más profunda de la estadística descriptiva y la for-

ma en que se pueden resumir y representar datos. La capacidad para organizar datos en grupos excluyentes, calcular frecuencias absolutas y relativas, y generar distribuciones acumuladas permite a los estudiantes desarrollar habilidades analíticas y críticas esenciales. Además, este conocimiento sienta las bases para futuros estudios estadísticos más avanzados y la comprensión de conceptos más complejos, como inferencia estadística y análisis multivariado.

### **1.3.6. Autoevaluación**

1.- ¿Cuál es la diferencia entre la frecuencia absoluta y la frecuencia relativa?

## **1.4. Distribución de frecuencias**

### **1.4.1. Datos no agrupados**

Los datos no agrupados son un conjunto de datos sin clasificar, donde cada valor se presenta individualmente en una tabla. Esto generalmente implica tener menos de 30 elementos ( $n < 30$ ) y una baja repetición de valores.

El tratamiento de estos datos no agrupados es sencillo. Se recopilan los datos de la población de estudio y se presentan en una tabla de datos sin la necesidad de crear categorías o clases. Los datos se mantienen con su identidad individual incluso después de haber sido organizados en una tabla de frecuencias para su análisis.

### **1.4.2. Conclusión**

Los datos no agrupados son aquellos que se presentan sin haber sido categorizados previamente, mostrándose en una tabla donde cada valor se presenta individualmente. Este tipo de conjunto suele constar de menos de 30 elementos ( $n < 30$ ) y con escasa o nula repetición.

El tratamiento de estos datos no agrupados es sencillo. Consiste en recopilar los datos de la población estudiada y organizarlos en una tabla de datos, analizándolos sin la necesidad de agruparlos en clases. Estos datos se distribuyen en una tabla de frecuencias, donde cada dato retiene su identidad única después de haber sido elaborada la distribución de frecuencias.

Con este conocimiento, los estudiantes pueden adquirir habilidades prácticas para recopilar, organizar y analizar datos de

manera eficiente, lo que les será útil en diversas áreas académicas y profesionales donde se requiera el análisis de información numérica.

### **1.4.3. Autoevaluación**

- ¿Qué son los datos no agrupados?
- ¿Ponga un ejemplo de datos no agrupados?



## **Capítulo 2**

Presentación de datos estadísticos e indicadores de tendencia central y variabilidad

## 2.1. Medidas de tendencia central

### 2.1.1. Conceptos básicos

#### Medía aritmética.

La medida de posición o promedio más comúnmente conocida, ampliamente utilizada y fácilmente comprensible por la mayoría es el promedio aritmético. Debido a su alta estabilidad, es la opción preferida en el muestreo, y sus fórmulas permiten un tratamiento algebraico.

El promedio aritmético se define como la “suma de todos los valores observados divididos por el número total de observaciones”. Esta definición se aplica únicamente a datos no agrupados, que también se conocen como datos originales.

#### Media aritmética simple.

Se emplean datos en su forma no agrupada o datos originales, donde:

##### 2.1.1.1 Media aritmética de la población

$$u = \bar{x} = \frac{\sum x}{N}; \text{ siendo } u = \frac{x_1 + x_2 + x_3 \dots x_n}{N} \quad \begin{array}{l} \text{Suma de los valores de todas las observaciones} \\ \text{Número de elementos de la población} \end{array}$$

##### 2.1.1.2 Media aritmética de la muestra

$$\bar{x} = \frac{\sum x}{n} \quad \text{donde } \bar{x} = \frac{x_1 + x_2 + x_3 \dots x_n}{n} \quad \begin{array}{l} \text{Suma de los valores de todas las observaciones} \\ \text{Número de elementos de la muestra} \end{array}$$

### *2.1.1.3. Media aritmética ponderada*

Se utiliza en datos que han sido agrupados, es decir, datos que se han organizado en una tabla de frecuencias, donde las frecuencias absolutas actúan como sus pesos o ponderaciones.

### *2.1.1.4. Mediana*

La mediana de un conjunto de datos es aquel valor que se sitúa exactamente en el centro cuando los valores se organizan en orden ascendente o descendente.

### *2.1.1.5. Moda*

La moda de un conjunto de datos es el valor que se presenta con mayor frecuencia, es decir, aquel que se repite más veces

### *2.1.1.6. Media geométrica*

La media geométrica de un conjunto de  $n$  números positivos se calcula como la raíz  $n$ -ésima del producto de estos  $n$  valores. La fórmula de la media geométrica se expresa de la siguiente manera:

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

## **2.1.2. Conclusión**

La media representa el promedio de todos los valores en el conjunto de datos y ofrece una visión general de su magnitud.

La mediana, por otro lado, es el valor central cuando los datos se ordenan de manera ascendente o descendente y es especialmente útil cuando hay valores atípicos que podrían distorsionar la media.

La moda se refiere al valor que aparece con mayor frecuencia en el conjunto de datos, lo que puede revelar picos o concentraciones en la distribución.

Comprender y aplicar estas medidas es fundamental para los estudiantes, ya que les capacita para resumir grandes conjuntos de datos en valores representativos y comprensibles. Esta habilidad es esencial en campos como la investigación científica, economía, sociología e ingeniería, donde se requiere analizar y comunicar información numérica de manera efectiva. Además, el conocimiento de estas medidas les permite a los estudiantes evaluar la variabilidad y estructura de los datos, lo que facilita la toma de decisiones basadas en evidencia y la resolución de problemas de manera informada.

### **2.1.3. Autoevaluación**

- 1.- ¿Qué es la media aritmética, la mediana y la moda?
- 2.- ¿Cuál es la diferencia entre la media aritmética y la media geométrica?

## 2.2. Medidas de dispersión: varianza, desviación típica y coeficiente de variación

### 2.2.1. Conceptos básicos

Se analiza varias medidas de dispersión. El rango se obtiene considerando los valores más altos y más bajos en el conjunto de datos, lo que significa que únicamente se tienen en cuenta dos valores extremos. Por contraste, la desviación media, la varianza y la desviación estándar se calculan a partir de las diferencias entre los valores individuales y la media aritmética

#### 2.2.1.1. Rango

La métrica más elemental para evaluar la dispersión es el rango, que se define como la diferencia entre el valor máximo y el valor mínimo presentes en un conjunto de datos. Esto se puede expresar de manera matemática de la siguiente manera:

Rango = Valor máximo–valor mínimo

#### 2.2.1.2. Desviación media

La desviación media es el promedio de los valores absolutos de las desviaciones en relación con la media aritmética.

$$\text{Desviación Media} = \mathbf{DM} = \frac{\sum |x - \bar{x}|}{n}$$

X es el valor de cada observación

es la media aritmética de los valores

N es el número de observaciones en la muestra

|| indica el valor absoluto

### 2.2.1.3. Desviación estándar o típica

La raíz cuadrada de la varianza de la población es la desviación estándar de la población.

$$\text{Desviación Estándar De La Población } \sigma = \sqrt{\frac{\sum(x-u)^2}{N}}$$

### 2.2.1.4. Varianza

Media aritmética de las desviaciones de la media elevadas al cuadrado

#### 2.2.1.4.1. Varianza de la población

La varianza poblacional se calcula de la siguiente manera:

$$\text{Varianza de la Población } \sigma^2 = \frac{\sum(x-u)^2}{N}$$

En esta fórmula:

$\sigma^2$  es la varianza de la población (es la letra minúscula griega sigma); se lee sigma al cuadrado.

X es el valor de una observación de la población.

u es la media aritmética de la población.

N es el número de observaciones de la población

#### 2.2.1.4.2. Varianza de la muestra

La fórmula para calcular la varianza de una muestra es la siguiente:

$$\text{Varianza Muestral } S^2 = \frac{\sum(x - \bar{x})^2}{n - 1}$$

donde:

$s^2$  representa la varianza muestral.

$X$  denota el valor de cada observación de la muestra.

$\bar{x}$  representa la media de la muestra.

$n$  es el número de observaciones en la muestra

#### 2.2.1.5. Coeficiente de variación

Se refiere al resultado de dividir la desviación estándar entre la media aritmética.

$$cv = \frac{s}{\bar{x}} \text{ (relativo)}$$

$$cv = \frac{s}{\bar{x}} * 100 \text{ (porcentual)}$$

### 2.2.2. Conclusión

Las medidas de dispersión, como el rango, la desviación estándar y el coeficiente de variación, ayudan a cuantificar la amplitud de los valores individuales con respecto a la medida central, lo que permite una comprensión más completa de la distribución de los datos.

El aporte de las medidas de dispersión para los estudiantes es significativo. Al comprender y aplicar estas medidas, los estudiantes pueden:

- Evaluar la consistencia de datos
- Comparar conjunto de datos
- Detectar valores atípicos
- Tomar decisiones informadas
- Evaluar la precisión de las medidas centrales

El conocimiento y la aplicación de las medidas de dispersión son esenciales para que los estudiantes desarrollen habilidades analíticas sólidas y una comprensión profunda de cómo interpretar y analizar conjuntos de datos de manera efectiva. Estas medidas no solo son relevantes en contextos académicos, sino también en la vida cotidiana y en una amplia gama de disciplinas y profesiones.

### **2.2.3. Autoevaluación**

- 1.- ¿Qué es un rango?
- 2.- ¿Qué es el coeficiente de variación?
3. ¿Qué es varianza, desviación estándar y desviación media?

## 2.3. Medidas de posición: cuartiles, deciles y percentiles

### 2.3.1. Conceptos básicos

#### 2.3.1.1. Cuartiles

Los cuartiles, representados por la letra Q, son tres números que parten una serie de datos en cuatro partes de igual tamaño.

Datos Pares

$$Q_k = k * n / 4$$

Datos Impares

$$Q_k = k * (n+1) / 4$$

#### 2.3.1.2. Deciles

Los deciles se representan con la letra D y son nueve valores que fragmentan una serie de datos en 10 partes de igual tamaño

Datos Pares

$$D_k = k * n / 10$$

Datos Impares

$$D_k = k * (n+1) / 10$$

#### 2.3.1.3. Percentiles

Los percentiles se representan con la letra P y son 99 valores que segmentan una serie de datos en 100 partes de igual tamaño.

Datos Pares

$$P_k = k * n / 100$$

Datos Impares

$$P_k = k * (n+1) / 100$$

### 2.3.2. Conclusión

Los cuartiles, deciles, percentiles son medidas de localización, su función es informar del valor de la variable que ocupará la posición (en tanto por cien) que nos interese respecto de todo

el conjunto de variables.

Estas medidas ofrecen una perspectiva crucial sobre cómo se distribuyen los datos y dónde se ubican los valores en términos de su orden. Al proporcionar información detallada sobre la posición de un valor en relación con el conjunto completo de datos, los cuantiles, deciles y percentiles son herramientas esenciales para analizar y comprender la variabilidad y la estructura de las distribuciones.

El estudio de cuartiles, deciles y percentiles brinda a los estudiantes habilidades analíticas sólidas y una comprensión profunda de cómo los datos están distribuidos y cómo interpretar esas distribuciones. Estos conceptos no solo son relevantes en contextos académicos, sino que también ofrecen a los estudiantes herramientas valiosas de análisis en su futura carrera profesional.

### **2.3.3. Autoevaluación**

- 1.- ¿Cuál es la diferencia entre cuartil y decil?
- 2.- Explique en un ejemplo como se calcula el cuartil

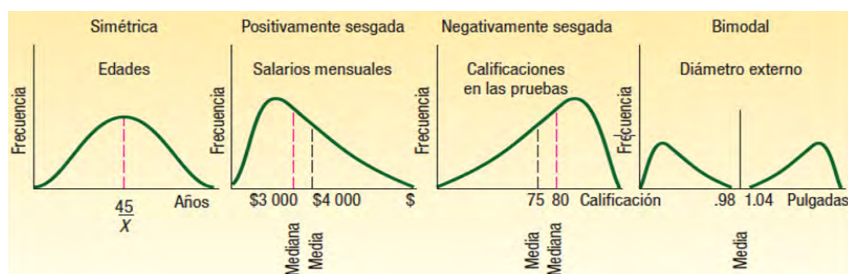
## 2.4 Momento, sesgo y curtosis

### 2.4.1 Conceptos básicos

#### Sesgo

Las representaciones gráficas de los datos en un conjunto pueden tener diferentes formas, algunas simétricas y otras asimétricas. En la gráfica 7, se observan distintos tipos de curvas con diferentes sesgos, lo que significa que los valores se concentran en un extremo del eje horizontal de la escala de medición. Esto indica que los valores no están uniformemente distribuidos. La curva A muestra un sesgo hacia la derecha, ya que gradualmente disminuye hacia el extremo derecho de la escala. Por otro lado, la curva B muestra un sesgo hacia la izquierda, ya que disminuye gradualmente hacia el extremo inferior de la escala.

