RESOLUCION DE CONSEJO DE COMISIÓN ORGANIZADORA N° 166-2025-CCO-UNJ

Jaén, 21 de marzo de 2025.

VISTOS:

El Oficio N° 431-2025-VPI-CO-UNJ, de fecha 17 de marzo de 2025, de la Vicepresidenta de Investigación, Oficio N° 039-2025-UNJ/VPI-DIITT, de fecha 14 de marzo de 2025, del Director de la Dirección de Investigación, Innovación y Transferencia Tecnológica, Carta N° 17-2025/DACBYA-UNJ/RYLLS, de fecha 14 de marzo de 2025, de la Dra. Rosario Yaqueliny Llauce Santamaria, Carta N° 16-2025/DACBYA-UNJ/RYLLS, de fecha 14 de marzo de 2025, de la Dra. Rosario Yaqueliny Llauce Santamaria, Carta N° 15-2025/DACBYA-UNJ/RYLLS, de fecha 14 de marzo de 2025, de la Dra. Rosario Yaqueliny Llauce Santamaria, Carta N° 14-2025/DACBYA-UNJ/RYLLS, de fecha 14 de marzo de 2025, de la Dra. Rosario Yaqueliny Llauce Santamaria, Acuerdo N° 216-2025-SO-CCO-UNJ, de Sesión Ordinaria de Consejo de Comisión Organizadora N° 011-2025-SO-CCO-UNJ, de fecha 20 de marzo de 2025, y;

CONSIDERANDO:

Que, conforme al 4to párrafo del Artículo 18°, de la Constitución Política del Estado, concordante con el Artículo 8°, de la Ley N.° 30220-Ley Universitaria, así como con el Artículo 6° del Estatuto de la Universidad Nacional de Jaén, el Estado reconoce la autonomía Universitaria en su régimen normativo, de gobierno, académico, investigación administrativo y económico;

Que, el Sr. Presidente de la Comisión Organizadora de la Universidad Nacional de Jaén, es el personero y representante legal de la Universidad conforme a lo dispuesto por la Ley Universitaria N° 30220, tiene a su cargo y Dedicación Exclusiva la Dirección, Conducción y Gestión del Gobierno Universitario en todos sus ámbitos. Y de acuerdo al Numeral 6.1.5, literal d) de la Norma Técnica "Disposiciones para la constitución y funcionamiento de las Comisiones Organizadoras de las Universidades Públicas en proceso de Constitución", aprobado mediante Resolución Viceministerial N° 244-2021-MINEDU, modificado por Resolución Viceministerial N° 055-2022-MINEDU, son funciones del Presidente de la Comisión Organizadora, Emitir resoluciones en los ámbitos de su competencia;

Que, mediante numeral 1.2.1. del Artículo 1° del Texto Único Ordenado de la Ley N° 27444 - Ley del Procedimiento Administrativo General, aprobado por Decreto Supremo N° 004-2019-JUS, (en adelante TUO de la LPAG), señala que: "Los actos de administración interna de las entidades destinados a organizar o hacer funcionar sus propias actividades o servicios. Estos actos son regulados por cada entidad, con sujeción a las disposiciones del Título Preliminar de esta Ley, y de aquellas normas que expresamente así lo establezcan";

Que, mediante numeral 1.1, establece sobre el Principio de Legalidad, del Artículo IV, del mismo cuerpo normativo, establece que: "Las autoridades administrativas deben actuar con respeto a la Constitución, a la Ley y al derecho, dentro de las facultades que le estén atribuidas y de acuerdo con los fines para los que les fueron conferidas";

Que, mediante numeral 73.3 del artículo 73° del Texto Único Ordenado de la Ley N° 27444 - Ley del Procedimiento Administrativo General, señala: "Cada Entidad es competente para realizar tareas materiales necesarias para el eficiente cumplimiento de su misión y objetivos;

Que, mediante Carta N° 14-2025/DACBYA-UNJ/RYLLS, de fecha 14 de marzo de 2025, la Dra. Rosario Yaqueliny Llauce Santamaria-Docente Ordinario Auxiliar A Tiempo Completo Adscrita al Departamento Académico de Ciencias Básicas y Aplicadas solicita al Director de Investigación, Innovación y Transferencia Tecnológica, la aprobación mediante acto resolutivo del Manual denominado "Distribución Binomial", adjuntando Carta N° 01-2025-UNJ/VPI-DIITT, Constancia URL, Constancia de Similitud Aceptable, Declaración Jurada de Conflicto de Interés y Material de Enseñanza;

Que, mediante Carta N° 15-2025/DACBYA-UNJ/RYLLS, de fecha 14 de marzo de 2025, la Dra. Rosario Yaqueliny Llauce Santamaria-Docente Ordinario Auxiliar A Tiempo Completo Adscrita al Departamento Académico de Ciencias Básicas y Aplicadas solicita al Director de Investigación, Innovación y Transferencia Tecnológica, la aprobación mediante acto resolutivo del Manual denominado "Análisis de la Varianza con un Factor (Anova)", adjuntando Carta N° 01-2025-UNJ/VPI-DIITT, Constancia de Similitud Aceptable, Declaración Jurada de Conflicto de Interés y Material de Enseñanza;

Que, mediante Carta N° 16-2025/DACBYA-UNJ/RYLLS, de fecha 14 de marzo de 2025, la Dra. Rosario Yaqueliny Llauce Santamaria-Docente Ordinario Auxiliar A Tiempo Completo Adscrita al Departamento Académico de Ciencias Básicas y Aplicadas solicita al Director de Investigación, Innovación y Transferencia Tecnológica, la aprobación mediante acto resolutivo del Manual denominado "Correlación y Regresión Lineal", adjuntando Carta N° 01-2025-UNJ/VPI-DIITT, Constancia URL, Constancia de Similitud Aceptable, Declaración Jurada de Conflicto de Interés y Material de Enseñanza;

Que, mediante Carta N° 17-2025/DACBYA-UNJ/RYLLS, de fecha 14 de marzo de 2025, la Dra. Rosario Yaqueliny Llauce Santamaria-Docente Ordinario Auxiliar A Tiempo Completo Adscrita al Departamento Académico de Ciencias Básicas y Aplicadas solicita al Director de Investigación, Innovación y Transferencia Tecnológica, la aprobación mediante acto resolutivo del Manual denominado "Análisis de Series de Tiempo", adjuntando Carta N° 01-2025-UNJ/VPI-DIITT, Constancia URL, Constancia de Similitud Aceptable, Declaración Jurada de Conflicto de Interés y Material de Enseñanza;

Que, mediante Oficio N° 039-2025-UNJ/VPI-DIITT, de fecha 14 de marzo de 2025, el Director de la Dirección de Investigación, Innovación y Transferencia Tecnológica solicita a la Vicepresidenta de Investigación, la emisión del acto resolutivo de los siguientes Manuales: "Distribución Binomial", "Análisis de la Varianza con un Factor (Anova)", "Correlación y Regresión Lineal" y "Análisis de Series de Tiempo", el cual presentan los siguientes integrantes:

- Dra. Rosario Yaqueliny Llauce Santamaria
- Dra. Marcela Yvone Saldaña Miranda
- Mg. Mario Félix Olivera Aldana

Que, mediante Oficio N° 431-2025-VPI-CO-UNJ, de fecha 17 de marzo de 2025, la Vicepresidenta de Investigación remite al Presidente de la Comisión Organizadora de la Universidad Nacional de Jaén, la solicitud presentada por el Dr. Juan Manuel Garay Román- Director de la Dirección de Investigación, Innovación y Transferencia Tecnológica, a fin de ser considerado en Sesión de Comisión Organizadora, para su posterior emisión de acto resolutivo, previa evaluación, a fin de oficializar los Manuales: "Distribución Binomial", "Análisis de la Varianza con un Factor (Anova)", "Correlación y Regresión Lineal" y "Análisis de Series de Tiempo", presentado por los integrantes, que se detallan en el presente documento;

Que, el pleno del Consejo de Comisión Organizadora UNJ, en Sesión Ordinaria N° 011-2025-SO-CCO-UNJ, de fecha 20 de marzo de 2025, emite el siguiente: Acuerdo N° 216-2025-SO-CCO-UNJ por **UNANIMIDAD**, APROBAR el Manual denominado: "Distribución Binomial", "Análisis de la Varianza con un Factor (Anova)", "Correlación y Regresión Lineal" y "Análisis de Series de Tiempo", presentados por los autores que se indican en la parte resolutiva;

En uso de las facultades y atribuciones conferidas por el Artículo 18°, de la Constitución Política del Perú, la Ley N° 30220-Ley Universitaria, a las "Disposiciones para la Constitución y funcionamiento de las Comisiones Organizadoras de las Universidades Públicas en proceso de Constitución", aprobada mediante RVM N° 244-2021-MINEDU, modificada con RVM N° 055-2022-MINEDU y RVM N° 053-2023-MINEDU, el Estatuto de la Universidad Nacional de Jaén, aprobado mediante Resolución N° 304-2020-CO-UNJ, de fecha 29 de setiembre de 2020, y;

SE RESUELVE:

ARTÍCULO PRIMERO.- APROBAR el **MANUAL** denominado: **"DISTRIBUCIÓN BINOMIAL", "ANÁLISIS DE LA VARIANZA CON UN FACTOR (ANOVA)", "CORRELACIÓN Y REGRESIÓN LINEAL"** y **"ANÁLISIS DE SERIES DE TIEMPO"**, los mismos que en anexo forman parte integrante de la presente resolución, presentado por los siguientes autores, Docentes de la Universidad Nacional de Jaén:

- Dra. Rosario Yaqueliny Llauce Santamaria
- Dra. Marcela Yvone Saldaña Miranda
- Mg. Mario Félix Olivera Aldana

ARTÍCULO SEGUNDO.- NOTIFICAR, a los autores y a las instancias correspondientes para su conocimiento y fines.

ARTÍCULO TERCERO.- DISPONER LA PUBLICACIÓN en el Portal Web Institucional de la Universidad Nacional de Jaén www.unj.edu.pe

REGÍSTRESE, COMUNÍQUESE Y CÚMPLASE;

UNIVERSIDAD NACIONAL DE JAÉN
Abg. Braian Alejandro Max Zegarra
SECRETARIO GENERAL

UNIVERSIDAD NACIONAL DE JAÉN
COMISIÓN ORGANIZADORA
Dr. Severino Apolinar Risco Zapata
PRESIDENTE

UNIVERSIDAD NACIONAL DE JAÉN

UNIVERSIDAD NACIONAL DE JAÉN

DEPARTAMENTO ACADÉMICO DE CIENCIAS BÁSICAS Y APLICADAS

INGENERÍA CIVIL

MANUAL

DISTRIBUCIÓN BINOMIAL

Autores:

Dra. Rosario Yaqueliny Llauce Santamaria

Dra. Marcela Yvone Saldaña Miranda

Mg. Mario Félix Olivera Aldana

Jaén – Perú, febrero 2025

Índice

15

1. Introducción

La probabilidad de éxitos y fallas se mide por la probabilidad discreta, que es la distribución binomial.

Existen numerosos casos en los que es probable que ocurra un evento particular en las e mpresas Esto puede ser exitoso o no exitoso sin superar a un punto medio. El resultado de la fabricación de un artículo puede ser positivo o negativo. El interés no es el factor principal que lo lleva En tales casos, se utiliza la distribución binomial.

El número de éxitos en una secuencia de prueba independiente se modela como una dist ribución de probabilidad discreta, conocida como distribución binomial, con cada proba bilidad de éxito como una constante

La aplicación de la distribución binomial en la Ingeniería civil se puede ver en una variedad de campos, incluida la evaluación de la confiabilidad de la estructura, la determinación de la probabilidad de falla del componente y la optimizació n de los sistemas de mantenimiento.

Por el contrario, la distribución binomial extiende la noción de la distribución de Bernoulli a múltiples experimentos independientes que producen dos resultados distintos ¿Cuántos éxitos hay en una serie de ensayos idénticos e independientes, dando un buen modelo para situaciones donde los experimentos se repiten en condiciones idénticas?

El enfoque del manual estará en las características matemáticas, aplicaciones y fórmulas que definen estas distribuciones, enfatizando su importancia y utilidad en las estadísticas y probabilidades

2. Objetivos

- Entender la teoría, propiedades y aplicaciones de la Distribución binomial en la carrera de Ingeniería
- Aplicar la distribución binomial para medir la fiabilidad de estructuras y sistemas.
- Emplear la Distribución binomial para estimar la probabilidad de falla de componentes y sistemas.

3. Distribución Binomial

La distribución binomial representa la probabilidad de obtener k éxitos en n ensayos de un experimento binomial. Si una variable aleatoria X sigue esta distribución, la probabilidad de que tome el valor k se determina utilizando la siguiente expresión matemática

$$P(X = x) = \binom{n}{x} p^x q^{n-x} \quad ; \quad 0 \leq x \leq n$$

Donde:

- n: número de intentos
- k: número de éxitos
- p: probabilidad de éxito en un ensayo dado
- c_k^n: el número de formas de obtener k éxitos en n intentos

Además:

$$\text{Media } E(X) = np$$

$$\text{Desviación estándar } \sigma = \sqrt{npq}$$

Para ilustrar, consideremos el escenario de lanzar una moneda al aire 10 veces. En este caso, podemos utilizar la distribución binomial para determinar la probabilidad de obtener exactamente 7 caras.

3.1. Propiedades de la Distribución Binomial.

Debe Cumplir con las siguientes propiedades:

- Dos posibles resultados: Imagina que estás lanzando una moneda. La variable aleatoria en este caso podría ser "obtener cara" o "obtener cruz". Estos son los dos resultados posibles en cada lanzamiento.
- Probabilidad constante: Supongamos que la probabilidad de obtener cara en un lanzamiento justo de la moneda es (p = 0.5). Esa probabilidad se mantiene constante en todos los lanzamientos. Siempre tienes la misma posibilidad de éxito.
- Independencia de ensayos: Cada lanzamiento de la moneda es independiente. Si obtuviste cara en el primer lanzamiento, eso no afecta la probabilidad en el segundo lanzamiento. Cada intento es como empezar de cero.
- Parámetro de probabilidad: La distribución de Bernoulli se define por un solo número: la probabilidad de éxito (p). Si conoces esa probabilidad, puedes modelar muchos experimentos binarios diferentes

3.2. Características

Se dice que X sigue una distribución Binomial de parámetros n y p, que se representa con la siguiente notación:

Si $X \sim B\ (n, p)$

$n = número\ de\ ensayos$

$p = probabilidad\ de\ éxito$

3.3. Su función de probabilidad viene definida por:

$$f(x) = P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}\ ,\ \ x = 0,1,2, \ldots\ldots, n$$

Donde, n, debe ser un entero positivo y p debe pertenecer al intervalo $0 \leq p \leq 1$, por ser una proporción.

3.4. Coeficiente binomial:

$$\binom{n}{x} = \frac{n!}{x!\,(n - x)!}$$

Su media, varianza y desviación estándar, vendrán dadas por las siguientes expresiones:

Media o valor esperado

$$\mu = E(X) = np$$

3.5. Varianza

$$\sigma^2 = V(X) = npq$$

3.6. Desviación estándar

$$\sigma = \sqrt{npq} = \sqrt{V(X)}$$

4. Ejemplos

Ejemplo 1

En una planta de energía, se utilizan interruptores automáticos para proteger el equipo contra sobrecargas. La probabilidad de que un interruptor funcione correctamente cuando se activa es del 95%. Un ingeniero selecciona 20 interruptores automáticos al azar para realizar una inspección de calidad

a) ¿Cuál es la probabilidad de que exactamente 18 interruptores funcionen correctamente?

b) ¿Cuál es la probabilidad de que 19 a más interruptores funcionen correctamente?

Desarrollo:

a) ¿Cuál es la probabilidad de que exactamente 18 interruptores funcionen correctamente?

Utilizamos la fórmula y representamos en los datos.

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

$$f(18) = P(X = 18) = \binom{20}{18}(0.95)^{18}(1-0.95)^{20-18}$$

Primeramente, calculamos el coeficiente binomial:

$$\binom{20}{18} = \frac{20!}{18!\,(20-18)!} = 190$$

Ahora, remplazamos y seguimos calculando en la fórmula de distribución binomial.

$$f(18) = P(X = 18) = 190 * (0.95)^{18}(1-0.95)^{20-18}$$

$$= 0.1885$$

Interpretación: La probabilidad de que exactamente 18 interruptores funcionen correctamente es aproximadamente 18.85%

b) ¿Cuál es la probabilidad de que 19 a más interruptores funcionen correctamente?