Gráfica 7. Tipos de curvas con diferentes sesgos.



Fuente: Lind et al., 2012, p. 120.

Gráfica 8. Comparación de dos curvas sesgadas y dos curvas con la misma posición central pero diferente curtosis.



Fuente: Levin & Rubin, 2004, p. 59.

## Curtosis

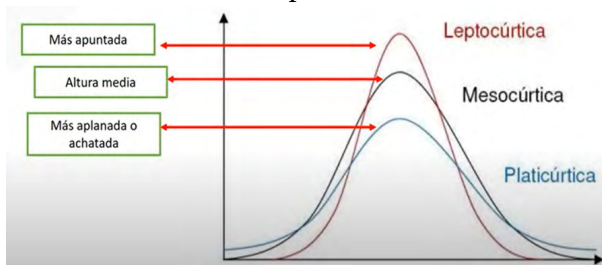
La curtosis de una distribución se refiere a cuán apuntada o achatada es la curva. En la gráfica 9 se ilustran diferentes tipos de curtosis. Esta medida se utiliza para evaluar la forma de la distribución en términos de su apuntamiento y puede categorizarse en:

Leptocúrtica, si  $Ap > 0$ ; es decir, es más apuntada que la normal. Los valores que toma la variable están muy concentrados en torno a su media y hay pocos valores extremos.

Mesocúrtica, si  $Ap = 0$ ; es decir, es tan apuntada como normal.

Platicúrtica, si  $Ap < 0$ ; es decir, es menos apuntada que la normal. Hay muchos valores extremos, las colas de la variable son muy pesadas (Levin & Rubin, 2004, p. 59).

Gráfica 9. Tipos de curtosis.



Fuente: <https://acortar.link/iuFOz4>

## Momentos estadísticos

Los momentos estadísticos son expresiones matemáticas utilizadas como parámetros en estadísticas, y varios de ellos son especialmente relevantes en el análisis de distribuciones de frecuencia, especialmente en relación con el sesgo y la curtosis. Estos momentos permiten una generalización de la teoría relacionada con los parámetros estadísticos y están vinculados a un amplio conjunto de estos.

El momento central o centrado de orden  $k$  en una distribución de datos estadísticos  $x_1, x_2, \dots, x_n$  se define como:

$$\mu_k = \frac{\sum_{i=1}^n (x_i - \bar{x})^k}{n}$$

### **2.4.2 Conclusión**

El estudio de momentos, sesgo y curtosis en estadística es de gran importancia debido a su papel fundamental en el análisis y la interpretación de conjuntos de datos. Estos conceptos proporcionan información esencial sobre la forma, la simetría y la concentración de los valores en una distribución, lo que permite a los analistas y profesionales comprender mejor la variabilidad de los datos y tomar decisiones informadas en diversos contextos.

### **2.4.3. Autoevaluación**

- 1.- ¿Qué significa estar sesgada a la derecha?
- 2.- ¿Qué es la curtosis?
- 3.- ¿Qué es momento?



## Capítulo 3

### Probabilidades, regresión y correlación

## 3.1 Teoría elemental de probabilidades

### 3.1.1 Conceptos básicos

**Probabilidad.** – Es un número que se encuentra en el rango de cero a uno, ambos valores incluidos, que indica la medida relativa de la posibilidad de que un evento ocurra.

**Experimento.** – Es un proceso que conduce a una de varias observaciones posibles, pero solo a una de ellas.

En relación con la probabilidad, un experimento implica la existencia de al menos dos resultados posibles, y no se tiene certeza sobre cuál de ellos se materializará.

**Resultado.** – Un resultado es un evento específico que ocurre como consecuencia de un experimento en particular. Por ejemplo, lanzar una moneda es un experimento en el que puedes observar el resultado, pero no puedes predecir con certeza si caerá cara o cruz.

**Evento o suceso.** – Se refiere a un grupo de uno o más resultados que pueden ocurrir como parte de un experimento. Este conjunto engloba todas las posibles consecuencias de un experimento aleatorio y constituye un subconjunto del espacio muestral, que es la totalidad de los resultados posibles de dicho experimento.

**Espacio muestral.** – El espacio muestral de un experimento se define como el conjunto que contiene todos los resultados potenciales que pueden ocurrir en dicho experimento.

$$S = \{\text{cara, cruz}\}$$

### 3.1.2 Enfoques de probabilidades

Es útil examinar dos enfoques para asignar probabilidades: los enfoques objetivo y subjetivo. La probabilidad objetiva se divide en dos categorías: a) probabilidad clásica y b) probabilidad empírica.

#### 3.1.2.1 Enfoque objetivo

##### Probabilidad clásica

La probabilidad clásica se basa en la premisa de que todos los resultados de un experimento tienen la misma probabilidad de ocurrir. Según esta perspectiva, para calcular la probabilidad de un evento en curso, se divide el número de resultados favorables entre el número total de resultados posibles.

##### Probabilidad Clásica

$$\text{Probabilidad de un evento} = \frac{\text{Número de resultados favorables}}{\text{Número total de posibles resultados}}$$

**Mutuamente excluyentes.** – Eventos mutuamente excluyentes se caracterizan por el hecho de que la realización de uno de ellos excluye la posibilidad de que los otros ocurran simultáneamente.

**Colectivamente exhaustivo.** – Por lo menos uno de los eventos debe ocurrir cuando se lleva a cabo un experimento.

##### Probabilidad empírica

La probabilidad de que un evento se produzca corresponde a una fracción de eventos similares que han tenido lugar en el pasado.

$$\text{Probabilidad empírica} = \frac{\text{Número de veces que el evento ocurre}}{\text{Número total de observaciones}}$$

**Ley de los grandes números.**–Establece que, en un gran número de experimentos, la probabilidad observada de un evento se acercará a su probabilidad real.

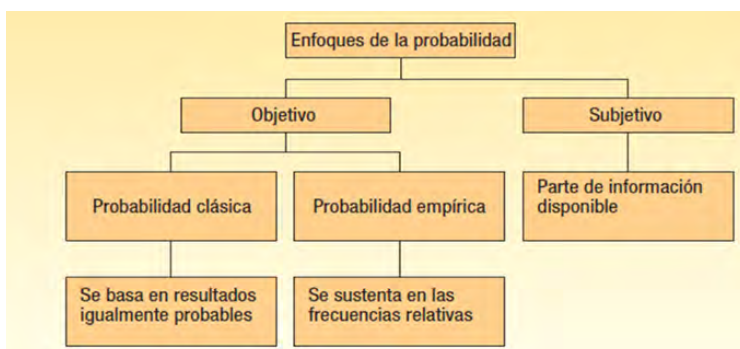
### 3.1.2.2 Probabilidad subjetiva

La probabilidad de un evento específico que un individuo determina utilizando cualquier información que tenga a su disposición.

Ejemplos:

- Determinar la probabilidad de casarse antes de cumplir 30 años.
- Estimar la probabilidad de que el déficit presupuestario de Estados Unidos disminuya a la mitad en los próximos 10 años.

Gráfica 10. Enfoques de la probabilidad



Fuente: Lind et al., 2006, p. 151.

### 3.1.3 Reglas para calcular probabilidades

#### 3.1.3.1 Reglas de la adición

Hay dos reglas relacionadas con la adición: la regla especial de la adición y la regla general de la adición.

##### Regla especial de la adición

La regla especial de la adición se aplica cuando tenemos dos eventos, A y B, que son mutuamente excluyentes, lo que significa que no pueden ocurrir al mismo tiempo. Según esta regla, la probabilidad de que ocurra uno de estos eventos es igual a la suma de sus probabilidades individuales. Esto se representa mediante la fórmula:  $P(A \text{ o } B) = P(A) + P(B)$ .

En el caso de tres eventos mutuamente excluyentes, A, B y C, la regla se extiende de la siguiente manera:  $P(A \text{ o } B \text{ o } C) = P(A) + P(B) + P(C)$ .

##### Regla del complemento

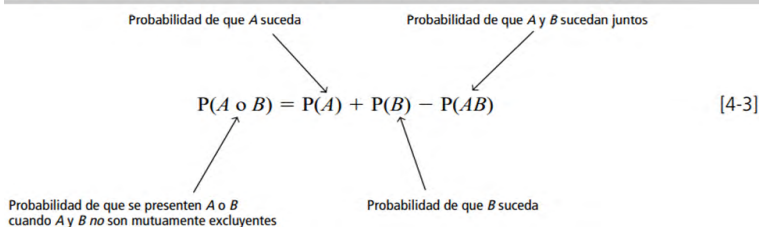
La Regla del Complemento es útil porque a veces es más fácil calcular la probabilidad de que un evento no ocurra y luego restar ese valor a 1 para encontrar la probabilidad de que ocurra el evento.

Se expresa como:

**Regla del complemento**  $P(A) = 1 - P(\neg A)$

##### **Regla de adición para eventos que no son mutuamente excluyentes**

Cuando dos eventos no son mutuamente excluyentes, existe la posibilidad de que ambos ocurran simultáneamente.

**Regla de adición para eventos que no son mutuamente excluyentes**

Fuente: Levin & Rubin 2004, p. 140.

Ejemplo.

Los empleados de una cierta compañía han elegido a cinco de ellos para que los representen en el consejo administrativo y de personal sobre productividad. Los perfiles de los cinco elegidos son:

| TIPO   | EDAD |
|--------|------|
| Hombre | 30   |
| Hombre | 32   |
| Mujer  | 45   |
| Mujer  | 20   |
| Hombre | 40   |

Este grupo opta por seleccionar a una persona que actúe como representante. Nuestra pregunta es, ¿cuál es la probabilidad de que el representante sea mujer o cuya edad esté por arriba de 35 años?

$P(\text{mujer o mayor de 35}) = P(\text{mujer}) + P(\text{mayor de 35}) - P(\text{mujer y mayor de 35})$

$$= 2/5 + 2/5 - 1/5$$

$$= 3/5$$

### 3.1.3.2 Reglas de la multiplicación

Para calcular la probabilidad de que dos eventos ocurran al mismo tiempo, se utiliza la regla de la multiplicación. Existen dos tipos de reglas de multiplicación: la regla especial y la regla general.

#### Regla especial de la multiplicación

La regla especial de la multiplicación requiere que dos eventos, A y B, sean independientes, y lo son si el hecho de que uno ocurra no altera la probabilidad de que el otro suceda.

**Independencia.** – Cuando un evento ocurre, no influye en la probabilidad de que otro evento se produzca.

Regla especial de la multiplicación  $P(A \text{ y } B) = P(A) P(B)$

En el caso de tres eventos independientes, A, B y C, la regla especial de la multiplicación que se utiliza para determinar la probabilidad de que los tres eventos ocurran es:

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

#### Regla general de la multiplicación

Si dos eventos no son independientes, implica que están vinculados o tienen relación de dependencia entre ellos.

La regla general de la multiplicación sostiene que cuando se trata de dos eventos, A y B, la probabilidad conjunta de que ambos eventos ocurran se halla multiplicando la probabilidad de que suceda el evento A por la probabilidad condicional de que ocurra el evento B, considerando que el evento A ya ha sucedido. En términos simbólicos, esto se expresa como:

Regla general de la multiplicación  $P(A \text{ y } B) = P(A)P(B|A)$

**Probabilidad condicional.** – Se refiere a la probabilidad de que

un evento específico ocurra después de que otro evento ya haya tenido lugar (Lind et al., 2012, pp. 160-161).

### **3.1.4 Conclusión**

La aplicación de los métodos de cálculo de la probabilidad posee una relevancia central al permitir la estimación y predicción de eventos futuros. La precisión de los resultados obtenidos se incrementa en proporción directa a la disponibilidad de datos para calcular la probabilidad de eventos específicos. Dado el nivel de complejidad inherente en los sistemas donde se emplea la teoría de la probabilidad, se requiere la utilización de modelos estadísticos e informáticos altamente avanzados.

La importancia fundamental de la teoría de la probabilidad se manifiesta en su esfuerzo por representar a través de conceptos matemáticos el desarrollo futuro de los eventos.

La probabilidad emerge como una herramienta de suma importancia en la planificación estratégica de movimientos que inciden en la esfera social, económica y laboral de la comunidad en su totalidad. La implementación de las reglas de probabilidad, tanto en términos de suma como de multiplicación de probabilidades, constituye un elemento fundamental para respaldar la obtención de resultados precisos y la toma de decisiones fundamentadas. La utilidad de las probabilidades radica en su capacidad para proporcionar una base sólida y cuantitativa para abordar la incertidumbre y facilitar la toma de decisiones informadas en diversos contextos.

### **3.1.5 Autoevaluación**

- 1.- ¿Qué es un evento?
- 2.- ¿Qué es el espacio muestral?
- 3.- ¿Qué es probabilidad condicional?

## 3.2 Principios y técnicas de conteo: permutaciones y combinaciones

### 3.2.1. Permutaciones

Permutación, cualquier distribución de  $r$  objetos seleccionados de un solo grupo de  $n$  posibles objetos.

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

$n$  = población

$r$  = muestra

$!$  = factorial

#### 3.2.1.1. Ejercicio en clase

El gerente de la tienda departamental quiere exhibir 6 marcas de un producto sobre el anaquel de un pasillo ¿de cuántas maneras puede ordenar las marcas?

$$n = 6$$

$$r = 6$$

$$6P6 = 6! / (6-6)! = 720$$

### 3.2.2 Combinaciones

El número de maneras en las que se pueden seleccionar  $r$  objetos de un conjunto con  $n$  elementos, sin importar el orden de selección, se denomina número de combinaciones de  $n$  objetos tomando  $r$  a la vez.

Fórmula de las combinaciones  $nCr = \frac{n!}{r!(n-r)!}$

n= población

r= muestra

!= factorial

### 3.2.3 Conclusión

Las permutaciones y combinaciones ofrecen un enfoque estructurado y preciso para resolver problemas de conteo y ordenamiento en una variedad de contextos. Estos conceptos son esenciales para la toma de decisiones fundamentadas, la optimización de procesos y la resolución eficiente de problemas en diversas disciplinas.

Las permutaciones y combinaciones ayudan a la resolución precisa de problemas de conteo, optimización de procesos, diseño y análisis de experimentos, desarrollo de algoritmos y programación.

### 3.2.4 Autoevaluación

¿Qué son la permutaciones y combinaciones?

### 3.3 Distribuciones probabilísticas

#### 3.3.1 Distribución de probabilidad

Lista de todos los resultados de un experimento y la probabilidad asociada a cada uno de ellos.

#### 3.3.2 Características de una distribución de probabilidad

- La probabilidad de un resultado específico está en el rango de 0 a 1, incluyendo ambos extremos.
- Los resultados son eventos mutuamente excluyentes
- La lista es exhaustiva. lo que significa que la suma de las probabilidades de todos los eventos equivale a 1.

##### Ejemplo

Suponga que le interesa el número de caras que aparecen en tres lanzamientos de una moneda. Tal es el experimento. Los posibles resultados son: cero caras, una cara, dos caras y tres caras. ¿Cuál es la distribución de probabilidad del número de caras?

##### Solución.

Hay ocho posibles resultados. En el primer lanzamiento puede aparecer una cara, una cruz en el segundo y otra cruz en el tercero. O puede obtener cruz, cruz y cara, en ese orden. Para obtener el conteo de resultados, aplique la fórmula de la multiplicación:  $(2)(2)(2)$ , es decir, 8 posibles resultados. Estos resultados se listan en seguida.

| Lanzamiento de la moneda |         |         |         |                 |
|--------------------------|---------|---------|---------|-----------------|
| Resultado posible        | Primero | Segundo | Tercero | Número de caras |
| 1                        | T       | T       | T       | 0               |
| 2                        | T       | T       | H       | 1               |
| 3                        | T       | H       | T       | 1               |
| 4                        | T       | H       | H       | 2               |
| 5                        | H       | T       | T       | 1               |
| 6                        | H       | T       | H       | 2               |
| 7                        | H       | H       | T       | 2               |
| 8                        | H       | H       | H       | 3               |

Observe que el resultado cero caras ocurre sólo una vez; una cara ocurre tres veces; dos caras, tres veces, y el resultado tres caras ocurre una sola vez. Es decir, cero caras se presentaron una de ocho veces. Por consiguiente, la probabilidad de cero caras es de un octavo; la probabilidad de una cara es de tres octavos, etc. La suma de las probabilidades de todos los eventos que pueden ocurrir equivale a 1.

Distribución de probabilidad de los eventos relativos a cero, una, dos y tres caras en tres lanzamientos de una moneda

| Número de caras X | Probabilidad de resultado P(x) |
|-------------------|--------------------------------|
| 0                 | $1/8 = 0,125$                  |
| 1                 | $3/8 = 0,375$                  |
| 2                 | $3/8 = 0,375$                  |
| 3                 | $1/8 = 0,125$                  |
| Total             | $8/8 = 1$                      |

Fuente: Lind et al., 2012, p. 188.

### 3.3.3 Conclusión

La distribución de probabilidad es una herramienta fundamental en la teoría de la probabilidad y la estadística, que permite modelar y entender cómo se distribuyen las probabilidades de diferentes resultados en un experimento. Esta distribución refleja la relación entre los posibles eventos y las probabilidades asociadas a cada uno de ellos, lo que resulta esencial en la toma de decisiones informadas y en la comprensión de la incertidumbre en una variedad de campos.

Las características clave de una distribución de probabilidad destacan su naturaleza y utilidad:

- **Rango de Probabilidad Limitado:** La probabilidad de cada resultado se sitúa en el intervalo entre 0 y 1, incluyendo ambos extremos. Esta característica refleja la noción de que los eventos imposibles tienen una probabilidad de 0, mientras que los eventos seguros tienen una probabilidad de 1.
- **Eventos Mutuamente Excluyentes:** En una distribución de probabilidad, los resultados son eventos mutuamente excluyentes, lo que significa que solo uno de los resultados posibles puede ocurrir. Esta propiedad es esencial para que la asignación de probabilidades sea consistente y coherente.
- **Lista Exhaustiva:** La lista de resultados abarca todas las posibilidades, asegurando que cada resultado potencial esté contemplado. Esto significa que la suma de las probabilidades de todos los eventos en la distribución es igual a 1.

- Estas características fundamentales permiten a los analistas y profesionales comprender cómo se distribuyen las probabilidades en diferentes situaciones y cómo estimar la probabilidad de resultados específicos. Las distribuciones de probabilidad se aplican en áreas como la estadística, la economía, la ingeniería y la investigación científica.

### **3.3.4 Autoevaluación**

¿Qué es una distribución de probabilidad?

¿Cuáles son las características de una distribución de probabilidad?

## 3.4 Distribución binomial y de Poisson

### 3.4.1 Conceptos básicos

**Variable aleatoria.** – Es una descripción cualitativa de los posibles resultados en un conjunto de muestras. Cada número o dato representa el resultado de un experimento y, por lo tanto, puede adquirir diferentes valores.

**Variable aleatoria discreta.** Es aquella cuyos valores son contables y pueden ser dispuestos en una secuencia, similar a los números enteros positivos. Esta variable solo puede tomar un conjunto finito de valores.

**Variable aleatoria continua.** Se presenta cuando puede adoptar cualquier valor dentro de un intervalo o unión de intervalos. Un ejemplo sería cualquier medida de longitud, ancho, o tiempo de realización de una tarea, donde estas variables pueden tomar valores fraccionarios.

### 3.4.2 Distribución binomial

Algunos la llaman el “método exacto” y se refiere a una distribución de una variable aleatoria discreta

Se puede decir que cuando realizamos múltiples ensayos, manteniendo constante la probabilidad de éxito en un solo ensayo como “ $p$ ” y la probabilidad de fracaso como “ $q$ ”, entonces la probabilidad “ $P$ ” de que se obtenga  $X$  éxitos en  $n$  ensayos, es el término del desarrollo binomial de  $(q+p)^n$

La fórmula general para cada término es:

$$p_{(x)} = \binom{n}{x} p^x q^{n-x}$$

n = Número de ensayos

X = Número de éxitos

p = probabilidad de éxito en cada ensayo

q = probabilidad de fracaso en cada ensayo

### *Ejercicio en clase*

En una fábrica de bombillas el 5% sale defectuosas. Determinar la probabilidad de que en una muestra de 12 se encuentren 2 bombillas defectuosas.

$$p_{(x)} = \binom{n}{x} p^x q^{n-x}$$

n = Número de ensayos

X = Número de éxitos

p = probabilidad de éxito en cada ensayo

q = probabilidad de fracaso en cada ensayo

n= 12

x= 2

p=0,05

q= 1-p  $\rightarrow$  1-0,5= 0.95

$$= \binom{12}{2} 0,05^2 * 0,95^{12-2} = \binom{12!}{2!(12-2)!} 0,05^2 * 0,95^{12-2}$$

$$= \left[ \frac{12 \times 11 \times 10}{2 \times 1} \right] 0,0025 \times 0,598736 = 132/2 * 0.0014968 = 0.0987888 \times 100$$

$$= 9.88\%$$

Interpretación: La probabilidad de que en una muestra de 12 se encuentren 2 bombillas defectuosas es del 9,88%

### 3.4.3 Distribución de Poisson

Esta distribución de probabilidad se refiere a una variable aleatoria discreta que nos brinda la probabilidad de que un evento específico ocurra cierto número de veces, denotado como “x”, en un intervalo predefinido de tiempo, longitud, área, u otra medida similar.

$\lambda$  = número medio de veces que ocurre nuestro suceso en el intervalo de tiempo, longitud, área, etc.

$$P_{(x)} = \frac{\lambda^x e^{-\lambda}}{x!}$$

$e = 2,71828$  (base de los logaritmos neperianos)

$\lambda = np$

$X$  = número de casos favorables

$P_{(x)}$  = Probabilidad que se va a calcular para un valor dado de  $X$

#### *Ejercicio en clase*

Una empresa de telecomunicaciones recibe llamadas a razón de 4 llamadas por minuto. Calcular la probabilidad de:

- Recibir 2 llamadas en 1 minuto
- No recibir ninguna llamada en un minuto
- Recibir menos de 3 llamadas en un minuto
- Recibir más de 3 llamadas en un minuto

$e$  = base de logaritmos neperianos = 2,71828

$\lambda = 4$  llamadas/minuto

$$X = 2$$

$$\begin{aligned} \text{a) } P(x=2) &= \frac{4^2 \times 2,71828^{-4}}{2!} = \frac{16 \times 0,0183156}{2} = 0,1465 = 14,65 \% \\ \text{b) } P(x=0) &= \frac{4^0 \times 2,71828^{-4}}{0!} = \frac{1 \times 0,0183156}{1} = 0,0183 = 1,83\% \end{aligned}$$

$$\text{c) } P(x=1) = \frac{4^1 \times 2,71828^{-4}}{1!} = \frac{4 \times 0,0183156}{1} = 0,0733 = 7,33\%$$

$$P(x < 3) = P(x=0) + P(x=1) + P(x=2) \rightarrow 0,0183 + 0,0733 + 0,1465 = 0,2381 = 23,81$$

$$\text{d) } P(x=3) = \frac{4^3 \times 2,71828^{-4}}{3!} \rightarrow \frac{64 \times 0,0183156}{6} \rightarrow 0,1954 \rightarrow 19,54 \%$$

$$P(x \leq 3) = P(x=0) + P(x=1) + P(x=2) + P(x=3) = 0,0183 + 0,0733 + 0,1465 + 0,1954 \rightarrow 0,4335 \rightarrow 43,35 \%$$

$$P(x > 3) = 1 - P(x \leq 3) = 1 - 0,4335 \rightarrow 0,5665 \rightarrow 56,65 \%$$

### 3.4.4 Conclusión

La distribución binomial y la distribución de Poisson son herramientas esenciales en la teoría de probabilidades y la estadística, que permiten modelar y analizar eventos discretos con gran precisión. Estas distribuciones se utilizan para estimar la probabilidad de ocurrencia de eventos en diferentes contextos y proporcionan métodos exactos para abordar problemas que involucran conteo y frecuencia de eventos en un intervalo determinado.

La distribución binomial es un método exacto para modelar la probabilidad de obtener un número específico de éxitos en una serie de ensayos independientes y mutuamente excluyentes. Esta distribución es especialmente útil cuando se trata de eventos que tienen dos resultados posibles, como éxito o fracaso. Es valiosa

en situaciones como pruebas de hipótesis, muestreo y control de calidad, proporcionando una base matemática sólida para tomar decisiones informadas.

La distribución de Poisson es un método exacto para estimar la probabilidad de ocurrencia de un número particular de eventos en un intervalo de tiempo, longitud, área u otra unidad. Esta distribución es especialmente relevante cuando se trata de eventos raros o poco frecuentes.

### **3.4.5 Autoevaluación**

¿Qué es una distribución binomial?