Para encontrar P $(X \geq 19)$, calculamos P(X=19) y P(X=20) y los sumamos:

Utilizamos la fórmula de distribución binomial:

Calculamos para P(X=19):

$$f(19) = P(X = 19) = \binom{20}{19}(0.95)^{19}(1-0.95)^{20-19}$$

$$f(19) = P(X = 19) = \frac{20!}{(20-19)! * 19!} * (0.95)^{19}(1-0.95)^{1}$$

$$f(19) = P(X = 19) = 0.3774$$

Calculamos para P(X=20):

Remplazamos los datos

$$f(20) = P(X = 20) = \binom{20}{20}(0.95)^{20}(1-0.95)^{20-20}$$

Ahora, remplazamos y seguimos calculando en la fórmula de distribución binomial

$$f(20) = P(X = 20) = \frac{20!}{(20-20)! * 20!} * (0.95)^{20}(1-0.95)^{0}$$

$$f(20) = P(X = 20) = 0.3584$$

Finalmente, encontramos el valor de P $(X \geq 19)$, sumando P(X=19) y P(X=20)

$P\,(X \geq 19) = P\,(X = 19) + P\,(X = 20)$

Remplazamos los valores encontrados en la ecuación

$P\,(X \geq 19) = 0.3774 + 0.3584 = 0.7358$

En porcentaje es:

Interpretación: La probabilidad de que 19 a más interruptores funcionen correctamente es 73.58%

Ejemplo 2

En una fábrica de engranajes, el 90% de los engranajes producidos son de alta calidad y el 10% son defectuosos. Se seleccionan al azar 15 engranajes para inspeccionarlos.

a) ¿Cuál es la probabilidad de que exactamente 12 engranajes sean de alta calidad?
b) ¿encuentra el valor esperado, varianza y desviación estándar?

Desarrollo:

Cada engranaje tiene una probabilidad de p=0.90 de ser de alta calidad y una probabilidad de q=1−p = 0.10 de ser defectuoso

a) ¿Cuál es la probabilidad de que exactamente 12 engranajes sean de alta calidad?

Queremos encontrar $P\ (X = 12)$

$$f(12) = P(X = 12) = \binom{15}{12}(0.90)^{12}(1-0.90)^{3}$$

$$f(12) = P(X = 12) = \frac{15!}{(15-12)!} * (0.90)^{12}(1-0.90)^{3}$$

$$f(12) = 0.1285$$

Interpretación: La probabilidad de que exactamente 12 de los 15 engranajes seleccionados sean de alta calidad es aproximadamente 12.85%.

b) ¿Encuentra el valor esperado, varianza y desviación estándar?

Para n = 15 y p = 0.90, calculamos estos valores:

Asimismo, encontramos media o valor esperado

Utilizamos la formula y remplazamos los datos

$$\mu = E(X) = np$$

$$\mu = E(X) = 15(0.90) = 13.5$$

Interpretación: En promedio, de cada 15 engranajes seleccionados, se espera que 13.5 sean de alta calidad.

Encontramos la varianza:

Utilizamos la fórmula y remplazamos los datos:

$$\sigma^2 = V(X) = npq$$

$$\sigma^2 = V(X) = 15(0.90)(0.10) = 1.35$$

Encontramos la desviación estándar:

Utilizamos la formula y remplazamos los datos

$$\sigma = \sqrt{V(X)}$$

$$\sigma = \sqrt{1.35} = 1.162$$

Ejemplo 3

En una fábrica de equipos topográficos el 5% de los equipos está defectuoso, determinar la probabilidad de que en una muestra de 12 se encuentre 2 equipos topográficos defectuosos.

Desarrollo:

X= n° de equipos topográficos defectuosos r=2

n=12

p=0.05 y q=1-0.05=0.95

$$P(X = 2) = \binom{12}{2}(0.05)^2(0.95)^{10}$$

$$P(X = 2) = \frac{12!}{(12-2)! * 2!} * (0.05)^2(0.95)^{10}$$

$$P(X = 2) = 0.0988$$

$\therefore 0.0988 * 100 = 9.88\%$

Interpretación: La probabilidad que existente en obtener 2 equipos topográficos defectuosos es de 9.88%

Ejemplo 4

Una empresa de construcción utiliza varillas de acero de un proveedor que, según registros históricos, produce un 5% de varillas defectuosas. La empresa realiza un control de calidad seleccionando 10 varillas al azar de cada lote.

a) ¿Cuál es la probabilidad de que exactamente 2 varillas en la muestra sean defectuosas?
b) ¿Cuál es la probabilidad de que menos de 3 varillas en la muestra sean defectuosas?
c) ¿Cuál es la probabilidad de que al menos 1 varilla sea defectuosa en la muestra?

Desarrollo:

Para resolver estas preguntas, aplicamos la fórmula de la Distribución Binomial:

Donde:

- $n = 10$ (tamaño de la muestra)
- $p = 0.05$ (probabilidad de que una varilla sea defectuosa)
- k es el número de varillas defectuosas

a) Probabilidad de que exactamente 2 varillas sean defectuosas.

$$P(X = 2) = \binom{10}{2}(0.05)^2(0.95)^{10-2}$$

$$P(X = 2) = 0.0746$$

Interpretación: La probabilidad de que exactamente 2 varillas sean defectuosas es 0.0746 o 7.46%.

b) Probabilidad de que menos de 3 varillas sean defectuosas

Esto implica la suma de las probabilidades para x = 0,1 y 2

$P\ (X < 3) = P(X = 0) + P(X = 1) + P(X = 2)$

$P\ (X < 3) = 0.9885$

Interpretación: La probabilidad de que menos de 3 varillas sean defectuosas es 0.9885 o 98.85%

c) Probabilidad de que al menos 1 varilla sea defectuosa

Utilizamos la propiedad del complemento:

$P\ (X \geq 1) = 1 - P\ (X = 0)$

$P\ (X \geq 1) = 0.9885$

Interpretación: La Probabilidad de que al menos 1 varilla sea defectuosa es 0.4013 o 40.13%

Ejemplo 5

Un Ingeniero Civil tiene la certeza de trabajar en una empresa es de 0.9 trabajando 3 h/d. Si tiene 3 empresas que lo solicitan trabajar.

a) ¿Cuál es la probabilidad de que no trabaje en ninguna de las empresas?

Sea X la variable aleatoria que denota el número de empresas

$$P(X = 0) = \binom{3}{0}(0.9)^0(0.1)^3$$

$$P(X = 0) = 0.001$$

Interpretación: la probabilidad que no trabaje en ninguna de las empresas es de 0.001

b) ¿Cuál es la probabilidad de que trabaje en dos empresas?

$$P(X = 2) = \binom{3}{2}(0.9)^2(0.1)^1$$

$$P(X = 2) = 0.243$$

Interpretación: la probabilidad que trabaje en dos empresas es de 0.243

8

c) ¿Cuál es la probabilidad de que trabaje en menos de dos empresas?

Aplicamos la condición P (X< 2)

$$P(X = 0) = \binom{3}{0}(0.9)^0(0.1)^3$$
$$P(X = 0) = \mathbf{0.001}$$

$$P(X = 1) = \binom{3}{1}(0.9)^0(0.1)^3$$
$$P(X = 1) = \mathbf{0.027}$$

Sumamos en la condición

P (X< 2) = P(X=0) + P(X=1)

P (X< 2) = 0.001 + 0.027 = 0.028

Interpretación: la probabilidad que trabaje en menos de dos empresas es de 0.028

d) Calcular la media y la varianza

Media o valor esperado

$$\mu = E(X) = np$$
$$\mu = E(X) = 3(0.9) = 0.27$$

Varianza

$$\sigma^2 = V(X) = npq$$
$$\sigma^2 = V(X) = 3(0.9)(0.1) = 0.27$$

Ejemplo 6

40 de 120 colegios de la ciudad de Jaén están con una infraestructura buena, se eligen 9 colegios al azar; Halle la probabilidad de que su infraestructura este en buen estado.

a) Exactamente 4 colegios
b) Más de 6 colegios
c) Calcule la media

Desarrollo

Sea X la variable que denota

n = 9

X: El número de colegios de la ciudad de Jaén que tienen una infraestructura buena

Éxito {infraestructura_ buena}: $p = \frac{40}{120} = 0.33$

7

Fracaso {infraestructura _mala}: $q = 0.67$

a) Exactamente 4 colegios de la ciudad de Jaén que tienen una infraestructura buena (x = 9)

Reemplazamos en la fórmula

$$P(X = 4) = \binom{9}{4}(0.33)^4(0.67)^{9-4}$$

$$P(X = 4) = 0.20$$

Interpretación: La probabilidad de que 4 colegios de un total de 9 tengan infraestructura buena es de 0.20

b) Más de 6 colegios (X> 6) tengan buena infraestructura:

P(X> 6) = P(X=7) +P(X=8) +P(X=9)

$$= \binom{9}{7}(0.33)^7(0.67)^{9-7} + \binom{9}{8}(0.33)^8(0.67)^{9-8} + \binom{9}{9}(0.33)^9(0.67)^{9-9}$$

Interpretación: La probabilidad de que más de 6 colegios de un total de 9 tengan infraestructura buena es de 0.008.

c) Calcule la media

$$\mu = E(X) = 9(0.33) = 2.97$$

P
J
μ

Ejemplo 7

Antes de realizar un estudio de mecánica de suelos por un especialista en Geotecnia, según la Norma Técnica E.070 de Albañilería confinada en una zona de sismicidad moderada (Zona 2) el suelo debe tener una capacidad portante mínima de $1kg/\ cm^2$. Sin embargo, el Ingeniero encargado de obra tiene la certeza de asignar una probabilidad de 0.7 de la capacidad portante mínima, si realiza 8 ensayos breves en el área donde se va a edificar.

a) ¿Cuál es la probabilidad de que salga menos del límite establecido en todos sus ensayos?

Sea x la variable aleatoria que denota el número de asignaturas.

$$P(X = 0) = \binom{8}{0}(0.7)^0(0.3)^8$$

$$P(X = 0) = 0.0001$$

Interpretación: La probabilidad de que salga menos del límite establecido en todos sus ensayos es de 0.0001

b) ¿Cuál es la probabilidad de que apruebe tres de sus ensayos?

6

$$P(X = 3) = \binom{8}{3}(0.7)^3(0.3)^8$$

$$P(X = 3) = 0.0013$$

Interpretación: La probabilidad de que apruebe tres de sus ensayos es de 0.0013

c) ¿Cuál es la probabilidad de que apruebe menos de tres de sus ensayos?

$$P(X \leq 2) = \binom{8}{0}(0.7)^0(0.3)^8 + \binom{8}{1}(0.7)^1(0.3)^7 + \binom{8}{2}(0.7)^2(0.3)^6$$

$$P(X \leq 2) = 0.0113$$

Interpretación: La probabilidad de que apruebe menos de tres de sus ensayos es de 0.0113

d) Calcular la media y la varianza.

Media

$$\mu = E(X) = np$$

$$\mu = E(X) = 8(0.7) = 5.6 \ \cong 6$$

Interpretación: La media es 6 ensayos.

Varianza

$$\sigma^2 = V(X) = npq$$

$$\sigma^2 = V(X) = 8(0.7)(0.3) = 1.68$$

Interpretación: La varianza es 1.68 ensayos.

Ejemplo 8

En un lote de 10 vigas de concreto, la probabilidad de que una viga tenga un defecto de resistencia es 0.1 ¿Cuál es la probabilidad de encontrar exactamente 2 vigas defectuosas en el lote?

Desarrollo

Usamos la distribución binomial, donde n = 10 y p = 0.1, para calcular la probabilidad de que haya exactamente X= 2 vigas defectuosas:

$$P(X = 2) = \binom{10}{2}(0.1)^2(0.9)^8$$

$$P(X = 2) = 0.1937$$

Interpretación: La probabilidad de encontrar exactamente 2 vigas defectuosas en el lote es aproximadamente 0.1937 o 19.37%.

Ejemplo 9

Un proyecto de construcción utiliza tubos de PVC para sistemas de alcantarillado. La probabilidad de que un tubo de PVC tenga defectos después de la instalación es de 0.05. Si se instalan 20 tubos en una sección del proyecto:

a) ¿Cuál es la probabilidad de que exactamente 2 tubos presenten defectos?

$$P(X = 2) = \binom{20}{2}(0.05)^2(0.95)^{18}$$

$$P(X = 2) = 0.189$$

Interpretación: La probabilidad de que exactamente 2 tubos presenten defectos es de 0.189.

b) ¿Cuál es la probabilidad de que más de 3 tubos presenten defectos

$P(X > 3) = 1 - P(X \leq 3)$

$P(X \leq 3) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)$

$$P(X \leq 2) = \binom{20}{0}(0.05)^0(0.95)^{20} + \binom{20}{1}(0.05)^1(0.95)^{19}$$
$$+ \binom{20}{2}(0.05)^2(0.95)^{18} + \binom{20}{3}(0.05)^3(0.95)^{17}$$

$P(X \leq 3) = 0.984$

Entonces $P(X > 3) = 1 - 0.984 = 0.016$

Interpretación: La probabilidad de que más de 3 tubos presenten defectos es de 0.016.

Ejemplo 10

Un ingeniero civil está analizando muestras de suelo para determinar su capacidad portante. Se estima que el 70% de las muestras cumple con los requisitos de residencia. Si el Ingeniero analiza 5 muestras al azar:

a) ¿Cuál es la probabilidad de que ninguna de las muestras cumpla con los requisitos?

Los valores son:

n = 5

X = 0

P = 0.7

q = 0.3

Además, aplicamos la fórmula

$$P(X = 0) = \binom{5}{0}(0.7)^2(0.3)^5$$

$$P(X = 0) = \frac{5!}{(5-0)! * 5!} * (0.7)^2(0.3)^5$$

4

$$P(X = 0) = 0.00243$$

Interpretación: la probabilidad de que ninguna de las muestras cumpla con los requisitos es 0.00243

b) Calcular la media y la varianza del número de muestras que cumplen con los requisitos.

Aplicamos, la Fórmula de la media

$$\mu = E(X) = np$$

$$\mu = E(X) = 5(0.7) = 3.5$$

Asimismo, la Fórmula de la varianza

$$\sigma^2 = V(X) = npq$$

$$\sigma^2 = V(X) = 5(0.7)(0.3) = 1.05$$

Interpretación: la media y la varianza son 3.5 y 1.05 respectivamente

Ejemplo 11

Un ingeniero estructural ha diseñado una serie de vigas de concreto. Se estima que el 85% de estas vigas pueden soportar la carga máxima esperada. Si se seleccionan 4 vigas al azar:

¿Cuál es la probabilidad de que al menos 3 vigas puedan soportar la carga máxima?

$$P(X = X) = \binom{n}{x} P^X q^{n-x}$$

$$P(X \geq 3) = P[X = 3] + P[X = 4]$$

$$P(X \geq 3) = \binom{4}{3} (0.85)^3 (0.15)^1 + \binom{4}{4} (0.85)^4 (0.15)^0$$

$$P(X \geq 3) = \frac{4!}{(4-3)! * 3!} * (0.85)^3 (0.15)^1 + \frac{4!}{(4-4)! * 4!} * (0.85)^4 (0.15)^0$$

$$P(X \geq 3) = 0.368 \quad + \quad 0.522$$

$$P(X \geq 3) = 0.89 \quad \text{ó } 89\%$$

Interpretación: la probabilidad de que al menos 3vigas puedan soportar la carga máxima es 0.89 o 89%

5. Ejemplos propuestos aplicados a la Ingeniería

1. Rudolf Diesel tiene 15 camiones de entrega, que emplea sobre todo para entregar motores y chasis de motocicletas modificados en Alemania, y hace entrega a

3

diferentes destinos. De los 15 camiones, 6 presentan problemas con los frenos. En forma aleatoria se seleccionó una muestra de 5 camiones.

a) ¿Cuál es la probabilidad de que 2 de los camiones probados presenten problemas eléctricos y frenos defectuosos?

b) Entonces dos de los camiones probados presenten problemas eléctricos y frenos defectuosos.

2. En una planta de energía, se utilizan interruptores automáticos para proteger el equipo contra sobrecargas. La probabilidad de que un interruptor funcione correctamente cuando se activa es del 95%. Un ingeniero selecciona 20 interruptores automáticos al azar para realizar una inspección de calidad.

a) ¿Cuál es la probabilidad de que exactamente 18 interruptores funcionen correctamente?

b) ¿Cuál es la probabilidad de que 19 a más interruptores funcionen correctamente?

3. En una fábrica de engranajes, el 90% de los engranajes producidos son de alta calidad y el 10% son defectuosos. Se seleccionan al azar 15 engranajes para inspeccionarlos.

a) ¿Cuál es la probabilidad de que exactamente 12 engranajes sean de alta calidad?

b) ¿encuentra el valor esperado, varianza y desviación estándar?

6. Conclusión

La distribución Binomial es un modelo probabilístico que se utiliza para escribir experimentos con dos posibles resultados, como el éxito y fracaso. Esta distribución asigna probabilidades a eventos discretos donde el éxito se presenta con el valor 1 y el fracaso con el valor 0. Por otro lado, la distribución binomial amplia este concepto al escribir la probabilidad de obtener un numero en específico de éxitos en múltiples repeticiones del experimento Bernoulli. La fórmula de la distribución binomial permite calcular la probabilidad de obtener K éxitos en n intentos, incorporando la probabilidad de éxito en un ensaño dado. Ambas distribuciones son fundamentales en la teoría de la probabilidad y encuentran aplicación en diversos campos para modelar y predecir resultados en situaciones con resultados binarios.

7. Bibliografía.

Devore, J. L. (2009). Probabilidad y estadística para ingeniería y ciencias. *Cengage Learning Editores*

Levin, R. I., & Rubin, D. S. (2004). *Estadística para administración y economía*. Pearson educación.

Martín Pliego, F.J. (2004). Introducción a la Estadística Económica y Empresarial. (Ed.) Thomson. Madrid.

Montiel, A.M.; Rius, F.; Barón F.J. (1997). Elementos básicos de Estadística Económica y Empresarial. (2ª Ed.) Prentice Hall, Madrid.