¿Cuál es la característica de la distribución de Poisson?



## **Capítulo 4**

### Probabilidades, regresión y correlación

## 4.1. Distribución normal

La distribución normal es ampliamente utilizada para modelar variables continuas en una variedad de contextos, como la altura y el peso de las personas, las puntuaciones en exámenes, mediciones científicas y la precipitación pluvial, entre otros.

Esta distribución es fundamental en estadística inferencial y proporciona información sobre la probabilidad de los resultados obtenidos a partir de un proceso de muestreo.

La distribución normal se representa visualmente mediante una curva en forma de campana.

Gráfica 11. Curva con forma de campana de la distribución normal.



Función De Densidad De Probabilidad Normal

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

Donde:

$\mu$  = media

$\sigma$  = desviación estándar

$\pi$  = 3.14159

$e$  = 2,71828

### 4.1.2. Características de la distribución normal

1. Las distribuciones normales se caracteriza y distingue a través de dos parámetros: la media representada por  $\mu$  y la desviación estándar simbolizada por  $\sigma$ .

2. El punto más alto en una curva normal se sitúa en la media, que es también igual a la mediana y la moda de la distribución.

3. La media de una distribución normal puede tomar diversos valores numéricos, ya sea negativos, nulos o positivos. Aquí se presentan tres distribuciones normales que comparten la misma desviación estándar, pero tienen medias distintas (-10, 0 y 20).

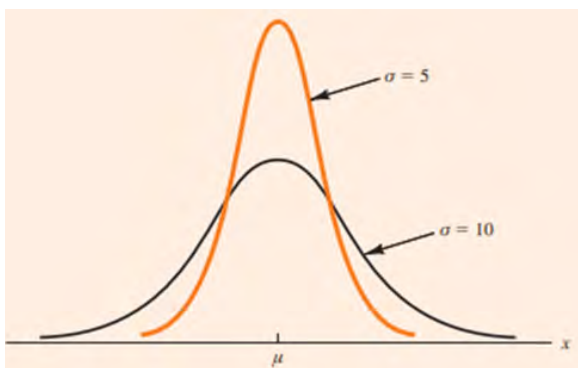
Gráfica 12. Distribución normal que tienen la misma desviación estándar.



4. La distribución normal es simétrica; esto significa que la forma de la curva a la izquierda de la media es un reflejo especular de la forma a la derecha de la media. Los extremos de la curva normal se extienden indefinidamente en ambas direcciones y, en teoría, nunca tocan el eje horizontal. Dado que son simétricas, las distribuciones normales no tienen sesgo; su medida de sesgo es cero.

5. La desviación estándar determina qué tan plana y ancha es la curva normal. Cuando la desviación estándar es grande, la curva se vuelve más ancha y plana, lo que indica una mayor variabilidad en los datos. A continuación, se presentan dos distribuciones normales con medias idénticas, pero desviaciones estándar distintas.

Gráfica 13. Distribución normal con la misma media y diferentes desviaciones estándar.

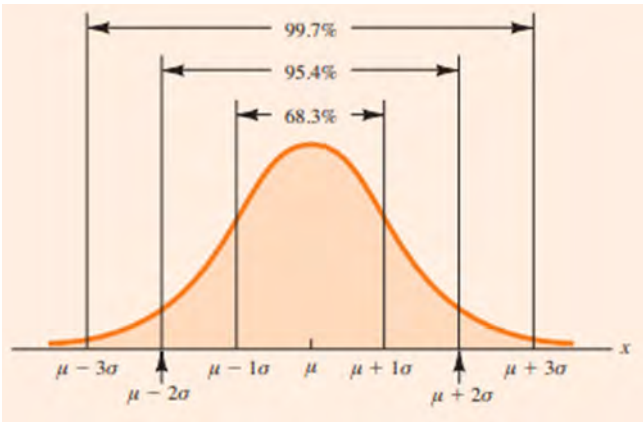


6. Las probabilidades asociadas con una variable aleatoria normal se representan mediante las áreas que se encuentran bajo la curva normal. El área total debajo de la curva de una distribución normal siempre suma 1. Debido a su simetría, el área bajo la curva a la izquierda de la media es 0.50 y el área a la derecha también es 0.50.

7. Los porcentajes de los valores en algunos intervalos de uso común son los siguientes.

- a) Alrededor del 68.3% de los valores en una distribución normal se encuentran a una desviación estándar de su media.
- b) Cerca del 95.4% de los valores en una distribución normal se ubican a dos desviaciones estándar de su media.
- c) Aproximadamente el 99.7% de los valores en una distribución normal están dentro de tres desviaciones estándar de su media.

Gráfica 14. Áreas bajo la curva de cualquier distribución normal.

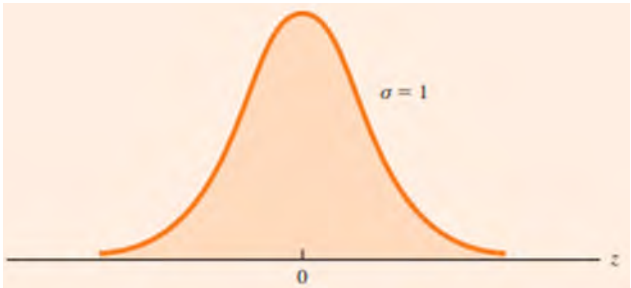


**4.1.3. Distribución de probabilidad normal estándar**

Una variable aleatoria que muestra una distribución normal con una media de cero y una desviación estándar de uno tiene una distribución de probabilidad normal estándar. La letra  $z$  se usa comúnmente para designar esta variable aleatoria normal.

La grafica muestra la distribución normal estándar, la cual tiene la misma apariencia que otras distribuciones normales, pero con las propiedades especiales de  $\mu = 0$  y  $\sigma = 1$ .

Gráfica 15. Distribución normal estándar.



Como  $\mu = 0$  y  $\sigma = 1$ , la fórmula para la función de densidad de probabilidad normal estándar es una versión más sencilla de la ecuación (6.2).

### Función De Densidad Normal Estándar

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

Donde:

$$\pi = 3.14159$$

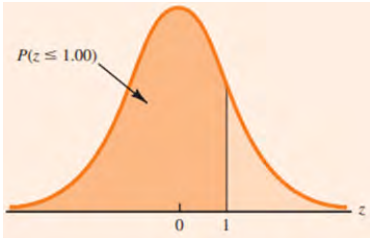
$$e = 2.71828$$

Los tres tipos de probabilidades que se necesita calcular incluyen:

- 1) La probabilidad de que la variable aleatoria normal estándar “z” sea igual o menor que un valor específico
- 2) la probabilidad de que z esté entre dos valores dados, y
- 3) la probabilidad de que z sea mayor o igual que un valor determinado. Para conocer cómo se usa la tabla de probabilidad acumulada de la distribución normal estándar con el propósito de calcular estos tres tipos de probabilidades, considere algunos ejemplos.

Primero se mostrará cómo calcular la probabilidad de que z sea menor o igual que 1.00, esto es,  $P(z \leq 1.00)$ . Esta probabilidad acumulada es el área bajo la curva normal a la izquierda de z “1.00 en la gráfica siguiente.

Gráfica 16. Áreas bajo la curva.



Luego, se consulta la tabla de probabilidad normal estándar. La probabilidad acumulativa correspondiente a  $z = 1.00$  se encuentra en la intersección de la fila etiquetada como 1.0 y la columna marcada como 0.00. En primer lugar, se localiza el valor 1.0 en la columna izquierda de la tabla y luego el valor 0.00 en la fila superior. Al examinar la tabla, se identifica que la fila 1.0 y la columna 0.00 se cruzan en el valor 0.8413. Por lo tanto,  $P(z \leq 1.00)$  es igual a 0.8413. El siguiente fragmento de la tabla de probabilidad muestra estos pasos.

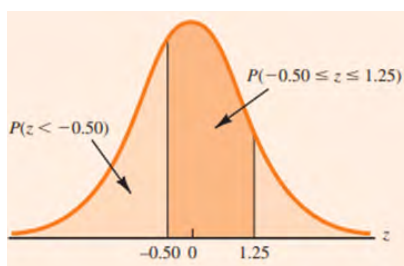
| z   | 0,00   | 0,01   | 0,02   |
|-----|--------|--------|--------|
| .   |        |        |        |
| .   |        |        |        |
| .   |        |        |        |
| 0,9 | 0.8159 | 0.8186 | 0.8212 |
| 1,0 | 0.8413 | 0.8438 | 0.8461 |
| 1,1 | 0.8643 | 0.8665 | 0.8686 |
| 1,2 | 0.8849 | 0.8869 | 0.8888 |
| .   |        |        |        |
| .   |        |        |        |
| .   |        |        |        |

$P(z \leq 1.00)$

Para el segundo escenario de cálculo de probabilidad, imaginemos que queremos averiguar la probabilidad de que “z” se

encuentre en el intervalo comprendido entre  $-0.50$  y  $1.25$ , es decir,  $P(-0.50 \leq z \leq 1.25)$ . El gráfico a continuación representa el área o probabilidad.

Gráfica 17. Áreas bajo la curva con intervalos.

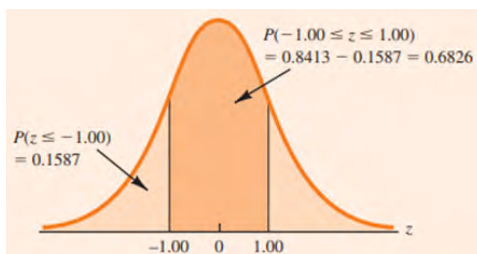


El cálculo de esta probabilidad implica tres pasos. En primer lugar, se determina el área bajo la curva normal a la izquierda de  $z = 1.25$ . En segundo lugar, se calcula el área bajo la curva normal a la izquierda de  $z = -0.50$ . Y, finalmente, restando el área a la izquierda de  $z = -0.50$  del área a la izquierda de  $z = 1.25$ , obtenemos  $P(-0.50 \leq z \leq 1.25)$ .

Para calcular el área bajo la curva normal a la izquierda de  $z = 1.25$ , primero ubicamos la fila 1.2 en la tabla de probabilidad normal estándar y luego seguimos hasta la columna 0.05. Dado que el valor que encontramos en la intersección de la fila 1.2 y la columna 0.05 es 0.8944, entonces  $P(z \leq 1.25)$  es igual a 0.8944. De manera similar, al calcular el área bajo la curva a la izquierda de  $z = -0.50$ , consultamos la tabla para encontrar el valor en la fila -0.5 y la columna 0.00, que resulta ser 0.3085. Por tanto,  $P(-0.50 \leq z \leq 1.25) = P(z \leq 1.25) - P(z \leq -0.50) = 0.8944 - 0.3085 = 0.5859$ .

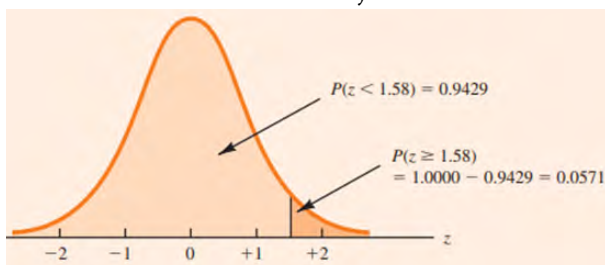
Otro caso de cálculo de probabilidad implica determinar la probabilidad de que “ $z$ ” se encuentre en el intervalo entre dos valores específicos.

Gráfica 18. Áreas bajo la curva con intervalos.



Para ilustrar cómo se realiza el tercer tipo de cálculo de probabilidad, consideremos el escenario en el que deseamos determinar la probabilidad de obtener un valor “ $z$ ” igual o mayor a 1.58, es decir,  $P(z \geq 1.58)$ . Al consultar la tabla de probabilidad acumulada, encontramos que el valor en la fila  $z = 1.5$  y la columna 0.08 es 0.9429, lo que significa que  $P(z \leq 1.58)$  es igual a 0.9429. Sin embargo, dado que el área total bajo la curva normal es siempre 1,  $P(z \geq 1.58)$  se calcula restando 0.9429 de 1, lo que resulta en 0.0571. Esta probabilidad se representa en la figura a continuación

Gráfica 19. Áreas bajo la curva.



Fuente: Anderson et al., 2012, pp. 238-244.

#### 4.1.4. Distribución de probabilidad normal estándar

Existen un número infinito de distribuciones normales, y cada una de ellas tiene distintas medias ( $\mu$ ), desviaciones estándar

dar ( $\sigma$ ), o incluso ambas características.