Peña, D. (2001). Fundamentos de Estadística. (Ed.) Alianza Editorial, S.A. Madrid. ISBN: 84206- 8696-4.

Field, A. (2013). Discovering statistics using IBM SPSS Statistics (4ª ed.). SAGE Publications.

Fisher, R. A. (1925). Statistical methods for research workers. Oliver & Boyd.

Howell, D. C. (2012). Statistical methods for psychology (8ª ed.). Wadsworth.

Montgomery, D. C. (2013). Design and analysis of experiments (8ª ed.). John Wiley & Sons.

Callister, W. D., & Rethwisch, D. G. (2018). *Materials science and engineering: An introduction (10th ed.).* Wiley.

Chen, X., & Chen, Y. (2020). *Energy efficiency and management in industrial applications. Springer.*

Groover, M. P. (2019). *Fundamentals of modern manufacturing: Materials, processes, and systems (7th ed.).* Wiley.

Jardine, A. K., L. D., & Banjevic, D. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing, 20*(7), 1483-1510.

UNIVERSIDAD NACIONAL DE JAÉN

UNIVERSIDAD NACIONAL DE JAÉN

DEPARTAMENTO ACADÉMICO DE CIENCIAS BÁSICAS Y APLICADAS

INGENIERÍA CIVIL

MANUAL

ANÁLISIS DE LA VARIANZA CON UN FACTOR (ANOVA)

Autores:

Dra. Rosario Yaqueliny Llauce Santamaria

Dra. Marcela Yvone Saldaña Miranda

Mg. Mario Félix Olivera Aldana

Jaén – Perú, febrero 2025

Índice

115

1. Introducción

El ANOVA de un solo factor es una herramienta utilizada por los científicos para determinar si existen diferencias significativas entre tres o más grupos. En áreas como la ingeniería, economía, salud, psicología y educación, este método se aplica para comparar tratamientos, condiciones o enfoques con el propósito de evaluar su efectividad y analizar el impacto de ciertas variables sobre una variable específica de interés.

Este análisis distingue dos tipos de variabilidad: la intragrupal, que surge debido a factores aleatorios o no controlados, y la intergrupal, atribuida al factor en estudio. Para determinar si las diferencias entre las medias son estadísticamente significativas, se emplea el estadístico F, permitiendo así extraer conclusiones fundamentadas.

En consecuencia, el ANOVA facilita la identificación de patrones y discrepancias relevantes en grandes volúmenes de datos, optimizando la toma de decisiones. Su aplicación contribuye a mejorar la precisión y eficiencia en la investigación y el análisis estadístico, proporcionando a los investigadores información clave para perfeccionar estrategias y metodologías en sus respectivos campos.

En el contexto del análisis de datos, se generaron representaciones gráficas, como los diagramas de caja (box plots), mediante el uso del lenguaje de programación Python. Estos gráficos fueron seleccionados por su capacidad para resumir visualmente la distribución de los datos, detectar valores atípicos y representar con precisión la variabilidad dentro de cada grupo o conjunto de datos.

2. Objetivos

2.1. Objetivo principal

Ejecute un ANOVA de un factor para detectar diferencias potenciales entre las medias de múltiples grupos, analizado por separado. Además, está diseñado para medir la influencia que un cierto factor puede tener en una variable de interés.

2.2. Objetivo particular

- Examinar el valor F y el valor p con el fin de determinar con precisión la relevancia estadística de los datos y la influencia del factor en cuestión.
- Proporcionar sugerencias al utilizar datos analíticos para fundamentadas para la investigación científica, respecto a cómo aplicar ANOVA de un solo factor

destacando la notabilidad de su uso en la toma de decisiones en áreas clave como la medicina, psicología, economía y educación.

3. Marco teórico

El ANOVA es una herramienta estadística empleada para valorar la presencia de las diferencias entre tres o más conjuntos de datos autónomos (Montgomery, 2013, p. 34). Este método examina el impacto de una variable categórica individual en una variable cuantitativa. Además, ofrece claridad acerca de cómo los diferentes niveles del componente influyen en la variable que está siendo evaluada. El análisis de varianza (ANOVA) es un método de gran magnitud en las consideraciones de decisiones en diferentes contextos, ya que permite desglosar la variabilidad total en distintos componentes explicativos.

Desarrollo cronológico y progreso del ANOVA

El método de ANOVA fue consolidado por Ronald A. Fisher en 1920 y se emplea para evaluar investigaciones en el ámbito agrícola. La técnica en cuestión posibilita a los investigadores la comparación de las disparidades existentes entre los diversos tratamientos empleados en la agricultura. La estrategia desarrollada por Fisher tuvo como objetivo superar las limitaciones de la prueba en estudios multigrupo, lo que posibilitaría la realización de evaluaciones estadísticas más precisas en diversos ámbitos. (Fisher, 1925, p. 12).

Variantes del ANOVA

Este artículo aborda principalmente el ANOVA de un solo factor; sin embargo, es fundamental reconocer las distintas versiones.

- ANOVA de dos vías: evalúa la influencia de dos variables categóricas y sus interacciones potenciales.
- ANOVA de evaluaciones repetidas: se utiliza cuando voluntarios idénticos se someten a muchas pruebas en entornos variados.
- El MANOVA (análisis de varianza multivariante) se utiliza cuando hay muchas variables dependientes presentes.

Fundamentos del ANOVA

La variabilidad total de las muestras se descompone en dos componentes básicos: variabilidad intergrupal y variabilidad intragrupal.

El término "variabilidad intergrupal" se refiere a los cambios que se producen por el efecto del componente que se está investigando. El componente que se está investigando es responsable de la irregularidad que se aprecia en las medias del grupo si las medias del grupo muestran diferencias significativas.

Cuando hablamos de variabilidad intragrupal, nos referimos a las diferencias individuales que se observan dentro de cada grupo como resultado de eventos aleatorios o incontrolables.

Determinar si las diferencias que se han observado son estadísticamente significativas o no utilizando la estadística F, por lo tanto, se realiza un análisis de la irregularidad que existe entre los grupos en balance con la irregularidad que existe dentro de los grupos. (Montgomery,2013,p.45).

Concepto e interpretación del estadístico F

El estadístico F se utiliza para analizar el vínculo entre la varianza explicada por el factor y la varianza residual en un estudio estadístico. Asimismo,la distribución F presenta asimetría, cuya configuración se define por los grados de libertad dentro y fuera de los mismos. La negación de la hipótesis nula se produce debido a un alto valor de la estadística F y un valor p por debajo del nivel de significancia (por ejemplo, $\alpha = 0{,}05$), señala que al menos un grupo presenta una diferencia significativa. (Tabachnick & Fidell, 2013, p. 78).

$$F = \frac{\frac{\mathrm{SSB}}{\mathrm{dfB}}}{\frac{\mathrm{SSW}}{\mathrm{dfW}}} \quad O \quad F = \frac{\mathrm{MSB}}{\mathrm{MSW}}$$

- **La sumatoria de los cuadrados entre grupos (SSB)**: mide las irregularidades de las medias de los grupos en relación con la media.
- **El número de grados de libertad entre grupos (dfB)**: evalúa g-1, donde g representa el número de grupos en el estudio.
- **Sumatoria de los cuadrados dentro de los grupos (SSW)**: evalúa la irregularidad de los datos dentro de cada grupo en comparación con la media del grupo.

- **El número de grados de libertad dentro de los grupos (dfW)**: evalúa la diferencia entre el total de muestras y el número de grupos (N - g), donde N es el total de muestras y g es el número de grupos.
- $\boldsymbol{MSB} = \frac{SSB}{dfB}$ forma abreviada de definir promedio de la variabilidad presente entre de los grupos.
- $\boldsymbol{MSW} = \frac{SSW}{\mathrm{dfW}}$ forma abreviada de definir promedio de la variabilidad presente dentro de los grupos.

4. Tabla de Anova

El análisis de varianza (ANOVA) se presenta de manera organizada en una estructura conocida, donde se detallan los cálculos principales y los resultados de la prueba. Esta tabla incluye las siguientes fórmulas:

Variación	Sumatoria de cuadrados	Grados de libertad	Promedio de los cuadrados	F
Factor	$SS_B = \sum_{k=1}^{g} (n_k * (X_k - \bar{X})^2)$	g - 1	$MS_B = \frac{SS_B}{(g-1)}$	$F = \frac{MS_B}{MS_W}$
Error	$SSE = \sum_{k=1}^{g} \sum_{i=1}^{nk} (\mathrm{X_i k} - \bar{\mathrm{X}}_\mathrm{k})^2$	N - g	$MS_W = \frac{SS_W}{(N-g)}$	
Total	$SST = \sum_{i=1}^{N} (\mathrm{X_i} - \bar{\mathrm{X}})^2$	N - 1		

Donde:

g: Número de grupos

n_k:Tamaño del grupo

X_k: Promedio aritmético de grupo

$\bar{X}$: Promedio aritmético general

4.1.Concepto e Interpretación del Valor p

El valor de p se define como la posibilidad de conseguir un estadístico de prueba similar de extremo o dentro del margen de lo observado en los datos de muestra, bajo la premisa de que la hipótesis nula (H_0) es verdadera.

Por consiguiente, es interpretado en propósito del nivel de significancia (α) y sirve como criterio para concluir si se aprueba o niega la hipótesis nula en un estudio. En investigaciones se requieren un mayor nivel de precisión estadística, es factible observar niveles de significancia inferiores al estándar comúnmente aceptado de 0.05.

Si $(p \leq \alpha)$

- Rechazamos (H_0).
- Se dispone de pruebas suficientes para poder negar la hipótesis nula.

Si $(p > \alpha)$

- No rechazamos (H_0).
- No se dispone de pruebas suficientes para admitir la existencia de una disparidad o impacto de relevancia.

Fórmula para el valor p:

$$p = P(F \geq Fobservado \mid dfB, dfW)$$

Donde:

- F es una variable que se distribuye de acuerdo a grados de libertad específicos para cada grupo de datos. Esta distribución es ampliamente empleada en el análisis de varianza con el fin de contrastar la variabilidad existente entre dos o más muestras.
- F observado es el valor calculado del estadístico F para los datos observados.
- dfB (grados de libertad entre los grupos): es el número de grupos menos se representa como: g−1, donde g es el número de grupos.
- dfW (grados de libertad dentro de los grupos): es la diferencia entre el total de muestras y el total de grupos se representa como N−g, donde N denota el total de muestras y g el total de grupos.

4.2. Distribución de F de FICHER

Esta tabla contiene valores críticos para la distribución F definida por

$P(F \geq F_{\alpha,x1,x2})$

$x1/_{x2}$	1	2	3	4	5	6	7	8	9	10	15	20
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88	245.95	248.01
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.43	19.46
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.70	8.66

4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.86	5.80
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.62	4.56
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	3.94	3.87
7	5.99	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.51	3.44
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.22	3.15
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.01	2.94
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.85	2.77
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.72	2.65
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.62	2.54
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.53	2.46
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.46	2.39
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.40	2.33
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.35	2.28
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.31	2.23
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.27	2.19
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.20	2.12
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.47	2.32	2.18	2.10
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.15	2.07
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.13	2.05
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.46	2.30	2.25	2.11	2.03
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.09	2.01
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.01	1.93
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	1.92	1.84
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.03	1.87	1.78
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.84	1.75
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.75	1.66
∞	3.85	3.00	2.61	2.38	2.22	2.10	2.01	1.94	1.88	1.84	1.67	1.58

La tabla muestra una pequeña parte de la tabla de distribución F de Fisher si necesita más valores puede visitar a distintos autores que proporcionen información más detallada.

5. Aplicaciones del ANOVA de un solo factor

A continuación, se enumeran algunos de los ámbitos que hacen un uso intensivo de este método (Field, 2013, p. 120):

- En el ámbito de la psicología, durante la realización de estudios para valorar la eficiencia de diversas terapias o intervenciones.
- En el ámbito de la medicina, se realizaría para evaluar la eficacia de diversas terapias o productos farmacéuticos.
- En el ámbito de la economía, el proceso de analizar los efectos de las políticas económicas en una serie de lugares o sectores diferentes.

5.1. Proceso del ANOVA de un solo factor

- La hipótesis nula (H_0) indica que no existe diferencias entre los grupos.

- La hipótesis alternativa (H_1) indica que existe como mínimo una diferencia entre muestras analizadas.

5.2. Análisis post hoc

Esta prueba es práctica para identificar discrepancias entre las medias. La prueba de Bonferroni es recomendable en situaciones en las que se pretende minimizar la posibilidad de producir un error de tipo I (Navarro, 2019, p. 145).

5.3. La prueba de Tukey

Es utilizado para reconocer cual grupos presenta discrepancias en sus medias.

También conocida como HSD, por sus siglas en inglés: Honest Significant Difference, se calcula siguiendo esta fórmula:

$$HSD = q \times \sqrt{\frac{MSE}{n}}$$

donde:

- q : Valor crítico de la distribución studentizada
- MSE : Cuadrado medio dentro de los grupos (del ANOVA).
- n : tamaño de los grupos.

5.4. Análisis de gráficos pertinentes para ANOVA de un solo factor

Para interpretar los resultados de un ANOVA de un solo factor, resulta fundamental realizar un análisis visual utilizando gráficos apropiados. Los diagramas de caja (box plots) se utilizan con frecuencia por su efectividad para visualizar la distribución de los datos y las diferencias entre los grupos analizados

6. Gráficos de diagnóstico

Para respaldar las suposiciones del ANOVA, se recomienda el uso de los siguientes gráficos:

- Diagrama de caja: Esta herramienta es ideal para detectar valores atípicos que podrían influir en los resultados del análisis.

La generación de estos gráficos puede realizarse mediante bibliotecas de software como Python (utilizando matplotlib y seaborn) o SPSS, lo que facilita la evaluación de los supuestos del ANOVA de manera más efectiva. (Howell, 2012).

7. Restricciones del ANOVA de un solo factor
    - Homogeneidad: Las varianzas entre grupos deben ser similares.
    - Las muestras deben ser autónomas entre ellas.

Las conclusiones pueden verse comprometidas si no se cumplen los criterios, lo que res ulta en resultados poco confiables La prueba Kruskal walis es una prueba no paramétrica que se utiliza en otros casos sin la necesidad de la normalidad d datos o la homogeneidad de las variaciones (Navarro, 2019).

8. Ejemplos

Ejemplo 1

Comparación de la Resistencia de Dos Materiales

Un ingeniero desea analizar la resistencia a la compresión de tres variedades de concreto para establecer si existen diferencias significativas entre ellas. Para ello, se recopilan datos de resistencia a partir de múltiples muestras de cada tipo de concreto.

Tipo de concreto	Muestra 1	Muestra 2	Muestra 3	Muestra 4	Muestra 5
Convencional	25	27	24	28	26
Estructural	30	31	29	32	30
Permeable	35	36	34	37	36

R
A
A

Paso 1: Calcular el promedio para cada grupo.