La distribución de probabilidad normal estándar es única, debido a su media de 0 y su desviación estándar de 1.

Toda distribución de probabilidad normal puede transformarse en una distribución de probabilidad normal estándar al restar la media de cada observación y luego dividir esa diferencia entre la desviación estándar. Estos resultados son conocidos como valores “z” o valores tipificados.

VALOR z Distancia con signo entre un valor seleccionado, designado X, y la media,  $\mu$ , dividida entre la desviación estándar,  $\sigma$

De esta manera, el valor z es la distancia de la media, medida en unidades de desviación estándar.

En términos de una fórmula,

Valor Normal Estándar  $z = (X - \mu) / \sigma$

En donde:

X es el valor de cualquier observación y medición.

$\mu$  es la media de la distribución.

$\sigma$  es la desviación estándar de la distribución

#### **4.1.5. Conclusión**

La distribución de probabilidad más importante para describir una variable continua es la distribución de probabilidad normal. Esta distribución ofrece una representación matemática fundamental para comprender y analizar la probabilidad de una amplia variedad de resultados en diversas aplicaciones, como la antropometría, la evaluación educativa, la investigación científica y la climatología. Además, su capacidad para describir la probabilidad de resultados en un proceso de muestreo es funda-

mental en la inferencia estadística, que es un tema central en la estadística. La distribución normal es una herramienta esencial para cuantificar la incertidumbre y entender la variabilidad en datos continuos, lo que la convierte en un elemento central en el análisis y la toma de decisiones basadas en datos.

#### **4.1.6. Autoevaluación**

¿Cuáles son las características de la distribución normal, comentando una de ellas?

## 4.2. Análisis de regresión

El análisis de regresión es un enfoque estadístico que permite investigar la conexión entre dos o más variables y determinar cuáles de ellas influyen de manera más significativa en un tema particular de interés.

En esta sección, se desarrolla una ecuación que describe la relación lineal entre dos variables, y se busca estimar el valor de la variable dependiente  $Y$  basado en un valor predefinido de la variable independiente  $X$ .

Este proceso se conoce como análisis de regresión y se utiliza para establecer relaciones y hacer predicciones en el contexto de la estadística.

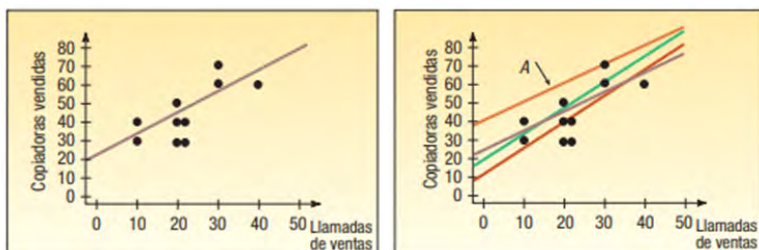
**Ecuación de Regresión.**—Ecuación que expresa la relación lineal entre dos variables

### 4.2.1. Principio de los mínimos cuadrados

Se establece una ecuación de regresión al minimizar la suma de los cuadrados de las diferencias verticales entre los valores reales de  $Y$  y los valores predichos de  $Y$ .

En el análisis de regresión, el propósito es emplear los datos para encontrar una línea que se ajuste de manera óptima a la relación entre las dos variables. Un primer paso consiste en utilizar un gráfico de dispersión para visualizar la posición de esta línea.

Gráfica 20. Análisis de regresión, diagrama de puntos.



El principio de los mínimos cuadrados se utiliza para encontrar la línea que mejor se ajusta a un conjunto de datos, lo que a menudo se denomina recta del mejor ajuste.

La ecuación de una recta tiene la forma:

FORMA GENERAL DE LA ECUACIÓN DE REGRESIÓN LINEAL =  $a + bX$

Donde:

$\hat{Y}$  = que se lee Y prima, es el valor de la estimación de la variable Y para un valor X seleccionado.

$a$  = es la intersección Y. Es el valor estimado de Y cuando  $X = 0$ . En otras palabras,  $a$  es el valor estimado de Y donde la recta de regresión cruza el eje Y cuando X es cero.

$b$  = es la pendiente de la recta, o el cambio promedio en por cada cambio de una unidad (ya sea aumento o reducción) de la variable independiente X.

$X$  = es cualquier valor de la variable independiente que se seleccione

La ecuación de regresión lineal tiene la misma estructura general que cualquier ecuación de una línea recta, con “a” representando la intersección en el eje Y y “b” como la pendiente. En el

análisis de regresión, el objetivo es calcular los valores específicos de “a” y “b” para crear una ecuación lineal que se adapte de manera óptima a los datos. Las fórmulas para calcular “a” y “b” son las siguientes:

$$\text{Pendiente de la recta de regresión} = b = r \frac{sy}{sx}$$

Donde:

$r$  representa el coeficiente de correlación.

$S_y$  corresponde a la desviación estándar de Y, que es la variable dependiente

$S_x$  se refiere a la desviación estándar de X, que es la variable independiente.

Intersección con el eje Y  $a = \bar{y} - b\bar{x}$

En esta fórmula:

$\bar{y}$  representa la media de Y, que es la variable dependiente

$\bar{x}$  representa la media de X, que es la variable independiente.

### 4.2.2. Conclusión

El análisis de regresión es una técnica estadística utilizada para representar la relación lineal entre dos variables. Su objetivo principal es estimar el valor de la variable dependiente (Y) en función de un valor específico de la variable independiente (X). Para lograrlo, se aplica el principio de los mínimos cuadrados, que busca encontrar la ecuación de regresión que minimice la suma de los cuadrados de las diferencias entre los valores reales de Y y los valores estimados de Y.

El principio de los mínimos cuadrados asegura que la línea de regresión obtenida sea la que mejor se ajusta a los datos disponibles, ya que reduce al mínimo el error cuadrático total. Esto

implica que la línea de regresión minimiza la distancia vertical promedio entre los puntos de datos reales y la propia línea.

#### **4.2.3. Autoevaluación.**

¿Qué es el análisis de regresión?

¿Cuál es el principio de los mínimos cuadrados?

### 4.3. Análisis de correlación

El análisis de correlación involucra una serie de técnicas para evaluar la relación entre dos variables. Su propósito fundamental radica en examinar la conexión entre dichas variables.

Por lo general, el proceso comienza con la representación de los datos en un gráfico de dispersión, lo que proporciona una visualización de la relación en cuestión. Luego, se procede a calcular el coeficiente de correlación, que proporciona una medida numérica de la fuerza de la relación entre ambas variables.

A modo de ejemplo, consideremos el caso de un gerente de ventas en “Copier Sales of America,” una empresa con una extensa fuerza de ventas en Estados Unidos y Canadá, que busca determinar si existe alguna relación entre el número de llamadas de ventas realizadas en un mes y la cantidad de copiadoras vendidas durante ese período (Lind et al., 2012, p. 463).

#### 4.3.1. Coeficiente de correlación

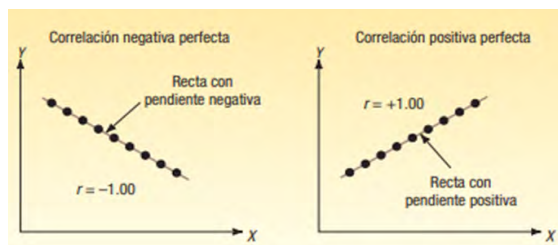
El coeficiente de correlación describe la fuerza de la relación entre dos conjuntos de variables que están en una escala de intervalo o razón. Se representa generalmente con la letra “ $r$ ” y a menudo se conoce como el coeficiente de correlación de Pearson o coeficiente de correlación producto-momento.

Este coeficiente puede tomar valores en el rango de -1.00 a +1.00, inclusive. Un coeficiente de correlación de -1.00 o +1.00 indica una correlación perfecta.

Por ejemplo, un coeficiente de correlación de +1.00 en el caso anterior indicaría que el número de llamadas de ventas y la cantidad de copiadoras vendidas por cada representante están perfectamente relacionados en una forma lineal positiva. En con-

traste, un valor de  $-1.00$  indicaría que las llamadas de ventas y la cantidad de copiadoras vendidas están perfectamente relacionadas de manera inversa en una forma lineal.

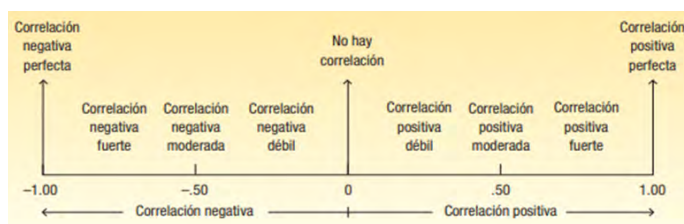
Gráfica 21. Diagrama de dispersión con correlación negativa perfecta y correlación positiva perfecta.



Cuando no existe una relación entre dos conjuntos de variables, el coeficiente de correlación de Pearson  $r$  es igual a cero. Un valor de coeficiente de correlación “ $r$ ” cercano a 0, ya sea positivo (como 0.08) o negativo ( $-0.08$ ), sugiere que la relación lineal entre las variables es muy débil. Los coeficientes de  $-0.91$  y  $+0.91$  tienen la misma magnitud y ambos indican una correlación muy fuerte entre las dos variables. En resumen, la fuerza de la correlación no depende de su dirección (positiva o negativa), sino de la magnitud del coeficiente.

El siguiente gráfico resume tanto la magnitud como la dirección del coeficiente de correlación.

Gráfica 22. Diagrama de correlación negativa y correlación positiva.



**Coefficiente De Correlación** medida que indica la intensidad de la relación lineal entre dos variables.

A continuación, se presentan las características del coeficiente de correlación.

### 4.3.2. Características del coeficiente de correlación

1. El coeficiente de correlación de la muestra se representa con la letra “r” en minúscula.
2. Ofrece información sobre la dirección y la intensidad de la relación lineal entre dos variables en una escala de intervalo o razón.
3. Se encuentra en el rango de -1 hasta +1, incluyendo ambos extremos
4. Un valor próximo a 0 indica una baja asociación entre las variables.
5. Un valor próximo a 1 refleja una relación positiva o directa entre las variables
6. Un valor cercano a -1 indica una relación negativa o inversa entre ellas.

$$\text{Coeficiente De Correlación } r = \frac{\sum(x-\bar{x})(y-\bar{y})}{(n-1)s_x s_y}$$

### 4.3.3. Conclusión

El coeficiente de correlación, conocido como “r” de Pearson o coeficiente de correlación de Pearson, es una herramienta estadística utilizada para cuantificar la magnitud y dirección de la relación lineal entre dos variables en una escala de intervalo o razón. Este coeficiente se encuentra en un rango de valores entre -1.00 y +1.00, inclusive. A continuación, se destacan algunas

conclusiones técnicas fundamentales sobre el coeficiente de correlación:

- **Magnitud de la relación:** El valor absoluto del coeficiente de correlación ( $|r|$ ) indica la intensidad de la relación entre las dos variables. Cuanto más cercano esté  $|r|$  a 1.00, ya sea positivo o negativo, mayor será la fuerza de la relación. Un valor cercano a 0 sugiere una relación débil.
- **Dirección de la relación:** El signo del coeficiente (positivo o negativo) indica la dirección de la relación. Un valor positivo de “ $r$ ” denota una relación positiva, lo que significa que cuando una variable aumenta, la otra tiende a aumentar también. Un valor negativo de “ $r$ ” señala una relación negativa, lo que indica que cuando una variable aumenta, la otra tiende a disminuir.
- **Correlación perfecta:** Un coeficiente de correlación de +1.00 o -1.00 representa una correlación perfecta. En el caso de +1.00, ambas variables están perfectamente relacionadas de manera positiva, lo que implica que aumentan juntas en una proporción constante. En el caso de -1.00, ambas variables están perfectamente relacionadas de manera negativa, lo que significa que una aumenta mientras que la otra disminuye en una proporción constante.
- **Limitaciones:** Es importante destacar que el coeficiente de correlación de Pearson solo evalúa relaciones lineales y no identifica relaciones no lineales. Además, no implica causalidad; simplemente señala que dos variables están relacionadas, pero no necesariamente una causa la otra.

#### **4.3.4. Autoevaluación.**

¿Cuáles son las características del coeficiente de correlación?

## 4.4. Pronósticos y series

### 4.4.1. Introducción

Un pronóstico es una estimación o una predicción de lo que puede ocurrir en el futuro, basada en datos, análisis y tendencias del pasado o del presente. Un pronóstico se puede aplicar a diferentes ámbitos, como el clima, la economía, la salud, la tecnología, etc. Un pronóstico tiene un grado de incertidumbre y un margen de error, por lo que no es una certeza ni una garantía de lo que va a pasar. Un pronóstico se puede hacer usando diferentes métodos, como el análisis estadístico, el juicio de expertos, la intuición, etc. Un pronóstico se puede expresar de forma cualitativa o cuantitativa, según el tipo de información que se quiera transmitir.