Media de A: $\frac{25+27+24+28+26}{5} = 26$

Media de B: $\frac{30+31+29+32+30}{5} = 30.4$

Media de C: $\frac{35+36+34+37+36}{5} = 35.6$

Media General (MG):

$$MG = \frac{25+27+24+28+26+30+31+29+32+30+35+36+34+37+36}{15}$$

$$MG = 30.67$$

Paso 2: Calcular la sumatoria de los cuadrados entre los grupos (SSB)

Utilizamos la fórmula:

$$SSB = n \times \sum (Mediadelgrupo - Mediageneral)^2$$

Para cada grupo:

Tipo A: $(26.0 - 30.67)^2 = 21.79$

Tipo B: $(30.4 - 30.67)^2 = 0.0729$

Tipo C: $(35.6 - 30.67)^2 = 24.30$

Calculamos SSB sin redondeos:

$$SSB = 5 \times (21.79 + 0.0729 + 24.30) = 5 \times 46.163 = 230.815$$

Paso 3: Calcular la sumatoria de los cuadrados dentro de los grupos (SSW)

Usamos la fórmula:

$$SSW = \sum (valorindividual - Mediadelgrupo)^2$$

Para el Tipo A:

$$(25 - 26)^2 + (27 - 26)^2 + (24 - 26)^2 + (28 - 26)^2 + (26 - 26)^2 = 10$$

Para el Tipo B:

$$(30 - 30.4)^2 + (31 - 30.4)^2 + (29 - 30.4)^2 + (32 - 30.4)^2 + (30 - 30.4)^2 = 4.8$$

Para el Tipo C:

$$(35 - 35.6)^2 + (36 - 35.6)^2 + (34 - 35.6)^2 + (37 - 35.6)^2 + (36 - 35.6)^2 = 5.6$$

Calculamos SSW sin redondeos:

$$SSW = 10 + 4.8 + 5.6 = 20.4$$

Paso 4: Calcular los Grados de Libertad (df)

Entre los grupos (dfB): g−1=3−1=2

Dentro de los grupos (dfW): N−g=15−3=12

Paso 5: Calcular MSB y MSW (Media de los Cuadrados)

MSB (Media de los cuadrados entre grupos):

$$MSB = \frac{SSB}{dfB} = \frac{230.815}{2} = 115.4075$$

MSW (Media de los cuadrados dentro de los grupos):

$$MSW = \frac{SSW}{dfW} = \frac{20.4}{12} = 1.7$$

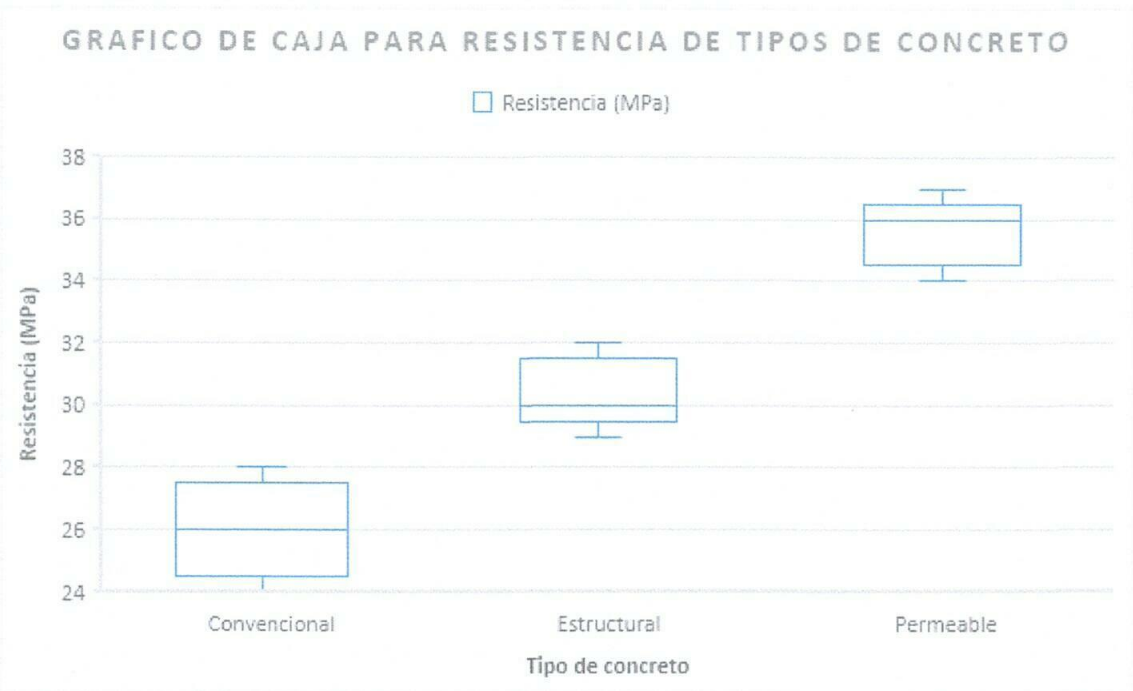
Paso 6: Calcular el Valor F

$$F = \frac{MSB}{MSW} = \frac{115.4075}{1.7} = 67.887$$

$$Valor\ F: 67.887$$

Interpretación: El valor F calculado es 67.887 es muy alto, indica que existe al menos una diferencia en la resistencia a la compresión entre los tipos de concreto. Asimismo, se confirma con un p-valor que será mucho menor que 0.05, obteniendo suficiente evidencia para negar la hipótesis nula.

Gráfico N°1: Box plot de resistencia de tipo de concreto



105

Fuente: Elaboración Propia.

Interpretación de la gráfica:

- Tipo C tiene la mayor resistencia, seguido por Tipo B y luego Tipo A.
- Tipo A presenta menos variabilidad en la resistencia, mientras que Tipos B y C tienen más dispersión.
- En general, Tipo C es más adecuado para proyectos que requieren alta resistencia, mientras que Tipo A podría usarse en aplicaciones menos exigentes.

Ejemplo 2:

Se quiere evaluar el efecto de tres tipos de cemento Pacasmayo (A), Inca (B), Mochica (C) en la resistencia a compresión de muestras de concreto. Los datos recopilados son:

Paso 1: Hipótesis

- **Nula (H_0)**: No hay diferencia en la resistencia a compresión entre los tipos de cemento Pacasmayo (A), Inca (B), Mochica (C).
- **Alternativa (H_1)**: Al menos un tipo de cemento tiene una resistencia significativamente diferente.

Paso 2: Datos

Los datos de resistencia a compresión para cada tipo de cemento están organizados en la tabla siguiente:

Tipo de Cemento	Muestra 1	Muestra 2	Muestra 3	Muestra 4	Muestra 5
Pacasmayo (A)	25	26	24	25	27
Inca (B)	30	31	29	32	30
Mochica (C)	35	36	34	35	36

Paso 3: Cálculo de la media general

$$\overline{X} = \frac{(25+26+24+25+27)+(30+31+29+32+30)+(35+36+34+35+36)}{15}$$

$$\overline{X} = 30.33$$

Paso 4: Cálculo de la suma de cuadrados

Suma de cuadrados total (SST)

$$SST = \sum_{i=1}^{N} \left(X_i - \overline{X}\right)^2$$

$$SST = (25-30.33)^2 + (26-30.33)^2 + (24-30.33)^2 + \cdots + (36-30.33)^2$$

Después de realizar los cálculos para cada observación, obtenemos:

$$SST = 253.33$$

Suma de cuadrados entre grupos (SStr)

$$SStr = \sum_{j=1}^{k} n_j \left(\overline{X}_j - \overline{X}\right)^2$$

Donde:

- n_j número de muestras en cada grupo.
- $\overline{X}_j$ media de cada grupo.
- $\overline{X}$ media general.

Para cada tipo de cemento, tenemos las medias:

- **Media de A** = 25.4
- **Media de B** = 30.4
- **Media de C** = 35.2

Aplicamos la fórmula:

$$SStr = 5 \times (25.4-30.33)^2 + 5 \times (30.4-30.33)^2 + 5 \times (35.2-30.33)^2$$

Realizando estos cálculos:

$$SSStr = 240.13$$

Sumatoria de cuadrados dentro de grupos (SSE)

Se obtiene mediante el siguiente cálculo:

$$SSE = SST - SStr$$

Sustituyendo los valores:

$$SSE = 253.33 - 240.13 = 13.20$$

Paso 5: Cálculo de los Grados de Libertad

- Grados de libertad entre grupos $(df_b) = g - 1 = 3 - 1 = 2$
- Grados de libertad dentro de los grupos $(df_w) = N - g = 15 - 3 = 12$
- Grados de libertad total $(df_t) = N - 1 = 15 - 1 = 14$

Paso 6: Cálculo de las medias cuadradas y el estadístico F

Media cuadrada entre grupos (MStr)

Se calcula como:

$$MStr = \frac{SStr}{df_b}$$

Sustituyendo los valores:

$$MStr = \frac{240.13}{2} = 120.07$$

Media Cuadrada dentro de los grupos (MSE)

Se calcula como:

$$MSE = \frac{SSE}{df_w}$$

Sustituyendo los valores:

$$MSE = \frac{13.20}{12} = 1.10$$

Estadístico F

Se obtiene como:

$$MSEF = \frac{MStr}{MSE}$$

Sustituyendo los valores:

$$F = \frac{120.07}{1.10} = 109.15$$

Paso 7: Interpretación

Con un valor de 109.15 para la prueba F, se observa que este valor es significativamente elevado. El valor observado de F supera el valor crítico para el nivel de significancia (α=0.05), por lo cual se tiene evidencia suficiente para negar la hipótesis nula, concluyendo que existe por lo menos una diferencia en la resistencia a compresión entre los tipos de cemento (Pacasmayo, Inca, Mochica),

Gráfico N°2: Gráfico de caja de la resistencia del cemento

101

Fuente: Elaboración Propia.

Ejemplo 3:

Analizar los métodos de compactación: Placa Vibratoria (X), Modificado Proctor (Y) y sismógrafo (Z) en la densidad del suelo.

Paso 1: Hipótesis

Nula (H_0): No hay diferencia en la densidad del suelo entre los métodos de compactación (Anillo de densidad, Densidad aparente y Proctor).

Alternativa (H_1): Al menos un método de compactación produce una densidad significativamente diferente.

Paso N°2: Datos

Los datos de densidad del suelo por cada tipo métodos de compactación están organizados en la tabla siguiente:

Método	Muestra 1	Muestra 2	Muestra 3	Muestra 4	Muestra 5
Placa Vibratoria (X)	1.85	1.87	1.86	1.84	1.88
Modificado Proctor (Y)	1.90	1.92	1.89	1.91	1.90
Proctor (Z)	1.95	1.97	1.96	1.94	1.95

Paso 3: Cálculo de la media general

$$\overline{X} = \frac{(1.90 + 1.92 + 1.89 + 1.91 + 1.90) + (1.95 + 1.97 + 1.96 + 1.94 + 1.95)}{15} + \frac{(1.85 + 1.87 + 1.86 + 1.84 + 1.88)}{15}$$

$$\overline{X} = 1.906$$

Paso 4: Cálculo de las sumas de cuadrados

Suma de cuadrados total (SST)

Se estima mediante el siguiente cálculo:

100

$$SST = \sum_{i=1}^{N} \left(X_i - \overline{X}\right)^2$$

$$SST = (1.85 - 1.906)^2 + (1.87 - 1.906)^2 + (1.86 - 1.906)^2 + \cdots + (1.95 - 1.906)^2$$

Luego de realizar estos cálculos obtenemos:

$$SST = 0.02416$$

Suma de cuadrados entre grupos (SStr)

El cálculo se realiza de la siguiente manera:

$$SStr = \sum_{j=1}^{k} n_j \left(\overline{X}_j - \overline{X}\right)^2$$

Donde:

- n_j número de muestras de cada grupo.
- $\overline{X}_j$ media de cada grupo.
- $\overline{X}$ media general.

Calculamos la media de cada método:

- **Media de X** = 1.86
- **Media de Y** = 1.90
- **Media de Z** = 1.95

Luego, aplicamos la fórmula:

$$SStr = 5 \times (1.86 - 1.906)^2 + 5 \times (1.90 - 1.906)^2 + 5 \times (1.95 - 1.906)^2$$

Después de realizar estos cálculos, obtenemos:

$$SStr = 0.02212$$

Suma de cuadrados dentro de los grupos (SSE)

Se obtiene mediante el siguiente cálculo:

$$SSE = SST - SStr$$

Sustituyendo los valores que calculamos anteriormente:

$$SSE = 0.02416 - 0.02212 = 0.00204$$

Paso 5: Cálculo de los grados de libertad

- Grados de libertad Entre Tratamientos $(\boldsymbol{df_b}) = g - 1 = 3 - 1 = 2$
- Grados de libertad del Error $(\boldsymbol{df_w}) = N - g = 15 - 3 = 12$
- Grados de libertad Total $(\boldsymbol{df_t}) = N - 1 = 15 - 1 = 14$

Paso 6: Cálculo de las medias cuadradas y el estadístico F

Media cuadrada entre gruposs (MStr)

Se calcula como:

$$MStr = \frac{SStr}{df_b}$$

Sustituyendo los valores:

$$MStr = \frac{0.02212}{2} = 0.01106$$

Media cuadrada dentro del grupo (MSE)

Se calcula como:

$$MSE = \frac{SSE}{df_{within}}$$

Sustituyendo los valores:

$$MSE = \frac{0.00204}{12} = 0.00017$$

Estadístico F

Se calcula como:

$$F = \frac{MStr}{MSE}$$

Sustituyendo los valores:

$$F = \frac{0.01106}{0.00017} = 65.06$$

Paso 7: Interpretación

Con un valor de 65.06 para la prueba F, se observa que este valor es significativamente elevado. El valor observado de la estadística F supera el valor crítico para un nivel de significancia (α=0.05), por lo cual se tiene evidencia suficiente para negar la hipótesis nula, lo que se concluye que existe al menos una diferencia entre densidades del suelo entre los distintos métodos de compactación (X, Y, Z).

Gráfico N°3: Gráfico de caja de Compactación en la densidad de suelos

Fuente: Elaboración Propia.

Ejemplo 4:

Evaluar el efecto de tres tipos de agregados :Arena de trituración (D), Arena de sílice (E), Polvo de piedra (F) en la durabilidad del concreto.

Hipótesis:

H_0: No hay diferencia en la durabilidad del concreto entre los tipos de agregados.

97

H₁: Al menos un tipo de agregado produce una durabilidad significativamente diferente.

Paso 2: Datos

Los datos de la durabilidad del concreto para cada tipo de agregados del concreto están organizados en la tabla siguiente:

Agregados	Muestra 1	Muestra 2	Muestra 3	Muestra 4	Muestra 5
Arena de trituración (D)	100	102	98	101	99
Arena de sílice (E)	110	112	109	111	110
Polvo de piedra (F)	120	121	119	122	120

Paso 3: Cálculo de la media general

$$\overline{X} = \frac{(110+112+109+111+110)+(120+121+119+122+120)}{15} + \frac{(100+102+98+101+99)}{15}$$

$$\overline{X} = 110.27$$

Paso 4: Cálculo de las sumas de cuadrados

Suma de cuadrados total (SST)

La fórmula es la siguiente:

$$SST = \sum_{i=1}^{N} (X_i - \overline{X})^2$$

Sustituyendo los valores:

$$SST = (100-110.27)^2 + (102-110.27)^2 + (98-110.27)^2 + \cdots + (120-110.27)^2$$

Realizando los cálculos, obtenemos:

$$SST = 1060.93$$

Suma de cuadrados entre grupos (SStr)

La fórmula utilizada para su cálculo es la siguiente:

$$SStr = \sum_{j=1}^{k} n_j \left(\overline{X}_j - \overline{X}\right)^2$$

Donde:

- jn_j número de muestras en cada grupo.
- $\overline{X}_j$ media de cada grupo.
- $\overline{X}$ media general.

Usando las medias de cada tipo de agregado:

- **Media de D** = 100.0
- **Media de E** = 110.2
- **Media de F** = 120.6

Realizando los cálculos:

$$SStr = 1040.53$$

Suma de cuadrados dentro de grupos (SSE)

Se obtiene mediante el siguiente cálculo:

$$SSE = SST - SStr$$

Sustituyendo los valores:

$$SSE = 1060.93 - 1040.53 = 20.40$$

Paso 5: Cálculo de los grados de libertad

- Grados de libertad Entre Tratamientos ($\boldsymbol{df_b}$) = g−1=3−1=2
- Grados de libertad del Error ($\boldsymbol{df_w}$) = N−g=15−3=12
- Grados de libertad Total ($\boldsymbol{df_t}$) = N−1=15−1=14

Paso 6: Cálculo de las medias cuadradas y el estadístico F

Media cuadrada entre grupos (MStr)

Se calcula como:

$$MStr = \frac{SStr}{\boldsymbol{df_b}}$$

Aplicamos en la fórmula:

$$SStr = 5 \times (100.0 - 110.27)^2 + 5 \times (110.2 - 110.27)^2 + 5 \times (120.6 - 110.27)^2$$

$$MStr = \frac{SStr}{\boldsymbol{df_b}}$$

Sustituyendo los valores:

$$MStr = \frac{1040.53}{2} = 520.27$$

Media Cuadrada dentro de los grupos (MSE)

Se calcula como:

$$MSE = \frac{SSE}{df_w}$$

Sustituyendo los valores:

$$MSE = \frac{20.40}{12} = 1.70$$

Estadístico F

Se obtiene como:

$$F = \frac{MStr}{MSE}$$

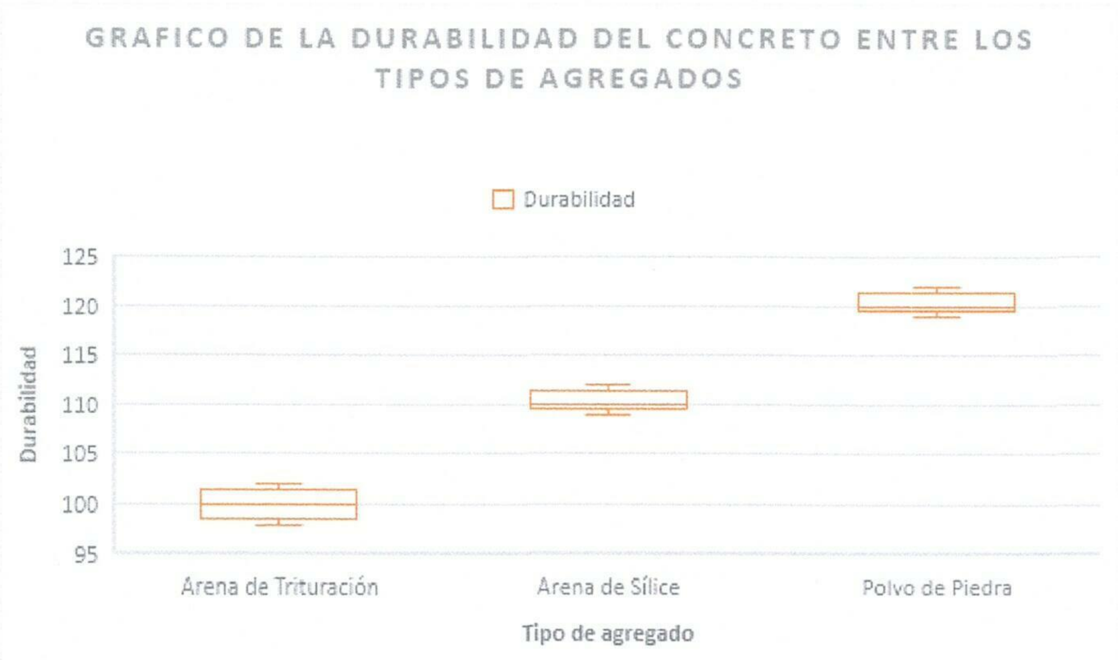
Sustituyendo los valores:

$$F = \frac{520.27}{1.70} = 306.04$$

Paso 7: Interpretación

El valor de F, que asciende a 306.04, se considera significativamente elevado. El valor observado de la estadística F supera el valor crítico para un nivel de significancia ($\alpha = 0.05$), por lo cual e tiene suficiente evidencia para rechazar la hipótesis nula, concluyendo que existe al menos una diferencia en la resistencia del concreto según el tipo de agregados utilizados (D, E, F).