Las estimaciones o pronósticos desempeñan un papel fundamental en la toma de decisiones en una variedad de situaciones, que abarcan desde la gestión del inventario en una pequeña tienda de calzado hasta la predicción de las ventas anuales de videojuegos. La precisión de estas previsiones está estrechamente vinculada a la información que podemos extraer y utilizar a partir de datos históricos. El análisis de series temporales se convierte en una herramienta cuantitativa que aplicamos para identificar patrones en datos recopilados a lo largo del tiempo. Este análisis nos posibilita detectar tendencias y cambios en la información estadística a intervalos regulares y emplear esta información para realizar proyecciones hacia el futuro. El análisis de series temporales nos ayuda a abordar la incertidumbre asociada con los eventos futuros.

Tabla 1. Serie de datos.

| Serie de tiempo para el número de buques cargados, en Morehead, Carolina del Norte |      |      |      |      |      |      |      |      |
|------------------------------------------------------------------------------------|------|------|------|------|------|------|------|------|
| AÑO                                                                                | 1988 | 1989 | 1990 | 1991 | 1992 | 1993 | 1994 | 1995 |
| NUMERO                                                                             | 98   | 105  | 116  | 119  | 135  | 156  | 177  | 208  |

**4.4.2. Variación en las series de tiempo**

La “serie” hace referencia a una secuencia de datos que re- presenta observaciones de una variable en relación con algún fac- tor o variable independiente. Esta secuencia de datos se emplea para examinar y comprender patrones, tendencias y relaciones dentro de los datos.

El término “serie de tiempo” se utiliza para caracterizar cualquier conjunto de datos estadísticos que se recopila en in- tervalos regulares. En el análisis de series de tiempo, se pueden identificar cuatro categorías de cambios o variaciones:

- 1. Tendencia secular
- 2. Fluctuación cíclica
- 3. Variación estacional
- 4. Variación irregular

*4.4.2.1. Tendencia secular*

En el primer tipo de variación, denominado “tendencia se- cular”, se observa que el valor de una variable tiende a mostrar un incremento o una disminución a lo largo de un período de tiem- po muy prolongado. Un ejemplo de esto podría ser el constante aumento en los costos de vida reflejados en el Índice de Precios al Consumidor (IPC). Aunque de un año a otro los costos de vida pueden experimentar variaciones significativas, si analizamos esta información a lo largo de un período extenso, notamos que la tendencia general es un aumento constante.

#### 4.4.2.2. *Fluctuación cíclica*

El segundo tipo de variación que podemos identificar en una serie de tiempo se conoce como “fluctuación cíclica”. Un ejemplo muy común de esto es el ciclo económico. Con el paso del tiempo, es posible observar años en los que la economía alcanza su punto más alto, superando la línea de tendencia, mientras que en otros años es probable que la actividad económica descienda por debajo de esta línea. Estos ciclos suelen tener una duración de al menos un año y, en algunos casos, pueden extenderse hasta 15 o 20 años. La gráfica (b), ilustra un patrón típico de fluctuación cíclica que oscila en relación con la línea de tendencia a largo plazo. Es importante destacar que estos movimientos cíclicos no siguen un patrón regular y a menudo son algo impredecibles.

#### 4.4.2.3. *Variación estacional*

El tercer tipo de variación que podemos identificar en los datos de una serie de tiempo se denomina “variación estacional”. En este contexto, nos referimos a los patrones de cambio que ocurren a lo largo de un año y tienden a repetirse de manera anual. Por ejemplo, un médico podría prever un aumento notable en el número de casos de gripe durante cada invierno y un incremento en los casos de tifoidea durante el verano. Dado que estos patrones siguen un ciclo regular, son altamente útiles para realizar predicciones sobre el futuro.

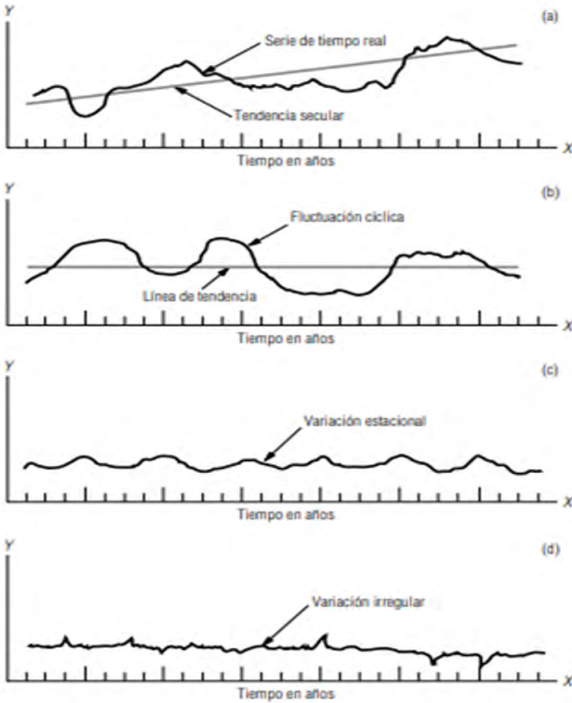
#### 4.4.2.4. *Variación irregular*

El cuarto tipo de variación que se identifica en el análisis de series de tiempo se conoce como “variación irregular”. En ocasiones, una variable puede experimentar cambios que son completamente impredecibles y aleatorios, y es precisamente esta variación irregular la que describe estos movimientos imprede-

cibles. Ejemplos de esta variación irregular incluyen los impactos imprevisibles que eventos como el conflicto en Medio Oriente en 1973, la situación en Irán entre 1979 y 1981, la disolución de la OPEP en 1986 y los acontecimientos en Irak en 1990 tuvieron en los precios de la gasolina en Estados Unidos.

Es importante destacar que, en la mayoría de los casos, las series de tiempo no presentan exclusivamente uno de estos tipos de cambios, sino que suelen contener varias de estas componentes. Por lo tanto, podemos describir la variación total en una serie de tiempo en términos de estas cuatro clases de variación

Gráfica 23. Variación en las series de tiempo.



**4.4.3. Análisis de tendencia**

Dentro de las cuatro componentes que conforman una serie de tiempo, la tendencia a largo plazo de la serie se refleja en la tendencia secular. Una manera de caracterizar esta componente es realizar un ajuste visual mediante una línea recta a un conjunto de puntos en un gráfico. No obstante, esta representación gráfica puede dar lugar a interpretaciones que difieren de una persona a otra. Una alternativa es emplear el método de mínimos cuadrados para realizar un ajuste de la línea de tendencia. En nuestro análisis, nos centraremos en el método de mínimos cuadrados, ya que el proceso de ajustar visualmente una línea recta a una serie de tiempo no es completamente fiable.

#### *4.4.3.1. Razones para estudiar tendencias*

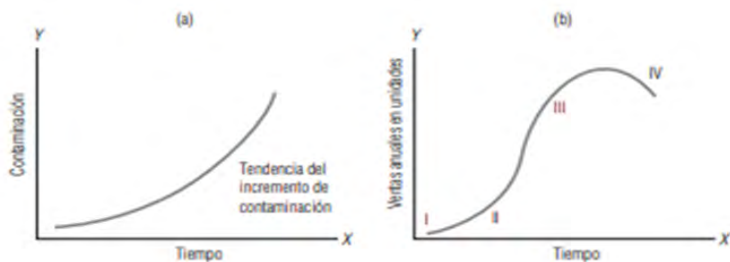
Existen tres razones fundamentales para llevar a cabo el estudio de las tendencias a largo plazo:

1. El análisis de tendencias a largo plazo nos proporciona una comprensión de los patrones históricos, y esta información puede aplicarse en una variedad de contextos para evaluar la efectividad de políticas previas. Por ejemplo, una universidad puede evaluar la eficacia de su programa de reclutamiento de estudiantes al examinar las tendencias en inscripciones pasadas.
2. Estudiar las tendencias a largo plazo nos permite proyectar patrones pasados hacia el futuro, lo que brinda valiosa información sobre lo que podemos anticipar en los próximos períodos. Por ejemplo, al analizar la tasa de crecimiento histórico de la población mundial, podemos hacer estimaciones sobre la población en el futuro.
3. En diferentes situaciones, analizar la tendencia a largo plazo en una serie de tiempo nos permite separar esa

parte de la serie, simplificando así el estudio de las otras tres componentes. Por ejemplo, si deseamos examinar la variación estacional en las ventas de equipos de esquí, al eliminar la tendencia a largo plazo, obtenemos una representación más precisa de la componente estacional.

Las líneas de tendencia pueden tomar diferentes formas, ya sea lineales o curvas. Es importante tener en cuenta que no todas las relaciones se ajustan a una línea recta; algunas siguen patrones curvilíneos. Por ejemplo, el aumento de contaminantes en el medio ambiente suele seguir una curva de crecimiento ascendente, como se ilustra en la gráfica. Otro ejemplo común de una relación curvilínea es el ciclo de vida de un nuevo producto comercial, que se representa en la gráfica (b) de la misma figura. Cuando un producto nuevo se introduce en el mercado, sus ventas son inicialmente bajas (I). A medida que el producto gana reconocimiento y éxito, las ventas unitarias aumentan a un ritmo cada vez más rápido (II). Luego, una vez que el producto está bien establecido, sus ventas unitarias crecen a un ritmo constante (III). Finalmente, cuando el producto alcanza el final de su ciclo de vida, las ventas unitarias comienzan a disminuir (IV).

Gráfica 24. Relaciones de tendencia curvilínea.



#### 4.4.3.2. Ajuste de la tendencia lineal con el método de mini-

## *mos cuadrados*

Además de las tendencias que pueden ser representadas mediante una curva, también existen tendencias que pueden ser expresadas mediante una línea recta. Estas últimas son conocidas como tendencias lineales. Antes de avanzar en la creación de la ecuación para una tendencia lineal, es importante revisar la ecuación general utilizada para estimar una línea recta, que se encuentra en la ecuación.

La fórmula para calcular una línea recta es:  $\hat{y} = a + bx$

En esta ecuación:

$\hat{y}$  representa el valor estimado de la variable dependiente.

X corresponde a la variable independiente, que en este contexto es el tiempo en el análisis de tendencia.

“a” representa la ordenada Y, es decir, el valor de Y cuando X es igual a cero.

“b” es la pendiente de la línea de tendencia recta.

Es factible representar la tendencia general de numerosas series de tiempo a través de una línea recta. No obstante, se presenta el desafío de determinar la recta o ecuación que logre un ajuste óptimo.

Pendiente de la recta de regresión de mejor ajuste

$$b = \frac{\sum xY - n\bar{X}\bar{y}}{\sum x_{-nX}^2}$$

Ordenada Y de la recta de regresión de mejor ajuste

$$a = \bar{y} - b\bar{x}$$

En esta fórmula:

$Y$  representa los valores de la variable dependiente.

$X$  corresponde a los valores de la variable independiente.

$\bar{y}$  simboliza la media de los valores de la variable dependiente.

$\bar{x}$  es la media de los valores de la variable independiente.

“ $n$ ” representa la cantidad de datos en la serie de tiempo.

“ $a$ ” representa la ordenada  $Y$ .

“ $b$ ” es la pendiente de la línea de tendencia.

A través de las 2 ecuaciones, podemos calcular la línea recta que ofrece la mejor descripción de los datos en la serie. Sin embargo, debido a la regularidad de los datos en la serie de tiempo, tenemos la oportunidad de simplificar los cálculos relacionados con estas ecuaciones. A continuación, vamos a explicar el proceso de simplificación.

**Pendiente de la línea de tendencia para valores de tiempo codificados**

$$b = \frac{\sum xy}{\sum x^2} \cdot 2$$

**Ordenada  $Y$  de la recta de tendencia para valores de tiempo codificados**

$$a = \bar{y}$$

#### **4.4.4. Variación Cíclica**

La variación cíclica en una serie de tiempo se caracteriza por fluctuar por encima y por debajo de la línea de tendencia a largo plazo en intervalos que superan un año. Para identificar esta variación cíclica, se emplea un enfoque conocido como el método de residuos.