Gráfico N°4: Gráfico de caja de la durabilidad del concreto entre los tipos de agregados



Fuente: Elaboración Propia

Ejemplo 5:

Analizar el efecto de tres condiciones climáticas Frío y Húmedo (A), Cálido y Húmedo (B), Costero (C) en el tiempo de fraguado del concreto.

Hipótesis:

H_0: No hay diferencia en el tiempo de fraguado del concreto entre las condiciones climáticas: Frío y Húmedo (A), Cálido y Húmedo (B), Costero (C) .

H_1: Al menos una de las condiciones climáticas produce un tiempo de fraguado significativamente diferente.

Paso 2: Datos

Los datos del tiempo de fraguado del concreto para cada tipo de condiciones climáticas están organizados en la tabla siguiente:

Clima	Muestra 1	Muestra 2	Muestra 3	Muestra 4	Muestra 5
Frío y Húmedo (A),	180	182	178	181	179
Cálido y Húmedo (B),	150	152	148	151	150
Costero (C)	120	122	119	121	120

Paso 3: Cálculo de la media general

$$\overline{X} = \frac{+(150 + 152 + 148 + 151 + 150) + (120 + 122 + 119 + 121 + 120)}{15} + \frac{(180 + 182 + 178 + 181 + 179)}{15}$$

$$\overline{X} = 150.2$$

Paso 4: Cálculo de las sumas de cuadrados

Suma de cuadrados total (SST)

Se calcula como:

$$SST = \sum_{i=1}^{N} (X_i - \overline{X})^2$$

$$SST = (180 - 150.2)^2 + (182 - 150.2)^2 + (178 - 150.2)^2 + \cdots + (120 - 150.2)^2$$

Después de realizar los cálculos para cada observación, obtenemos:

$$SST = 8904.4SST \ = \ 8904.4SST = 8904.4$$

Suma de cuadrados entre grupos (SStr)

$$SStr = \sum_{j=1}^{k} n_j(\overline{X}_j - \overline{X})^2$$

Donde:

- n_j número de muestras en cada grupo.
- $\overline{X}_j$ media de cada grupo.
- $\overline{X}$ media general.

Para cada clima, tenemos las medias:

- **Media de A** = 180.0
- **Media de B** = 150.2
- **Media de C** = 120.4

Aplicamos la fórmula:

$$SStr = 5 \times (180.0 - 150.2)^2 + 5 \times (150.2 - 150.2)^2 + 5 \times (120.4 - 150.2)^2$$

Realizando estos cálculos:

$$SStr = 8880.4SStr\ =\ 8880.4SStr = 8880.4$$

Suma de cuadrados del error (SSE)

Se obtiene mediante el siguiente cálculo:

$$SSE\ =\ SST\ -\ SStr$$

Sustituyendo los valores:

$$SSE = 8904.4 - 8880.4 = 24.0$$

Paso 5: Cálculo de los grados de libertad

- Grados de libertad Entre Tratamientos $(\boldsymbol{df_b}) =\ g - 1 = 3 - 1 = 2$
- Grados de libertad del Error $(\boldsymbol{df_w}) =\ N - g = 15 - 3 = 12$
- Grados de libertad Total $(\boldsymbol{df_t}) = N - 1 = 15 - 1 = 14$

Paso 6: Cálculo de las medias cuadradas y el estadístico F

Media cuadrada entre grupos (MStr)

$$MStr = \frac{SStr}{df_b}$$

Sustituyendo los valores:

$$MStr = \frac{8880.4}{2} = 4440.2$$

Media cuadrada dentro de los grupos (MSE)

Se obtiene como:

$$MSE = \frac{SSE}{df_w}$$

Sustituyendo los valores:

$$MSE = \frac{24.0}{12} = 2.0$$

Estadístico F

Se obtiene como:

$$MSEF = \frac{MStr}{MSE}$$

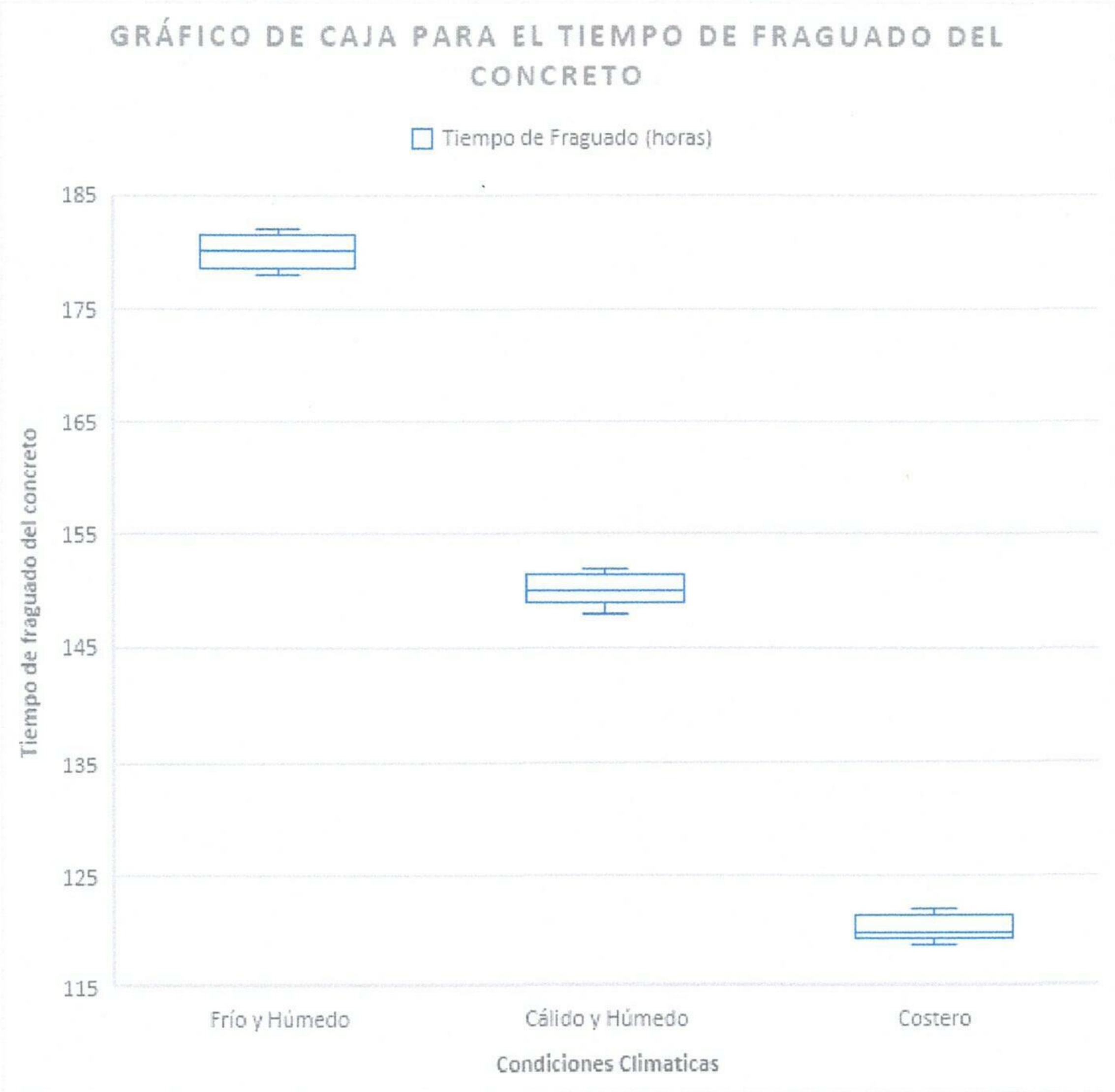
Sustituyendo los valores:

$$F = \frac{4440.2}{2.0} = 2220.1$$

Paso 7: Interpretación

Con un valor de 2220.1 para la prueba F, se puede afirmar que este valor es significativamente elevado. El valor observado de la estadística F supera el valor crítico correspondiente para un nivel de significancia ($\alpha = 0.05$), con ello se tiene suficiente prueba para negar la hipótesis nula concluyendo que existe en al menos un grupo diferencia en el tiempo de fraguado del concreto bajo distintas condiciones climáticas (A, B, C).

Gráfico N°5: Gráfico de caja para el Tiempo de Fraguado del Concreto



Fuente: Elaboración Propia.

9. Conclusiones

- Este estudio comprende plenamente los conceptos teóricos del ANOVA de un solo factor), con miras a la eficiencia de este enfoque estadístico para detectar variaciones notables entre grupos independientes. Demostrando el poderoso instrumento estadístico para examinar el efecto de una sola variable categórica en un resultado cuantitativo, los investigadores pueden obtener resultados más exactos tanto en estudios experimentales como observacionales.

10. Recomendaciones

- Los modelos mixtos, son recomendados para el análisis de datos con estructuras jerárquicas o mediciones repetidas. Al combinar ambos tipos de efectos, estos modelos ofrecen un enfoque más flexible y equilibrado.
- Cuando los datos no cumplen con los supuestos de normalidad y homogeneidad de varianzas requeridos para el ANOVA, se sugiere utilizar métodos alternativos como el MANOVA, que permite analizar múltiples variables independientes.

11. La creación de scripts en Python con el propósito de analizar ANOVA de un solo factor posee la capacidad de incrementar significativamente tanto la eficacia como la reproducibilidad de investigaciones subsiguientes. Códigos Utilizados en Python para el Análisis ANOVA

Para ejecutar códigos es necesario tener instalado Python y una serie de bibliotecas:

1. Instalar `numpy` para trabajar con arrays y funciones matemáticas:

```
pip install numpy
```

2. Instalar `pandas` para la manipulación y análisis de datos:

```
pip install pandas
```

3. Instalar `matplotlib` para la visualización de gráficos:

```
pip install matplotlib
```

4. Instalar `statsmodels` para realizar el análisis ANOVA:

```
pip install statsmodels
```

Código del programa

```
import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import statsmodels.api as sm

from statsmodels.formula.api import ols
```

```
def ingresar_datos():
    """Función para que el usuario ingrese datos para varios grupos."""
    grupos = {}
    print("Ingrese los datos para cada grupo. Escriba 'fin' cuando haya terminado.")
    while True:
        nombre_grupo = input("\nNombre del grupo (o 'fin' para terminar): ")
        if nombre_grupo.lower() == 'fin':
            break
        datos = input(f"Ingrese los valores para {nombre_grupo} separados por comas: ")
        try:
            grupos[nombre_grupo] = list(map(float, datos.split(',')))
        except ValueError:
            print("Error: asegúrese de ingresar números separados por comas.")
    return grupos
def realizar_anova(grupos):
    """Función para calcular la ANOVA y mostrar los resultados junto al gráfico."""
    # Crear un DataFrame con los datos ingresados por el usuario
    data = pd.DataFrame({
        'Valor': [valor for grupo in grupos.values() for valor in grupo],
        'Grupo': [grupo for grupo, valores in grupos.items() for _ in valores]
    })
    # Realizar el análisis ANOVA
    model = ols('Valor ~ C(Grupo)', data=data).fit()
    anova_table = sm.stats.anova_lm(model, typ=2)
```

```
try:
    f_value = anova_table['F'][0]
    p_value = anova_table['PR(>F)'][0]
    sum_sq_between = anova_table['sum_sq'][0]
    df_between = anova_table['df'][0]
    sum_sq_within = anova_table['sum_sq'][1]
    df_within = anova_table['df'][1]
except KeyError:
    print("No se pudo calcular la estadística F y el valor p. Verifique los datos ingresados.")
    return
# Crear un gráfico con matplotlib
fig, (ax1, ax2) = plt.subplots(1, 2, figsize=(14, 6))
# Gráfico de caja personalizado
boxplot = ax1.boxplot([valores for valores in grupos.values()], labels=grupos.keys(), patch_artist=True)
# Personalizar el gráfico
for patch in boxplot['boxes']:
    patch.set_facecolor('orange')  # Color de la caja (IQR)
for median in boxplot['medians']:
    median.set_color('red')
    median.set_linewidth(2)
ax1.set_title('Distribución de valores por grupo')
ax1.set_xlabel('Grupo')
ax1.set_ylabel('Valor')
```

```
    ax1.grid(True)
    # Mostrar los resultados como una lista en el lado derecho
    ax2.axis('off')
    ax2.set_title('Resultados del análisis ANOVA')
    # Crear la lista de resultados
    resultados_texto = (
        "Resultados del ANOVA:\n\n"
        f"- Suma de cuadrados entre grupos: {sum_sq_between:.4f}\n"
        f"- Grados de libertad entre grupos: {df_between}\n"
        f"- Suma de cuadrados dentro de grupos: {sum_sq_within:.4f}\n"
        f"- Grados de libertad dentro de grupos: {df_within}\n"
        f"- Estadística F: {f_value:.4f}\n"
        f"- Valor p: {p_value:.4f}\n"
        "\nInterpretación:\n"
        "- Un valor p bajo (< 0.05) indica diferencias significativas entre los grupos.\n"
        "- Un valor p alto (>= 0.05) sugiere que no hay diferencias significativas.\n"
    )
    # Mostrar el texto en el gráfico
    ax2.text(0.1, 0.5, resultados_texto, fontsize=12, va='center', wrap=True)
    plt.tight_layout()
    plt.show()
# Ingresar datos y realizar ANOVA
grupos = ingresar_datos()
if grupos:
```

UNIVERSIDAD NACIONAL DE JAÉN

UNIVERSIDAD NACIONAL DE JAÉN

DEPARTAMENTO ACADÉMICO DE CIENCIAS BÁSICAS Y

APLICADAS

INGENERÍA MECÁNICA Y ELÉCTRICA

Manual de correlación y regresión Lineal

Autores:

Dra. Rosario Yaqueliny Llauce Santamaria

Dra. Marcela Yvone Saldaña Miranda

Mg. Mario Félix Olivera Aldana

JAÉN – PERÚ, FEBRERO 2024

Índice

47

Índice de Tablas

Introducción

El coeficiente de correlación y la regresión lineal examinan la relación entre dos variables continuas X e Y. El coeficiente de correlación mide el grado de asociación lineal entre X e Y sin asumir una dirección específica en la relación entre ambas variables. Por otro lado, la regresión lineal simple cuantifica cómo cambia el valor promedio de la variable Y(dependiente o respuesta) en función de los cambios en la variable X(independiente o explicativa)

En este manual se incluyen ejemplos prácticos y aplicaciones en diversos campos, como la física y la ingeniería, para resolver situaciones problemáticas la vida real. Se discutirán en detalle las aplicaciones de ingeniería, particularmente en ingeniería mecánica y eléctrica. Estos campos hacen un uso extensivo de la correlación para analizar el rendimiento de los componentes y evaluar la eficiencia energética, así como la regresión para modelar el comportamiento de los materiales y optimizar los procesos de fabricación. Por ejemplo, la correlación se puede utilizar para comprender cómo los cambios en la corriente eléctrica afectan la temperatura de un motor, mientras que la regresión puede ayudar a predecir cómo se comportará un material en determinadas condiciones de tensión y temperatura. Finalmente, incluiremos una sección de ejercicios aplicados que le permitirá practicar y reforzar sus conocimientos, resolver problemas del mundo real, reforzar conceptos teóricos y demostrar la versatilidad y el poder predictivo de estos métodos estadísticos.

Los Autores

45

1. Correlación

La correlación tiene como objetivo evaluar la dirección y la fuerza de la asociación entre dos variables cuantitativas. Esto nos permite entender la intensidad de su relación y determinar si el aumento en el valor de una variable está asociado con un incremento o una disminución en el valor de la otra variable.

Para evaluar la asociación entre dos variables, la aproximación inicial generalmente se realiza mediante un diagrama de dispersión

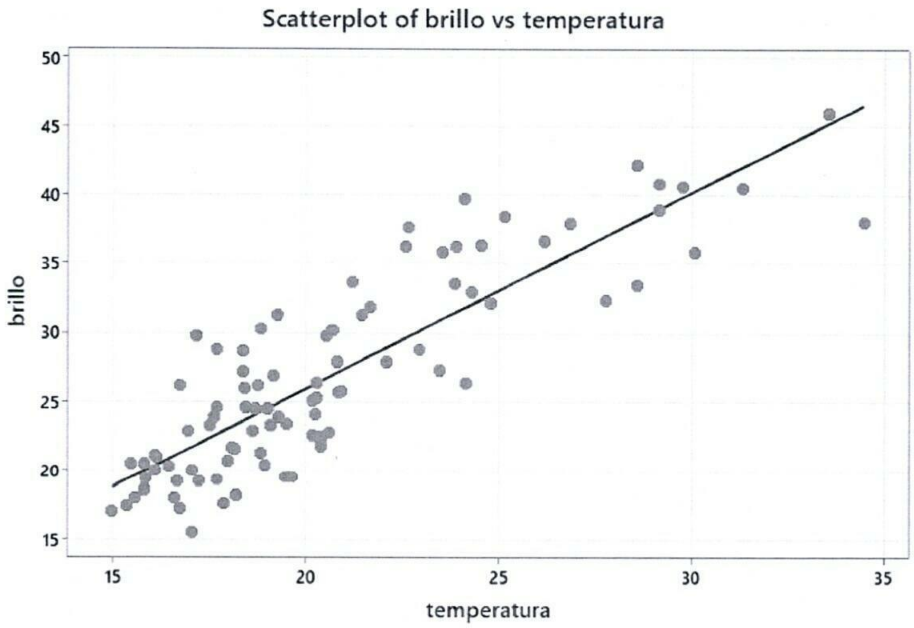


Figura 1. Scatterplot of brillo vs temperatura

En el diagrama de dispersión se observa una posible relación lineal entre la temperatura y el Scatterplot.

La nube de puntos proporciona una visualización inicial de la relación entre dos variables. Sin embargo, para medir la fuerza y dirección de esta asociación, es necesario calcular un coeficiente de correlación que cuantifique la relación entre las variables.

2. Coeficiente de Correlación

Existen dos coeficientes de correlación: el coeficiente de Pearson (paramétrico) se emplea cuando las variables cumplen con los criterios de normalidad y evalúa específicamente la adecuación de la relación lineal entre dos variables cuantitativas y el coeficiente de Spearman (no paramétrico) se usa cuando las variables no cumplen con los criterios de normalidad o son ordinales, y mide cualquier tipo de asociación, no necesariamente lineal.

3. Coeficiente de Correlación de Pearson

También conocido como coeficiente de correlación lineal, evalúa la fuerza y la dirección de la relación lineal entre dos variables cuantitativas. Este coeficiente, representado por la letra "r", oscila entre -1 y 1, donde:

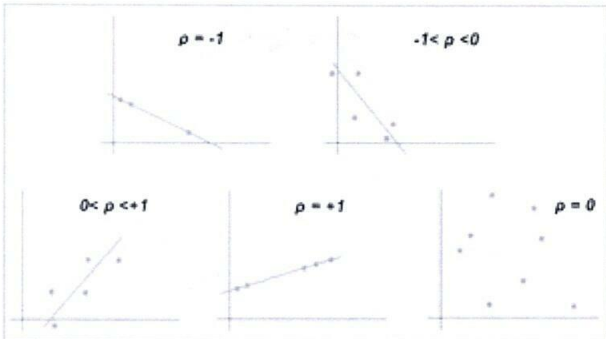


Figura 2. Diagramas de dispersión

Valores y su Interpretación:

- r =1 correlación positiva perfecta, a medida que una variable aumenta, la otra también lo hace en una relación lineal exacta.
- r = −1 correlación negativa perfecta, Cuando una variable aumenta, la otra disminuye de manera perfectamente lineal.
- r = 0 no hay correlación lineal, no existe una relación lineal entre las variables, aunque podría haber otro tipo de relación (no lineal).
- **Valores positivos (0 < r< 1):** Correlación lineal positiva. A medida que una variable aumenta, la otra también tiende a aumentar, pero no de manera perfecta. El valor de r indica la fuerza de la asociación: valores cercanos a 1 sugieren una relación lineal fuerte y positiva, mientras que valores cercanos a 0 indican una relación lineal más débil

43

- **Valores negativos (-1 < r < 0):** Correlación lineal negativa. A medida que una variable aumenta, la otra tiende a disminuir. Valores cercanos a -1 indican una relación lineal fuerte y negativa, mientras que valores cercanos a 0 indican una relación lineal más débil

Y se calcula usando:

$$r = \frac{n(\sum x_i y_i) - (\sum x_i)(\sum y_i)}{\sqrt{[n\sum x_i^2 - (\sum x_i)^2][n\sum y_i^2 - (\sum y_i)^2]}}$$

Donde:

- n = es el número de pares de datos.
- $\sum xy$ suma del producto de los pares de datos.
- $\sum x$ suma de los valores de la variable x
- $\sum y$ suma de los valores de la variable y
- $\sum x^2$ suma de los cuadrados de los valores de x
- $\sum y^2$ suma de los cuadrados de los valores de y

4. Covarianza

Es una medida estadística que indica la dirección de la relación lineal entre dos variables cuantitativas. Mide cómo varían juntas dos variables respecto a sus medias. A diferencia del coeficiente de correlación, que estandariza la medida, la covarianza proporciona una medida en las unidades de los variables originales.

La fórmula de covarianza entre dos variables X e Y es:

$$Cov(X,Y) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})$$

Donde:

- X_i e Y_i son valores individuales de las variables X e Y
- $\bar{X}$ y $\bar{Y}$ corresponde a las medias de las variables X e Y, respectivamente.

- n es el número de pares de datos

Interpretación de la covarianza:

- **Covarianza positiva**: Indica que las dos variables tienden a aumentar o disminuir juntas.
- **Covarianza negativa**: una variable tiende a aumentar mientras que la otra disminuye.
- **Covarianza cercana a cero**: no hay una relación lineal clara entre las dos variables.

Asimismo, tener en cuenta que la covarianza no proporciona una medida de la fuerza de la relación, ya que su valor depende de las unidades de las variables. Por esta razón, se suele usar el coeficiente de correlación, que estandariza la covarianza para ofrecer una medida de la fuerza y dirección de la relación.

5. Regresión Lineal simple

Es una técnica estadística utilizada para modelar y analizar la relación entre dos variables cuantitativas: independiente (X) y dependiente (Y). Su objetivo es encontrar la mejor línea recta que prediga el valor de Y en función de X

Modelo de Regresión Lineal Simple

La fórmula general es:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

Donde:

y: V. dependiente

x: V. independiente

β_0 intersección o el término constante

β_1 pendiente de la línea de regresión, indica el cambio en Y que no se explica por X.

ε término de error, que captura la variabilidad en Y que no se explica por X

6. Ejercicios Desarrollados

Ejemplo 1:

Se tienen los siguientes datos sobre las horas de estudio y las calificaciones en un examen para 5 estudiantes:

Tabla 1

Datos de Horas de Estudio (X) y Calificación (Y)

Estudiante	Horas de Estudio (X)	Calificación (Y)
1	2	65
2	3	70
3	5	85
4	1	55
5	4	80

R
I
A

Determinar el coeficiente de correlación de Pearson entre las horas de estudio y las calificaciones:

Calculamos:

- Media de X = $\frac{2+3+5+1+4}{5}$= 3
- Media de Y= $\frac{65+70+85+55+80}{5}$= 71

1. **Desviaciones estándar:**

- Para X:

$$\sigma_X = \sqrt{\frac{(2-3)^2 + (3-3)^2 + (5-3)^2 + (1-3)^2 + (4-3)^2}{5}}$$

$$= 1.414$$

40

- Para Y:

$$\sigma_y = \sqrt{\frac{(65-71)^2+(70-71)^2+(85-71)^2+(55-71)^2+(80-71)^2}{5}} = 10.68$$

2. **Covarianza:**

$$Cov(X,Y) = \frac{(2-3)(65-71)+(3-3)(70-71)+(5-3)(85-71)+(1-3)(55-71)+(4-3)(80}{5}$$

$$= 12.6$$

3. **Coeficiente:**

$$r = \frac{Cov(X,Y)}{\sigma X \sigma Y} = 0.83$$

Conclusión:

El coeficiente es aproximadamente 0.83, lo que muestra una fuerte correlación positiva entre las horas de estudio y las calificaciones.

Ejemplo 2 (Teoría):

Se ha encontrado un coeficiente de correlación de Pearson de -0.65 entre el número de horas que los empleados pasan en redes sociales durante el trabajo y su rendimiento en una evaluación de productividad.

¿qué implica este valor del coeficiente?

Solución: El coeficiente -0.65 expresa una relación negativa moderada del número de horas que los empleados pasan en redes sociales y su rendimiento en la evaluación de productividad. En otras palabras, más tiempo pasa en las redes sociales, es probable que tu rendimiento en tareas relacionadas con la productividad baje.

Ejemplo 3:

Imaginemos que queremos estimar cuánto costará una casa según su tamaño en metros cuadrados. Contamos con la siguiente información:

Tabla 2

Datos de Tamaño (m^2) y Precio (USD)

Tamaño (m^2	*Precio (USD)*
50	150,000
60	180,000
70	210,000
80	240,000
90	270,000

Queremos ajustar una regresión lineal simple para determinar el precio y basado en el tamaño x.

Solución:

1. **Definir el modelo:**

Fórmula , modelo de regresión lineal simple:

$$y = \beta_0 + \beta_1 x + \epsilon$$

Donde:

β_0 intersección

β_1 pendiente, y ϵ es el error.

2. **Calcular los parámetros:**

Primero, necesitamos calcular la media de X y Y:

$$x = \frac{50 + 60 + 70 + 80 + 90}{5} = 70$$

$$y = \frac{150{,}000 + 180{,}000 + 210{,}000 + 240{,}000 + 270{,}000}{5} = 210{,}000$$

30

Luego, calculamos la pendiente β_1 y la intersección β_0:

$$\beta_1 = \frac{\sum_{i=1}^{n}(x_i - x)(y_i - y)}{\sum_{i=1}^{n}(x_i - x)^2}$$

$$\beta_1 = \frac{(50-70)(150{,}000-210{,}000)+(60-70)(180{,}000-210{,}000)+(70-70)(210{,}000-210{,}000)}{(50-70)^2+(60-70)^2+(70-70)^2+(80-70)^2+(90-70)^2}\ldots$$

$$\ldots.\frac{+(80-70)(240{,}000-210{,}000)+(90-70)(270{,}000-210{,}000)}{(50-70)^2+(60-70)^2+(70-70)^2+(80-70)^2+(90-70)^2} = 3000$$

Para β_0:

$$\beta_0 = \bar{y} - \beta_1\bar{x} = 210{,}000 - 3{,}000 \cdot 70 = 210{,}000 - 210{,}000 = 0$$

Así que la ecuación de la regresión es:

$$y = 0 + 3{,}000x$$

Interpretación:

La pendiente $\beta_1 = 3000$ indica que, por cada metro cuadrado adicional, el precio de la casa aumenta en 3,000 USD. La intersección $\beta_0 = 0$ no tiene sentido práctico en este caso, ya que un tamaño de 0 m² no corresponde a una casa real, pero es parte del modelo matemático.

Ejemplo 4: Regresión Lineal Múltiple

Se desea averiguar el precio de una casa teniendo en cuenta dos factores: su tamaño en metros cuadrados x_1 y número de habitaciones x_2. Los datos son los siguientes:

Tabla 3

Datos de Tamaño (m²), Habitaciones y Precios (USD)

Tamaño (m²)	Habitaciones	Precios (USD)
50	2	150,000
60	3	180,000
70	3	210,000
80	4	240,000
90	4	270,000

Queremos ajustar un modelo de regresión lineal múltiple.

Solución:

Definir el modelo:

El modelo de regresión lineal múltiple es:

$$y = \beta_0 + \beta_1 x + \epsilon$$

Calcular los parámetros:

Los cálculos exactos son más complejos y generalmente se realizan usando software estadístico. El proceso incluye resolver el sistema de ecuaciones de los parámetros.

Aquí presentamos una forma simplificada de hacerlo con software o herramientas estadísticas (como R, Python, o incluso Excel). Sin embargo, el método manual involucra matrices y álgebra lineal, lo cual puede ser extenso para detallar aquí.

Interpretación:

Supongamos que obtienes los siguientes parámetros:

$$\beta_0 = 500{,}000$$

$$\beta_1 = 2{,}000 (pormetrocuadrado)$$

$$\beta_2 = 10{,}000 (por\ habitacion)$$

Entonces la ecuación sería:

$$y = 50{,}000 + 2{,}000x_1 + 10{,}000x_2$$

Esto indica que el precio base de la casa es 50,000 USD. Cada metro cuadrado adicional aumenta el precio en 2,000 USD, y cada habitación adicional aumenta el precio en 10,000 USD.

5. Aplicaciones de la Correlación y Regresión en la Ingeniería Mecánica y Eléctrica

5.1. Correlación

Análisis de Desempeño de Componentes

El análisis de desempeño de componentes en sistemas mecánicos y eléctricos frecuentemente utiliza la correlación para identificar relaciones entre variables críticas. Por ejemplo, en un motor eléctrico, la correlación entre la corriente de entrada y la temperatura del motor puede revelar cómo varía el rendimiento con respecto al calor generado (Montgomery & Runger, 2018)

Evaluación de la Eficiencia Energética

La eficiencia energética en sistemas eléctricos puede ser evaluada mediante técnicas de correlación. Por ejemplo, la relación entre el voltaje y el consumo de energía en distintos dispositivos puede ser analizada para optimizar su uso y reducir pérdidas energéticas ((Chen & Chen, 2020).

5.2. Regresión

Modelado de Comportamiento de Materiales

En ingeniería mecánica, la regresión se utiliza para modelar el comportamiento de materiales bajo diferentes condiciones de estrés y temperatura. Los modelos de regresión pueden predecir cómo un material se deformará bajo ciertas cargas, lo cual es crucial para el diseño de estructuras seguras y eficientes (Callister & Rethwisch, 2018)

Optimización de Procesos de Manufactura

La regresión es un método clave usado en optimizar procesos de manufactura. Permite desarrollar modelos predictivos para ajustar parámetros de proceso, como la velocidad de corte y la alimentación, para minimizar defectos y maximizar la eficiencia (Groover, 2019)

Análisis de Fallas

Este análisis en sistemas eléctricos y mecánicos puede beneficiarse del uso de la regresión. Por ejemplo, la regresión logística puede emplearse para modelar la probabilidad de falla de componentes en función: la edad del componente, las condiciones de operación y el historial de mantenimiento (Jardine, Lin,, & Banjevic, 2006)

5.2.1. Regresión Múltiple y Sistemas Complejos

Predicción de Vida Útil de Componentes

La regresión múltiple se utiliza para determinar la vida útil de componentes en sistemas complejos, considerando múltiples variables influyentes. Este tipo de análisis es importante para el mantenimiento predictivo y la gestión de activos (Meeker & Escobar, 1998).

Modelos de Automatización y Control

En este modelo la regresión múltiple puede ser utilizada para crear modelos que explican la relación entre diversas variables de entrada y salida. Estos modelos son fundamentales para diseñar controladores que aseguren el rendimiento óptimo del sistema (Ogata, 2010)

6. Ejercicios aplicados a la Carrera

En una fábrica, se desea analizar el vínculo que tiene el esfuerzo con la deformación de cierto material que desea trabajar en factoría

Ejemplo 5

Tabla 4

Datos de Material, Esfuerzo (psi) y Deformación (mm)

Material	1	2	3	4	5	6	7	8	9	10	11
Esfuerzo (psi)	35	34	37	36	35	36	35	33	37	36	36
Deformación (mm)	2	0	3	4	3	4	3	1	4	3	5

Utilizaremos a continuación el diagrama de dispersión:

Diagrama de dispersión

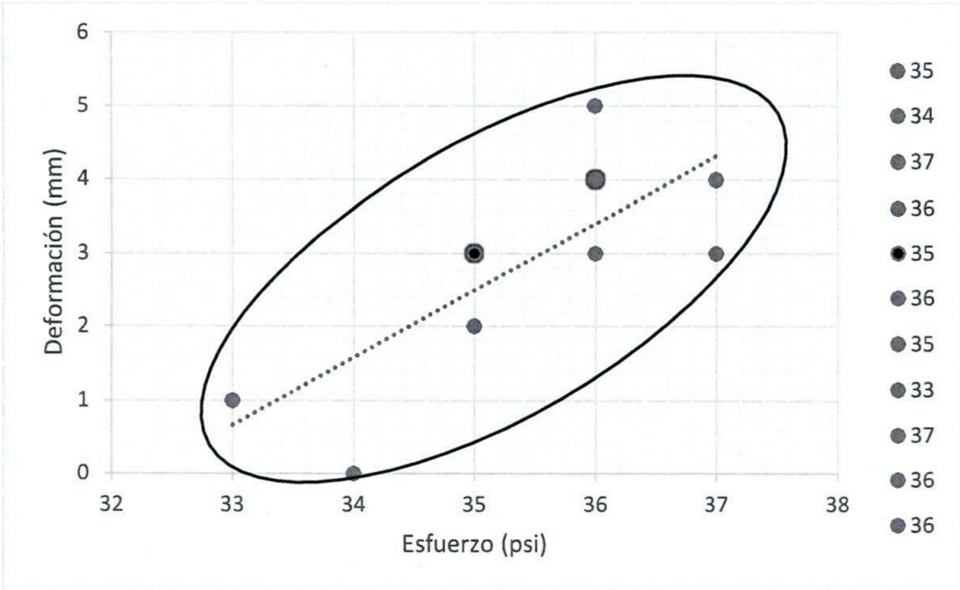


Figura 3

Tenemos una tendencia positiva. Por ende, la relación es positiva.

Calculamos y/o Trazamos la Línea de Regresión, junto con ello el Coeficiente de Regresión.