#### 4.4.4.1. Método de residuos

Al analizar una serie de tiempo que consiste en datos anuales, nuestra atención se enfoca en las componentes de tendencia secular, cíclica e irregular. Esto se debe a que la variación estacional sigue un patrón anual predecible que no difiere significativamente de un año a otro. Dado que podemos describir la tendencia secular mediante una línea de tendencia, podemos aislar las componentes cíclicas e irregulares de la variación que no se explica por la tendencia. Hacemos la suposición de que la mayor parte de esta variación no explicada por la tendencia se debe a la componente cíclica, aunque es importante destacar que, en muchas series de tiempo reales, esta suposición no siempre se cumple. En tales casos, se emplean métodos como el análisis de Fourier y el análisis espectral para analizar la componente cíclica de estas series de tiempo.

Al trabajar con una serie de tiempo que incluye datos anuales, podemos calcular la fracción que representa la tendencia dividiendo el valor real ( $Y$ ) entre el valor estimado de la tendencia ( $\hat{Y}$ ) en cada punto de la serie de tiempo. Luego, multiplicamos este resultado por 100 para expresar la variación cíclica como un porcentaje de la tendencia.

#### Porcentaje de tendencia

$$\frac{y}{\hat{y}} \times 100$$

Donde:

$Y$  = valor real de la serie de tiempo

$\hat{y}$  = valor de tendencia estimado a partir del mismo punto de la serie de tiempo

El residuo cíclico relativo representa otra forma de medir

la variación cíclica. En este enfoque, se calcula el porcentaje de cambio con respecto a la tendencia para cada valor.

**Residuo cíclico relativo**

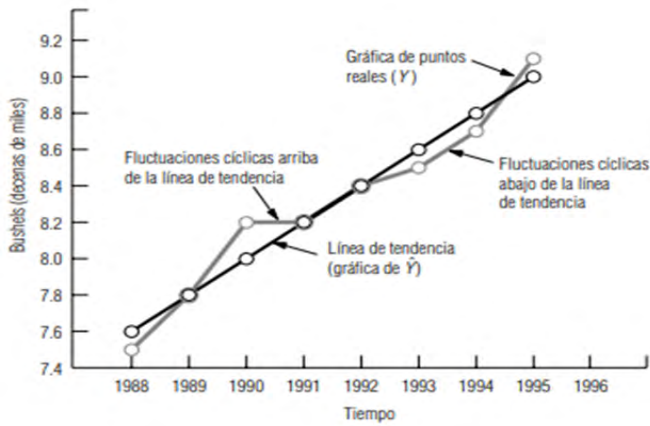
$$\frac{y - \hat{y}}{\hat{y}} * 100$$

En esta fórmula:

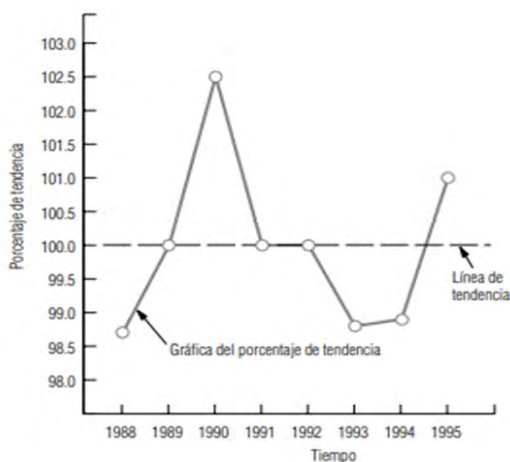
Y= representa el valor real de la serie de tiempo

$\hat{y}$  valor de tendencia estimado a partir del mismo punto de la serie de tiempo

Gráfica 25. Fluctuaciones cíclicas alrededor de la línea de tendencia.



Gráfica 26. Gráfica del porcentaje de tendencia alrededor de la línea de tendencia.



#### 4.4.5 Variación estacional

Aparte de la tendencia a largo plazo y la variación cíclica, también se encuentra presente en una serie de tiempo la variación estacional. Esta forma de variación se caracteriza por movimientos repetitivos y previsibles alrededor de la línea de tendencia dentro de un período de un año o menos. Para identificar esta variación estacional, es necesario desglosar los intervalos de tiempo en unidades más pequeñas, como días, semanas, meses o trimestres.

Hay tres motivos fundamentales para investigar y comprender la variación estacional:

1. Facilita la identificación de los patrones de cambios previos, lo que proporciona un método para contrastar dos períodos que, de lo contrario, serían muy distintos. Por ejemplo, si una escuela de formación de pilotos desea determinar si una disminución en la demanda duran-

te el mes de diciembre es una tendencia común, puede analizar los comportamientos estacionales en años previos y obtener la información requerida.

2. Resulta valioso proyectar los patrones anteriores hacia el futuro. Mientras que el análisis de tendencia a largo plazo puede ser apropiado para decisiones que abarcan un período extenso, la habilidad de anticipar las variaciones estacionales generalmente se vuelve esencial en decisiones a corto plazo. Por ejemplo, una cadena mayorista de alimentos que busca mantener niveles de inventario adecuados en su gama de productos necesita la capacidad de prever patrones a corto plazo, como la demanda de pavos en la temporada navideña, dulces en el Día del Niño o duraznos en verano, con el fin de gestionar eficazmente su cadena de suministro.
3. Luego que hemos reconocido y definido el patrón estacional en una serie de tiempo, podemos proceder a eliminar su impacto en esa serie. Este ajuste nos posibilita calcular la variación cíclica que tiene lugar cada año. Al suprimir la influencia de la variación estacional en una serie de tiempo, estamos ejecutando una técnica denominada desestacionalización de la serie.

#### **4.4.6. Método de razón de promedio móvil**

El método de razón de promedio móvil se utiliza comúnmente para evaluar la variación estacional. Este enfoque proporciona un índice que describe el nivel de variación estacional, utilizando una base de referencia establecida en 100. Se mide el grado de estacionalidad en función de las desviaciones con respecto a esta base. Por ejemplo, al analizar la estacionalidad de los alquileres de canoas en un hotel de verano, podríamos encontrar

que el índice para el trimestre de primavera es 142. Esto significa que durante la primavera, se producen el 142% de los alquileres trimestrales promedio. Si la administración registró un total de 2,000 alquileres de canoas durante el año anterior, entonces el promedio trimestral sería de 2,000 dividido entre 4, lo que da como resultado 500 alquileres por trimestre. Dado que el índice para el trimestre de primavera es 142, podemos estimar la cantidad de alquileres de canoas de la siguiente manera.

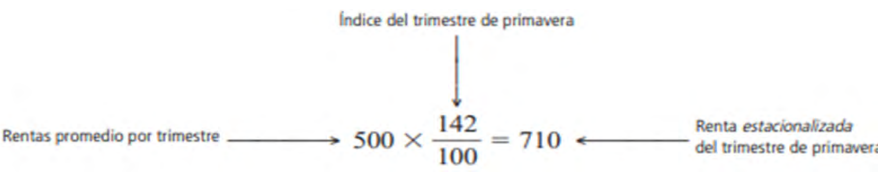


Diagram illustrating the calculation of the seasonalized rent for the spring quarter:

Índice del trimestre de primavera

Rentas promedio por trimestre  $\longrightarrow 500 \times \frac{142}{100} = 710 \longleftarrow$  Renta estacionalizada del trimestre de primavera

#### 4.4.7. Variación irregular

La fase final de una serie de tiempo corresponde a la variación irregular. Incluso después de haber eliminado las influencias de la tendencia, la variación cíclica y la estacionalidad en una serie de tiempo, todavía permanece una componente impredecible. Esta variación irregular típicamente se manifiesta en intervalos breves y sigue un patrón aleatorio (Levin et al., 2004, pp. 674-699).

#### 4.4.8. Conclusión

En el último segmento de este tema, hemos examinado las cuatro componentes fundamentales de una serie de tiempo. Hemos detallado el proceso de proyectar tanto la tendencia histórica como la variación estacional hacia el futuro, reconociendo las posibles imprecisiones que pueden surgir en este análisis. Además, hemos subrayado que, aunque las componentes cíclicas e irregu-

lares influyen en el comportamiento futuro, son factores volátiles y complejos de incorporar en pronósticos.

Es crucial tener en cuenta que el enfoque mecánico del análisis de series de tiempo está sujeto a errores y puede experimentar cambios significativos. Los administradores deben complementar estos métodos simples con su comprensión de otros factores para desarrollar pronósticos efectivos. Los analistas deben estar preparados para revisar, actualizar y ajustar continuamente sus pronósticos. Si deseamos gestionar el futuro con éxito, es esencial adoptar esta mentalidad.

Al poner en práctica los métodos descritos en este capítulo, es fundamental enfocar nuestra atención de manera particular en dos aspectos importantes:

1. En los pronósticos, proyectamos tendencias históricas y variaciones cíclicas hacia el futuro. Es fundamental cuestionarnos la regularidad y duración de estas tendencias pasadas y considerar si existe la posibilidad de que estos patrones estén experimentando cambios.
2. La precisión de los datos históricos que utilizamos en el análisis de series de tiempo es crucial. Por ejemplo, si una empresa cambió de un sistema de contabilidad de inventario PEPS (primero en entrar, primero en salir) a un sistema UEPS (último en entrar, primero en salir) en un período analizado, los datos, como las ganancias trimestrales, antes y después del cambio, no son comparables y no son muy útiles para fines de pronóstico.

#### **4.4.9. Autoevaluación**

¿Qué es un pronóstico?

¿Qué es una serie?



## Referencias

- Anderson, D., & García, C. (2019). *Estadística para negocios y economía*. Cengage Learning.
- Anderson, D., Sweeney, D., & Williams, T. (2012). *Estadística para negocios y economía*. Editores, S.A.
- Anderson, D., Sweeney, D., Williams, T., Camm, J., & Cochran, J. (2016). *Estadística para negocios y economía*. Editores, S.A.
- Espejo, I. (2016). *Inferencia estadística: teoría y problemas*. Servicio de Publicaciones de la Universidad de Cádiz.
- Gamero, C. (2017). *Estadística I: elementos de estadística descriptiva y teoría de la probabilidad*. Universidad de Málaga.
- Gamero, C. (2022). *Estadística*. Universidad de Málaga.
- García, J., Ramos, C., & Ruiz, G. (2016). *Estadística empresarial*. Servicio de Publicaciones de la Universidad de Cádiz.
- Gutiérrez, E., & Vladimirovna, O. (2016). *Estadística inferencial*. Grupo Editorial Patria.
- Hernández, M. (2019). *Estadística inferencial 2*. Grupo Editorial Patria.
- Levin, R., & Rubin, D. (2004). *Estadística para administración y economía*. Pearson educación.
- Lind, D., Marchal, W., & Wathen, S. (2012). *Estadística aplicada a los negocios y la economía*. Interamericana editores, S.A.
- Llinás, H. (2018). *Estadística Inferencial*. Editorial Universidad del Norte.
- Marín, J. (2020). *Estadística: gradon en ciencia y tecnología de los alimentos*. Editum–Ediciones de la Universidad de Murcia.
- Martínez, C. (2012). *Estadística y muestreo*. Ecoe ediciones.
- Martínez, C. (2019). *Estadística básica aplicada*. Ecoe Ediciones.
- Martínez, C. (2019). *Estadística y muestreo*. Ecoe Ediciones.

- Martínez, E. (2020). *Estadística*. Universidad Abierta para Adultos.
- Miller, I., & Freund, J. (2021). *Probabilidad y estadística para Ingenieros*. Editorial Reverte.
- Mode, E., & García, R. (2021). *Elementos de probabilidad y estadística*. Editorial Reverté.
- Obando, J., Arango, N., & Misas, M. (2019). *Probabilidad y estadística*. Fondo editorial EIA.
- Puente, C. (2018). *Estadística descriptiva e inferencial*. Ediciones IDT.
- Salazar, L. (2018). *Probabilidad y estadística: para bachilleratos tecnológicos*. Grupo Editorial Patria.
- Sánchez, E., Callejas, J., Inzunza, S., Martínez, J., Polo, A., & Vargas, G. (2019). *Probabilidad y estadística 1: Serie integral por competencias*. Grupo Editorial Patria.
- Spiegel, M., & Stephens, L. (2009). *Estadística*. Interamericana editores, S.A.
- Triola, M. (2013). *Estadística: Actualización Tecnológica*. Pearson Educación de Mexico, S.A.
- Villada, D., & Beltrán, O. (2021). *Elementos de estadística descriptiva y probabilidad*. Universidad Piloto de Colombia.

## **Anexo**

## **Anexo 1. Solucionario Capítulo 1**

### **1. ¿Cuál es la finalidad de la estadística?**

- Conocer la realidad de una observación o fenómeno
- Determinar lo típico o normal de esa observación
- Determinar los cambios que presenta el fenómeno
- Relacionar dos o más fenómenos
- Determinar las causas que originan el fenómeno
- Hacer estimativos sobre el comportamiento futuro del fenómeno
- Obtener conclusiones de un grupo menor (muestra), para hacerlas extensivas a un grupo mayor (población).