Formula a emplear: $Y_c = a + bX$, Donde:

- ✓ Y_c: Y dado un valor de X.
- ✓ a : Punto de intersección con el eje y cuando x=0
- ✓ b : Dependencia de la línea de regresión (coeficiente de regresión)

$a = \bar{Y} - b\bar{X}$ $\quad \bar{Y} = \frac{\sum Y}{n}$ $\quad \bar{X} = \frac{\sum X}{n}$ $\quad b = \frac{n(\sum xy) - (\sum x)(\sum y)}{n \sum x^2 - (\sum x)^2}$

Ejemplo 6

Tabla 5

Cálculo de $(X_i)^2$

Material	X	Y	$X_i * Y_i$	$(X_i)^2$
1	35	2	70	1225
2	34	0	0	1156
3	37	3	111	1369
4	36	4	114	1296
5	35	3	105	1225

33

6	36	4	144	1296
7	35	3	105	1225
8	33	1	33	1089
9	37	4	148	1369
10	36	3	108	1296
11	36	5	180	1296
Suma	390	32	1118	13842

Ahora reemplazamos en $b = \frac{n(\sum xy)-(\sum x)(\sum y)}{n\sum x^2-(\sum x)^2}$

$$b = \frac{11(1118)-(390)(32)}{11(13842)-(390)^2}$$

$$b = -\frac{91}{81} = -1.12$$

Remplazamos en $a = \bar{Y} - b\bar{X}$

$$a = \frac{\sum Y}{n} - b\frac{\sum X}{n}$$

$$a = \frac{32}{11} - b\frac{390}{11}$$

$$a = 2.909 - (-1.12)(35.454)$$

$$a = 42.617$$

Remplazamos en $Y_c = a + bX$

$$Y_c = 42.617 - 1.12X$$

Ahora graficamos la línea de regresión:

Calculamos Y_c , para X=33 y X=37

Ejemplo 7

Tabla 6

Datos de X y Y_c

32

X	Y_c
33	5.657
37	1.177

Diagrama de dispersión

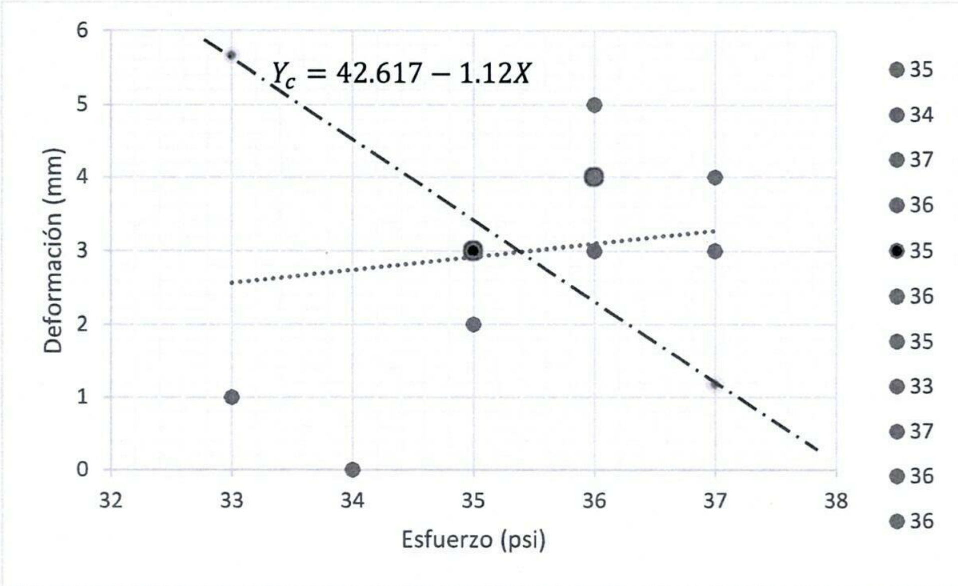


Figura 4

Caculo de coeficiente de correlación de Pearson

Tenemos la siguiente formula: $r = \frac{n(\sum xy)-(\sum x)(\sum y)}{\sqrt{([n\sum x^2-(\sum x)^2][n\sum y^2-(\sum y)^2])}}$

Tabla 7

Cálculo de $(Y_i)^2$

Materiales	X	Y	$X_i * Y_i$	$(X_i)^2$	$(Y_i)^2$
1	35	2	70	1225	4
2	34	0	0	1156	0
3	37	3	111	1369	9
4	36	4	114	1296	16
5	35	3	105	1225	9
6	36	4	144	1296	16
7	35	3	105	1225	9

8	33	1	33	1089	1
9	37	4	148	1369	16
10	36	3	108	1296	9
11	36	5	180	1296	25
Suma	390	32	1118	13842	114

Remplazamos $r = \frac{11(1118)-(390)(32)}{\sqrt{([11(13842)-(390)^2][11(114)-(32)^2])}}$

$$r = -0.942$$

$$|r| = 0.942$$

El valor de r se asemeja a uno; la relación lineal entre las variables es bastante fuerte.

Ejemplo 8

Determinar el vínculo entre la temperatura de operación de un motor eléctrico y su eficiencia energética

Temperatura de operación (°C)

Eficiencia energética (%)

Haremos la recopilación de datos:

Supongamos que se han recogido los siguientes datos de un motor eléctrico en diferentes condiciones de operación:

Tabla 8

Datos de Temperatura de operación (°C) y Eficiencia energética (%)

Temperatura de operación (°C)	Eficiencia energética (%)
20	92
25	91
30	89
35	88
40	85

30

45	83
50	80
55	78
60	75

Una vez recopilada lo datos, vamos a hacer el análisis correspondiente

Análisis de Correlación

Cálculo del coeficiente de correlación.

$$[r = \frac{\sum(X-\overline{X})(Y-\overline{Y})}{\sqrt{\sum(X-\overline{X})^2\sum(Y-\overline{Y})^2}}]$$

Análisis:

Cálculo de la regresión lineal:

La fórmula de la recta de regresión es:

$$[Y = a + bX]$$

Donde:

- (Y) Eficiencia energética.
- (X) Temperatura de operación.
- (a) Ordenada al origen (intercepto).
- (b) Pendiente de la recta (coeficiente de regresión).

Resultados del Ejercicio de Correlación y Regresión

Coeficiente de Correlación:

$$[r = -0.992]$$

Este valor indica una fuerte correlación negativa entre la temperatura de operación y la eficiencia energética. A medida que aumenta la temperatura, la eficiencia energética disminuye.

Regresión Lineal:

$$[\text{Eficiencia} = 101.89 - 0.433 \times \text{Temperatura}]$$

Donde:

El intercepto (a) es 101.89.

La pendiente (b) es -0.433.

Interpretación

Intercepto (101.89): Cuando la temperatura de operación es 0°C, la eficiencia energética esperada del motor es 101.89%. Aunque este valor no es realista en términos físicos, es útil para la ecuación de regresión.

Pendiente (-0.433): Por cada aumento de 1°C en la temperatura de operación, la eficiencia energética disminuye en 0.433%.

Análisis de Resultados

R-cuadrado (0.984): el valor de (R^2) expresa que el 98.4% de la variación en la eficiencia energética se puede explicar por la variación en la temperatura de operación.

Probabilidad del F-Statistic (1.49e-07Este valor muestra que el modelo de regresión es muy relevante y confiable.

Este tipo de análisis es útil en ingeniería mecánica y eléctrica para optimizar las condiciones operativas y mejorar el rendimiento de los equipos.

Ejemplo 9

Una muestra aleatoria de 9 trabajadores eléctricos aparentemente buenos en sus trabajos entre 10 a 15 trabajadores produjo los siguientes datos respecto a los trabajadores y la eficiencia en porcentaje:

Trabajadores	Eficiencia %
10	86
10.5	88
11	92
12.5	93
13	95
13.5	96
14	97
14.5	98
15	99

- Encontrar la ecuación de la regresión lineal.
- En el diagrama de dispersión, trazar la recta de la ecuación de regresión lineal.
- Si el trabajador 13, ¿cuál es la eficiencia?

Solución:

Para determinar los coeficientes en una regresión lineal, se emplean las ecuaciones para la pendiente y la intersección,

X es trabajadores.

Y es la eficiencia.

Sumatorias:

$$\sum X = 10 + 10.5 + 11 + 12.5 + 13 + 13.5 + 14 + 14.5 + 15 = 114$$

$$\sum Y = 86 + 88 + 92 + 93 + 95 + 96 + 97 + 98 + 99 = 844$$

$$\sum XY = 10 * 86 + 10.5 * 88 + 11 * 92 + 12.5 * 93 + 13 * 95 + 13.5 * 96 + 14 * 97 + 14.5 * 98 + 15 * 99 = 10753.5$$

$$\sum X2 = 100 + 110.25 + 121 + 156.25 + 169 + 182.25 + 196 + 210.25 + 225$$
$$= 1470$$

Pendiente:

$$m = \frac{n(\sum xy) - ((\sum x) \ (\sum y)}{n(\sum x^2) - (\sum x)^2}$$

$$m = \frac{9 \times (10753.5) - (114)(844)}{9 \times (1470) - (114)^2} = 2.42$$

Intersección:

$$b = \frac{844 - (2.42)(114)}{9} = 63.12$$

La ecuación de regresión es:

$y = 63.12 + 2.42(x)$

Si el trabajador 13, ¿cuál es la eficiencia?

$y = 63.12 + 2.42 \ x\ 13 = \ 63.12 + 31.46 = 94.58\%$

La eficiencia del trabajador 13 es de 94.58%

Ejemplo 10

En una entidad eléctrica trabajan cuatro técnicos operarios. Los años que han pasado de sus títulos y la cantidad de meses que han trabajado en esta empresa, aquí se muestran cada uno:

años(x)	3	4	5	6
meses(y)	4	3	2	1

- calcular el valor del coeficiente de correlación lineal e interpretarlo.

solución:

Xi	Yi	Xi*Yi	X_i^2	Y_i^2
3	4	12	9	16

4	3	12	16	9
5	2	10	25	4
6	1	6	36	1
18	10	40	86	30

$$X = \frac{18}{4} = 4{,}5$$

$$Y = \frac{10}{4} = 2{,}5$$

$$\sigma_x^2 = \frac{86}{4} - 4.5^2 = 1.25$$

$$\sigma_y^2 = \frac{30}{4} - 2.5^2 = 1.25$$

$$\sigma_{xy} = \frac{40}{4} - 4.5 * 2.25 = -1.25$$

$$r = \frac{-1.25}{\sqrt{1.25} * \sqrt{1.25}} = \frac{-1.25}{1.25} = -1$$

interpretación:

Por lo tanto , la correlación es perfecta e inversa.

7. Conclusiones

- La correlación y la regresión son métodos estadísticos clave para entender y modelar las relaciones entre variables cuantitativas Estas técnicas nos ayudan a identificar patrones y hacer predicciones, asimismo facilita la toma de decisiones que se basan en datos en diferentes áreas, especialmente en ingeniería.
- En este manual, hemos explorado cómo la correlación evalúa el vínculo entre variables, mientras que la regresión facilita la predicción de resultados. Estas técnicas son fundamentales para que los procesos de ingeniería optimicen y mejoren el rendimiento de los sistemas, demostrando su utilidad práctica en aplicaciones del mundo real.
- Dominar la correlación y la regresión permite a los profesionales convertir los datos en información valiosa, impulsando la innovación y la mejora continua. Estas

UNIVERSIDAD NACIONAL DE JAÉN

UNIVERSIDAD NACIONAL DE JAÉN

DEPARTAMENTO ACADÉMICO DE CIENCIAS BÁSICAS Y APLICADAS

INGENERÍA MECÁNICA Y ELÉCTRICA

MANUAL

ANÁLISIS DE SERIES DE TIEMPO

Autores:
Dra. Rosario Yaqueliny Llauce Santamaria
Dra. Marcela Yvone Saldaña Miranda
Mg. Mario Félix Olivera Aldana

JAÉN – PERÚ, FEBRERO 2025

ÍNDICE

75

1. Introducción

El análisis de series de tiempo es una disciplina fundamental en estadística y probabilidad, utilizada para comprender y prever el comportamiento de datos recopilados en intervalosde tiempo sucesivos. Las series de tiempo son omnipresentes en diversos campos como la economía, la meteorología, la ingeniería, y la ciencia, donde se necesita analizar y predecir datos temporales para tomar decisiones informadas.

Una serie de tiempo se puede descomponer en varios componentes básicos que ayudan adesentrañar la estructura subyacente y los patrones recurrentes en los datos. Estoscomponentes incluyen la tendencia, la estacionalidad, el ciclo y la componente aleatoriao error. Cada uno de estos componentes desempeña un papel crucial en la explicación delas variaciones observadas en la serie de tiempo y su identificación precisa es esencial para un análisis eficaz y una predicción precisa.

La correcta identificación y separación de estos componentes permiten una mejor comprensión de los datos históricos y mejoran la capacidad de realizar predicciones futuras. A través de la descomposición de una serie de tiempo en sus componentes fundamentales, se puede desarrollar modelos más precisos y útiles para diversas aplicaciones prácticas, desde la planificación económica hasta la gestión de inventarios yla predicción del clima.

Este informe explorará en detalle cada uno de estos componentes, los métodos utilizadospara su identificación y análisis, y cómo su descomposición puede mejorar nuestra comprensión y capacidad de predicción de las series de tiempo.

2. Objetivos

- Identificar y definir claramente los componentes básicos de una serie de tiempo:tendencia, estacionalidad, ciclo y componente aleatoria.
- Describir cómo cada componente influye en la estructura y variabilidad de unaserie de tiempo

74

3. Desarrollo

3.1.Tendencia (T)

En el análisis de series de tiempo, la tendencia es el patrón subyacente a largo plazo querepresenta el crecimiento o disminución en la serie a lo largo del tiempo. No necesariamente es lineal, y puede mostrar una evolución ascendente o descendente (Estadística, 2024).

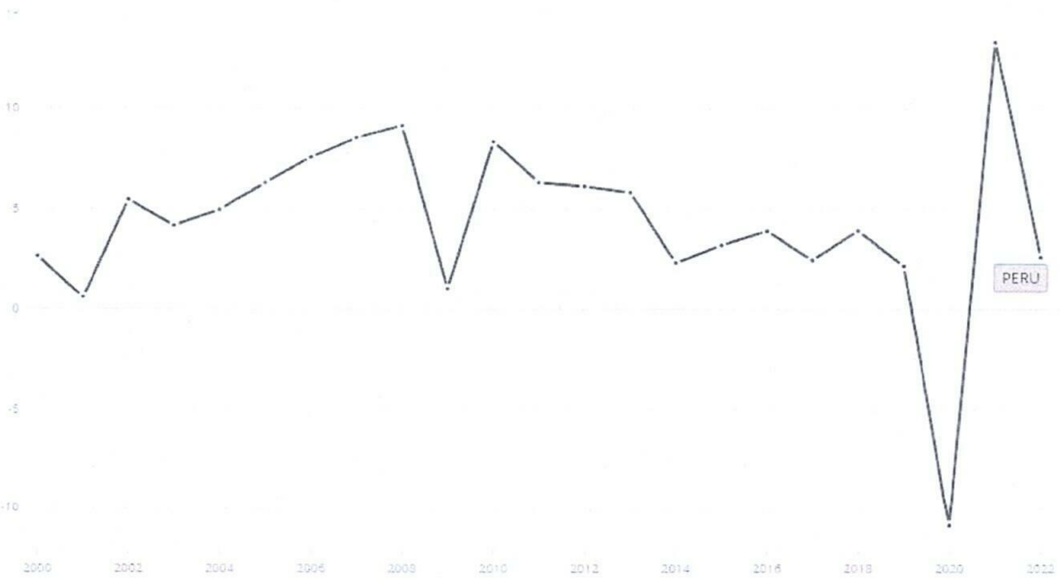
3.1.1. Aplicaciones

Crecimiento Económico de un País:

- Aumento continuo del Producto Interno Bruto (PIB) debido a factores como lainversión, la innovación tecnológica, y el aumento de la productividad laboral.

Gráfico 1:

Crecimiento del PBI (% anual) en Perú, 2000-2022



Nota: Datos de cuentas nacionales del Banco Mundial y archivos de datos de CuentasNacionales de la OCDE (*World Bank Open Data*, s. f.)

Aumento de la Población en una ciudad

Incremento sostenido de la población de una ciudad como resultado de la migración, el crecimiento natural (nacimientos menos muertes), y la expansión urbana

Incremento de Ventas de una Empresa:

Crecimiento sostenido en las ventas de una empresa debido a la expansión de mercado, la mejora de productos, y el aumento de la demanda.

Métodos de Identificación:

Regresión Lineal:

Descripción: La regresión lineal es un método estadístico utilizado para ajustar una línea recta a los datos de una serie de tiempo. Este método modela la relación entre una variable dependiente (el valor de la serie en un momento dado) y una variable independiente (el tiempo).