### **2. ¿Cuántos tipos de estadística existen?**

El estudio de la estadística se divide en dos categorías: la estadística descriptiva y la estadística inferencial.

- Estadística Descriptiva.–Métodos para organizar, resumir y presentar datos de manera informativa.
- Estadística Inferencial.–Métodos que se emplean para determinar una propiedad de una población con base en la información de una muestra de ella

### **3. ¿Cuál es la diferencia entre la estadística descriptiva y estadística inferencial?**

La estadística descriptiva se enfoca en la descripción y resumen de datos observados, mientras que la estadística inferencial se centra en la toma de decisiones y la obtención de conclusiones sobre poblaciones más amplias a partir de datos muestrales, utilizando métodos estadísticos y probabilísticos.

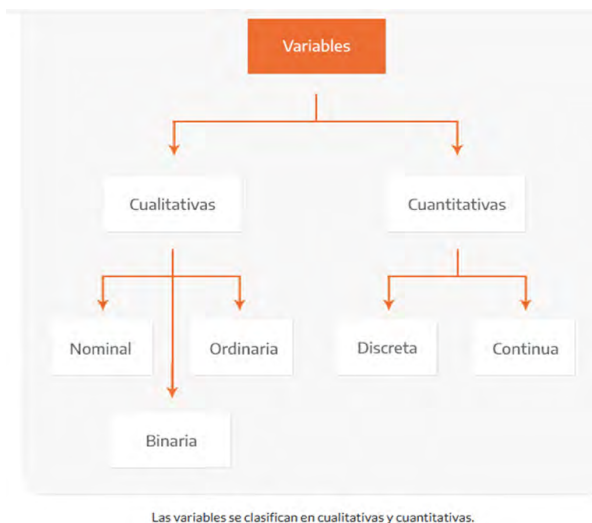
### **4. ¿Cuáles son los conceptos de variables cualitativas y cuantitativas?**

Variable cualitativa.–Cuando la característica que se estudia es de naturaleza no numérica, recibe el nombre de variable cualitativa o atributo. Algunos ejemplos de variables cualitativas son el género, la filiación religiosa, tipo de automóvil que se posee, estado de nacimiento y color de ojos.

Variable cuantitativa.–Cuando la variable que se estudia

aparece en forma numérica, se le denomina variable cuantitativa. Ejemplos de variables cuantitativas son el saldo en su cuenta de cheques, las edades de los presidentes de la compañía, la vida de la batería de un automóvil aproximadamente 42 meses y el número de hijos que hay en una familia.

## 5. ¿Cuál es la clasificación de las variables?



## 6. ¿Cuántos tipos de gráficos conoce?

Existen 6 tipos de gráficos que se utilizan frecuentemente en estadística.

- Gráfica de barras
- Gráfica de pastel
- Histograma
- Polígono de frecuencias
- Ojiva o polígono de frecuencias acumuladas
- Diagrama de puntos

**7. ¿Cuál es la diferencia entre la frecuencia absoluta y la frecuencia relativa?**

La frecuencia absoluta es el número de veces que se repite un dato o valor en un conjunto de datos. La frecuencia relativa es la proporción o el porcentaje que representa la frecuencia absoluta respecto al total de datos. Por ejemplo, si en una clase hay 20 estudiantes y 8 de ellos tienen el pelo rubio, la frecuencia absoluta del pelo rubio es 8 y la frecuencia relativa es 0,4 o el 40%. Para calcular la frecuencia relativa, se divide la frecuencia absoluta por el número total de datos.

**8. ¿Qué son los datos no agrupados?**

Los datos no agrupados son aquellos que no han sido clasificados ni ordenados según algún criterio. Son los datos que se obtienen directamente de una fuente o de una recolección, sin haber sido procesados ni organizados.

**9. ¿Ponga un ejemplo de datos no agrupados?**

Si se mide la altura de 10 personas, los datos no agrupados serían los 10 valores numéricos sin ordenar ni agrupar.

## Anexo 2. Solucionario Capítulo 2

### 1. ¿Qué es la media aritmética, la mediana y la moda?

**La media aritmética** es el promedio de todos los valores numéricos de un conjunto de datos.

**La mediana** es el valor que ocupa la posición central de un conjunto de datos ordenados de menor a mayor o de mayor a menor.

**La moda** es el valor que más se repite o que tiene mayor frecuencia en un conjunto de datos.

### 2. ¿Cuál es la diferencia entre la media aritmética y la media geométrica?

La media aritmética y la media geométrica se utilizan para diferentes propósitos según el tipo y la distribución de los datos. Por ejemplo, la media aritmética se usa para calcular el promedio simple de una serie de datos, mientras que la media geométrica se usa para calcular el promedio compuesto o el rendimiento efectivo de una serie de datos que varían en el tiempo.

### 3. ¿Qué es un rango?

El rango es una medida de dispersión que indica la diferencia entre el valor máximo y el valor mínimo de un conjunto de datos.

### 4. ¿Qué es el coeficiente de variación?

El coeficiente de variación es una medida estadística que sirve para determinar la dispersión relativa de un conjunto de datos respecto a su media.

### 5. ¿Qué es varianza, desviación estándar y desviación media?

**La varianza** mide la variabilidad o el grado de dispersión de los datos.

**La desviación estándar** mide la desviación típica o el promedio de las distancias entre cada dato y la media del conjunto.

**La desviación media** es el promedio de las distancias en valor absoluto entre cada dato y la media del conjunto. **La desviación media** mide la desviación absoluta o el promedio de las diferencias sin elevarlas al cuadrado.

**6. ¿Cuál es la diferencia entre cuartil y decil?**

El cuartil divide un conjunto de datos ordenados en cuatro partes iguales, mientras que el decil lo divide en diez partes iguales. Los cuartiles y los deciles son medidas de posición que se usan para describir la distribución de los datos y compararlos entre sí.

**7. Explique en un ejemplo como se calcula el cuartil**

Tenemos los siguientes datos de las edades de 12 personas: 15, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27 y 28

Para calcular el primer cuartil (Q1), que representa el 25% de los datos, usamos la fórmula:

$$Q1 = 25(12+1) / 100 = 325 / 100 = 3.25$$

Como el resultado es un número decimal, lo redondeamos al entero más próximo, que es 3. Luego, tomamos el dato en la posición 3, que es 19. Por lo tanto, el primer cuartil es 19.

**8. ¿Qué significa estar sesgada a la derecha?**

Estar sesgada a la derecha significa que una distribución de datos tiene una forma asimétrica en la que la mayoría de los datos se concentran en el lado izquierdo del gráfico y la cola se extiende hacia el lado derecho. Esto implica que hay valores más altos y más separados de la media en el lado derecho. Una distribución sesgada a la derecha también se llama distribución sesgada positivamente, porque el signo del coeficiente de asimetría es positivo.

**9. ¿Qué es la curtosis?**

La curtosis es una medida estadística que indica el grado de concentración de los valores de una variable alrededor de la media. La curtosis también se conoce como el grado de pico o apuntamiento de una distribución.

La curtosis es una medida útil para describir y comparar la forma y el comportamiento de diferentes distribuciones.

#### 10. **¿Qué es momento?**

El momento es un concepto que se usa en estadística para medir diferentes aspectos de una distribución de datos.

Los momentos se relacionan con parámetros estadísticos como la media, la varianza, la asimetría y la curtosis.

### **Anexo 3. Solucionario Capítulo 3**

**1. ¿Qué es un evento?**

Un evento o suceso es un subconjunto de un espacio muestral, es decir, un conjunto de posibles resultados que se pueden dar en un experimento aleatorio.

**2. ¿Qué es el espacio muestral?**

Se llama espacio muestral del experimento al conjunto de todos los resultados posibles de un experimento.

**3. ¿Qué es probabilidad condicional?**

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

**4. ¿Qué es una distribución de probabilidad?**

Lista de todos los resultados de un experimento y la probabilidad asociada a cada uno de ellos.

**5. ¿Cuáles son las características de una distribución de probabilidad?**

- La probabilidad de un resultado en particular se encuentra entre 0 y 1, inclusive.
- Los resultados son eventos mutuamente excluyentes
- La lista es exhaustiva. Por lo tanto, la suma de las probabilidades de los diversos eventos es igual a 1.

**6. ¿Qué es una distribución binomial?**

Una distribución binomial es una distribución de probabilidad discreta que describe el número de éxitos en una serie de ensayos independientes con una probabilidad fija de ocurrencia de éxito

**7. ¿Cuál es la característica de la distribución de Poisson?**

La característica de la distribución de Poisson es que describe la probabilidad de que ocurra un determinado número

ro de eventos en un intervalo fijo, cuando la probabilidad de cada evento es pequeña y constante, y los eventos son independientes entre sí.

## **Anexo 4. Solucionario Capítulo 4**

### **1. ¿Cuáles son las características de la distribución normal, comente una de ellas?**

Las características de la distribución normal son las siguientes:

- Es una distribución de probabilidad continua que describe el comportamiento de una variable aleatoria que puede tomar cualquier valor real.
- Tiene una forma simétrica y acampanada, que se conoce como campana de Gauss o función gaussiana.
- Está determinada por dos parámetros: la media y la desviación estándar, que indican el valor central y la dispersión de los datos, respectivamente.
- La media, la mediana y la moda coinciden y se ubican en el centro de la curva.
- El área bajo la curva es igual a 1, lo que significa que la suma de todas las probabilidades es 100%.
- Tiene algunas propiedades matemáticas que facilitan su cálculo y su uso en diferentes campos de estudio.
- Una de estas características es la simetría, que implica que los datos se distribuyen de forma equilibrada a ambos lados de la media

### **2. ¿Qué es el análisis de regresión?**

Es estimar el valor de la variable dependiente Y con base en un valor seleccionado de la variable independiente X. La técnica para desarrollar la ecuación y proporcionar las estimaciones se denomina análisis de regresión.

El análisis de regresión es un método estadístico que permite examinar la relación entre dos o más variables e identificar cuáles son las que tienen mayor impacto en un tema de interés.

### 3. **¿Cuál es el principio de los mínimos cuadrados?**

El principio de los mínimos cuadrados es un criterio que se usa para encontrar la función que mejor se ajusta a un conjunto de datos, minimizando la suma de los cuadrados de las diferencias entre los valores observados y los valores estimados por la función. Este principio se basa en la idea de que cuanto menor sea el error cuadrático, mayor será la precisión del ajuste. El principio de los mínimos cuadrados se aplica a diferentes tipos de funciones, como rectas, polinomios, exponenciales, logarítmicas, etc.

### 4. **¿Cuáles son las características del coeficiente de correlación?**

1. El coeficiente de correlación de la muestra se identifica con la letra minúscula  $r$ .
2. Muestra la dirección y fuerza de la relación lineal (recta) entre dos variables en escala de intervalo o en escala de razón.
3. Varía de  $-1$  hasta  $+1$ , inclusive.
4. Un valor cercano a  $0$  indica que hay poca asociación entre las variables.
5. Un valor cercano a  $1$  indica una asociación directa o positiva entre las variables.
6. Un valor cercano a  $-1$  indica una asociación inversa o negativa entre las variables.

### 5. **Qué es un pronóstico?**

Un pronóstico es una estimación o una predicción de lo que puede ocurrir en el futuro, basada en datos, análisis y tendencias del pasado o del presente. Un pronóstico se puede aplicar a diferentes ámbitos, como el clima, la economía, la salud, la tecnología, etc. Un pronóstico tiene un grado de incertidumbre y un margen de error, por lo que no es una certeza ni una garantía de lo que va a pasar. Un pronóstico se puede hacer usando diferentes métodos, como el análisis estadístico, el juicio de expertos, la intuición, etc. Un pronóstico se puede expresar de forma cualitativa o cuantitativa, según el tipo de información que se quiera transmitir.

## **6. ¿Qué es una serie?**

Una “serie” se refiere a una secuencia de datos que representa observaciones de una variable en función de algún factor o variable independiente. Esta secuencia de datos se utiliza para analizar y comprender patrones, tendencias y relaciones dentro de los datos. Las series pueden ser unidimensionales o multidimensionales, dependiendo de la cantidad de variables involucradas.





Religación  
**Press**

Ideas desde el Sur Global



Religación  
**Press**  
Ideas desde el Sur Global

**int** Instituto  
Nelson  
Torres

ISBN: 978-9942-642-39-4



9 789942 642394