3.1.2. Proceso

1. **Modelado de la Relación o proyección de tendencia**: Se establece una relación lineal entre el tiempo t y el valor de la serie Y_t mediante la ecuación de la línea de regresión:

$$Y_t = a + bt$$

$$a = \bar{Y} - bt$$

$$b = \frac{\sum tY - n\bar{t}\bar{Y}}{\sum t^2 - n\bar{t}^2}$$

Donde:

- a es la intersección o el valor de Y cuando $t = 0$
- b es la pendiente de la línea, que indica la tasa de cambio de Y por unidad de tiempo
- t es periodo

Ejemplo 1

Consumo de eléctrico en la ciudad de lima

Contexto: Considere la serie de tiempo formada por el consumo de electricidad emitida por electro norte sobre los últimos 12 años

Datos: La cual se muestra en tabla:

Tabla 1

Registro de consumo eléctrico

AÑO	**Periodo (t)**	**Consumo (KWh) (y)**	**(ty)**	x^2
2013	1	1045	1045	1
2014	2	1123	2246	4
2015	3	1287	3861	9
2016	4	1234	4936	16
2017	5	1098	5490	25
2018	6	1256	7536	36
2019	7	1013	7091	49
2020	8	1145	9160	64
2021	9	1300	11700	81
2022	10	1244	12460	100
2023	11	1272	13684	121

72

2024	12	1182	14184	144
Total	$\bar{x}$= 6.5	$\bar{y}$ = 1183.25	$\sum$= 93393	$\sum$= 650

Calculamos *b*

Reemplazamos en la fórmula:

$b = \sum xy \; - n\bar{x}\bar{y} / \sum x^2 - n\bar{x}^2$

$$b = \frac{93393-(12)*(6.5)*(1183.25)}{650-(12)*(6.5)^2}$$

b = 7.7

Reemplazamos para "a"

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

a = 1183.25 – (7.7) *(6.5)

a = 1133.2

Reemplazamos para $\bar{y}$

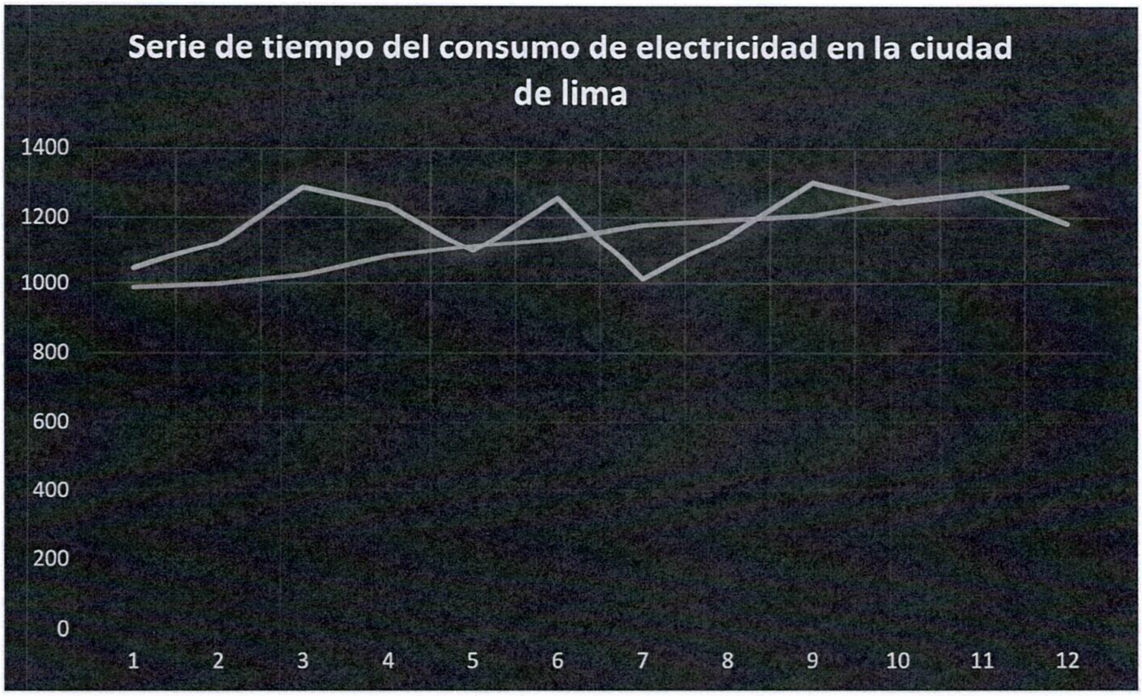
$\bar{y} = a + bx$

$\bar{y}$= *1133.2 + (7.7) *(13)*

$\bar{y}$= *1233.*

Resultado:

Para el 2025 o el periodo 13 se consumirán 1233.3 KWh en la ciudad de lima



Fuentes: Elaboración propia

3.2.Suavización Exponencial

Descripción: El suavizado exponencial es un método que asigna más peso a los

71

datos más recientes, lo que permite capturar tendencias y patrones más recientes en la serie detiempo.

3.2.1. Proceso

1. **Cálculo del Valor Suavizado**: El valor suavizado en el tiempo t se calcula utilizando la fórmula.

$$S_t = S_{t-1} + \alpha(Y_{t-1} - S_{t-1})$$

Donde:

- S_t es el valor suavizado en el tiempo t.
- Y_i es el valor observado en el tiempo t
- α es el parámetro de suavizado ($0 < \alpha < 1$), que determina el peso asignado a los datos más recientes

2. **Elección del Parámetro** α: Se selecciona un valor de α que balancee la respuesta rápida a los cambios recientes y la estabilidad general del suavizado. Un α alto responde más rápidamente a los cambios, mientras que un α bajo proporciona unsuavizado más estable.

Ejemplo 2

Predicción del Tráfico Vehicular

Se dispone de datos históricos de la cantidad de vehículos que pasan por una calle en un horario específico durante los últimos 10 días. Traza la tendencia de la tabla ypredice el día 11.

Datos: Los datos de tráfico vehicular (en número de vehículos) son los siguientes:

Tabla 2:

Datos de tráfico vehicular en la calle Cajamarca

Día	**Vehículos**
1	**200**
2	**210**
3	**195**
4	**205**
5	**220**
6	**230**
7	**240**
8	**235**
9	**225**
10	**215**

11	S_{11}

Fuente: elaboración propia

Se utiliza la formula usando de manera arbitraria $\alpha = 0.3$:

$$S_t = S_{t-1} + \alpha\,(Y_{t-1} - S_{t-1})$$

$$Día\ 2 = 200 + 0.3(200 - 200) = 200$$

Se obtiene la siguiente tabla repitiendo hasta el día 10:

Tabla 3:

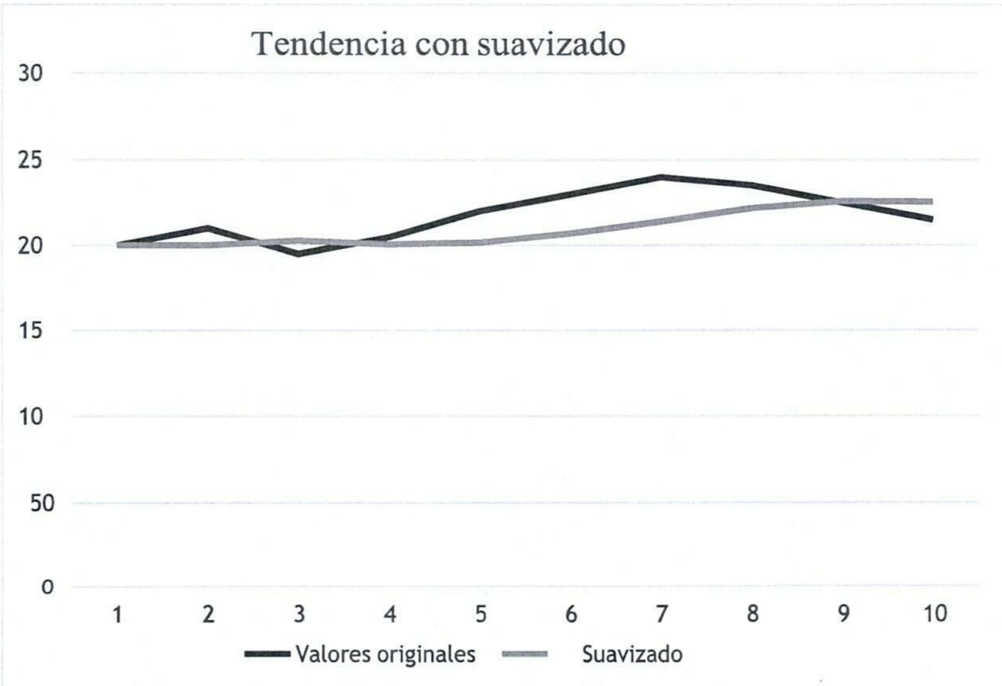
Valores de S_t según el día.

Día	S_t
1	200
2	200
3	203
4	200.6
5	201.92
6	207.34
7	214.14
8	221.89
9	225.82
10	225.58
11	S_{11}

Fuentes: Elaboración propia

Con los datos obtenidos de las tablas se hace una comparación de la tabla 1 y la tendencia encontrada con gráficas usando EXEL

Gráfica 3:



69

Fuentes: Elaboración propia

Calculando el día 11:

$$S_{11} = S_{10} + \alpha(Y_{10} - S_{10})$$

$$S_{11} = 225.58 + 0.3(215 - 225.58) = 222.41$$

Interpretación: Aproximadamente 222.41 vehículos pasarán por la carretera en el día 11.

Ahora se observará como la gráfica cambia usando un valor mayor a alfa, α = 0.90

Gráfico 4:

Fuentes: Elaboración propia

3.2.2. Aplicación

Útil para Series con Tendencia No Lineal y Fluctuaciones Menores: Este método es eficaz para capturar tendencias en series de tiempo que no siguen un patrón lineal claro y tienen fluctuaciones menores.

Ejemplo 3: Suavizado de datos de ventas mensuales para identificar la tendencia subyacente. El suavizado exponencial puede ayudar a destacar la tendencia general en las ventas, minimizando las fluctuaciones menores.

3.3.Medias Móviles

Descripción: Las medias móviles son un método que calcula la media de los valores de una serie de tiempo dentro de una ventana móvil. Este método ayuda a suavizar las fluctuaciones a corto plazo y resaltar la tendencia subyacente

3.3.1. Proceso

Cálculo de la Media Móvil: La media móvil en el tiempo t se calcula como:

$$MA_t = \frac{1}{n}\sum_{i=1}^{n} Y_{t-i}$$

donde:

- MA_t es la media móvil en el tiempo t.
- n es el tamaño de la ventana.
- Y_{t-i} son los valores observados en los tiempos $t-1, \ldots, t-(n-1)$

Ejemplo 4

Análisis del Uso de cables eléctricos en una Obra de Construcción

Contexto: Disponemos de datos históricos del uso diario de cables eléctricos (en metros)en una obra de construcción durante los últimos 10 días. Queremos utilizar las medias móviles para suavizar las fluctuaciones diarias y entender la tendencia subyacente en el uso de los cables, así como para predecir el uso para el día 11.

Datos: La cantidad de cables eléctricos utilizada en los últimos 10 días es la siguiente:

Tabla 4:

Cantidad de concreto utilizado en los últimos 10 días.

Día	Uso de cables eléctricos (m)
1	20
2	22
3	18
4	21
5	23
6	24
7	25
8	27
9	26
10	28
11	MA_{11}

Fuentes: Elaboración propia

3.3.2. Parámetro de la media Móvil

Vamos a elegir un tamaño de ventana, n. Supongamos que elegimos $n = 2$ días.

Proceso de Cálculo:

1. **Inicialización**: Para calcular la media móvil, tomamos la media de los

67

valoresde la ventana seleccionada.

2. **Cálculo de medias móviles**: Utilizamos la fórmula de la media móvil paracalcular las medias móviles para los días desde el día 3 hasta el día 10

$$MA_t = \frac{1}{n}\sum_{i=1}^{n} Y_{t-i}$$

$$Día\ 3: MA_3 = {}_2(Y_2 + Y_1) = {}_3(22 + 20) = 21$$

Siguiendo el patrón para cada día desde el día 3 hasta el día 10 se obtiene:

Tabla 5:

Cálculo de medias móviles según el día

Día	MA_t
1	
2	
3	21
4	20
5	19.5
6	22
7	23.5
8	24.5
9	26
10	26.5
11	

Fuentes: Elaboración propia

Para calcular el pronóstico en el día 11 se usa:

$$MA_{11} = \frac{1}{2}\sum_{i=1}^{2} Y_{t-i}$$

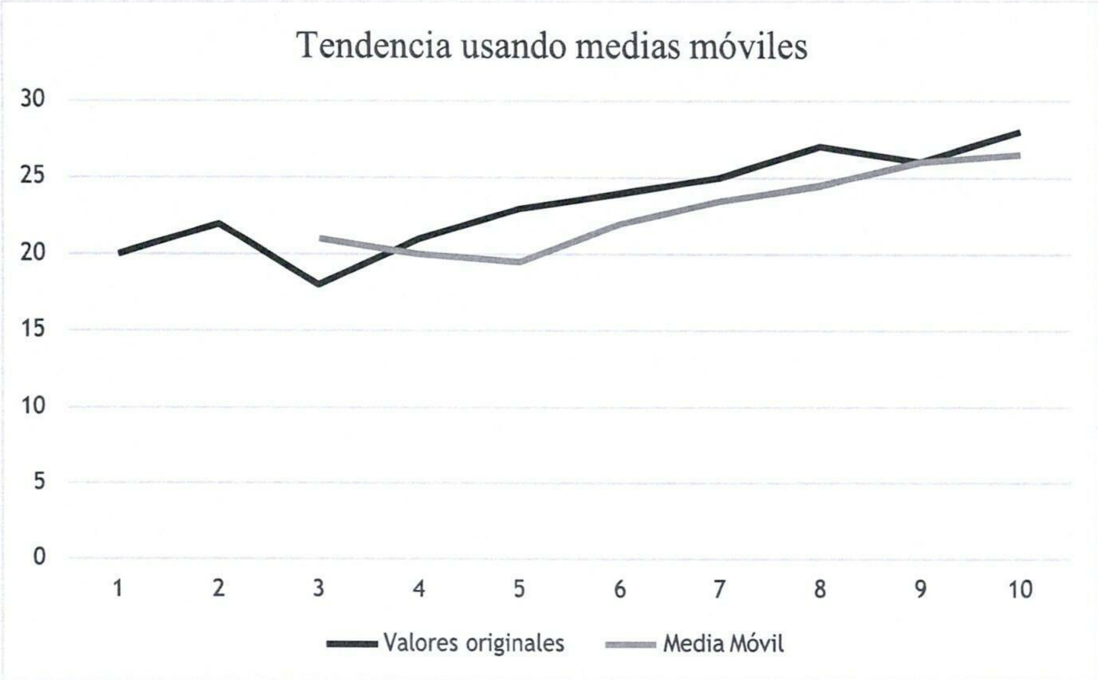
$$\mathrm{MA}_{11} = \frac{28 + 26}{2} = 27$$

Interpretación: Se pronostica que el día 11 se usaran 27 m de cables eléctricos

Usando EXEL, relacionamos los valores originales con la tendencia usando mediasmóviles

Gráfico 5:

66



3.3.3. Tipos de las medias móvil

- **Media Móvil Simple (SMA)**: Calcula la media aritmética de los valores en la ventana.
- **Media Móvil Ponderada (WMA)**: Asigna diferentes pesos a los valores en la ventana, dando más importancia a los datos más recientes.

3.3.4. Aplicación

- **Adecuada para Datos con Fluctuaciones Periódicas o Aleatorias**: Las medias móviles son útiles para suavizar series de tiempo con fluctuaciones periódicas o aleatorias, ayudando a identificar la tendencia subyacente.
- **Ejemplo 5**: Análisis de tendencias en datos de tráfico web diario utilizando una ventana de 7 días para calcular la media móvil semanal. La media móvil ayuda asuavizar las variaciones diarias y resaltar la tendencia semanal

3.4.La Variación Estacional

La estacionalidad es un concepto fundamental. Se refiere a las fluctuaciones periódicasy predecibles que ocurren durante una serie de tiempo debido a factores estacionales. Las estaciones del año, las festividades, los días de la

65

semana y otros eventos repetidosafectan estas variaciones estacionales.

3.4.1. Características de la Estacionalidad

1. **Periodicidad**: Las fluctuaciones se repiten en intervalos regulares y conocidos,como mensuales, trimestrales, o anuales.
2. **Patrón Recurrente**: Los cambios en los datos siguen un patrón que se repite enciclos fijos.
3. **Causa Subyacente**: La estacionalidad es impulsada por factores externos ypredecibles, como el clima, eventos culturales, vacaciones, etc.
4. **Constancia en el Tiempo**: Aunque la magnitud puede variar, el patrónestacional generalmente permanece constante a lo largo del tiempo.

3.4.2. Métodos para Identificar y Analizar la Estacionalidad

Descomposición de Series Temporales:

- **Descomposición Clásica**: Divide una serie temporal en componentes de tendencia (T), estacionalidad(S), y ruido(R), . Utiliza métodos aditivoso multiplicativos.

3.4.3. Modelo Aditivo

$$Y_t = T_t + S_t + R_t$$

Donde:

- Y_t = Es el valor observado de la serie temporal en el tiempo.
- T_t = Es el componente de tendencia en el tiempo.
- S_t = Es el componente estacional en el tiempo.
- R_t = Es el componente de ruido o aleatorio en el tiempo.

64

3.4.4. Modelo Multiplicativo

$$Y_t = T_t x S_t x C_t x E_t$$

- Y_t = Es el valor observado en el tiempo.
- T_t=Es la tendencia en el tiempo.
- S_t=Es la estacional en el tiempo.
- C_t=Es el ciclo en el tiempo
- E_t = Es el componente de error en el tiempo.

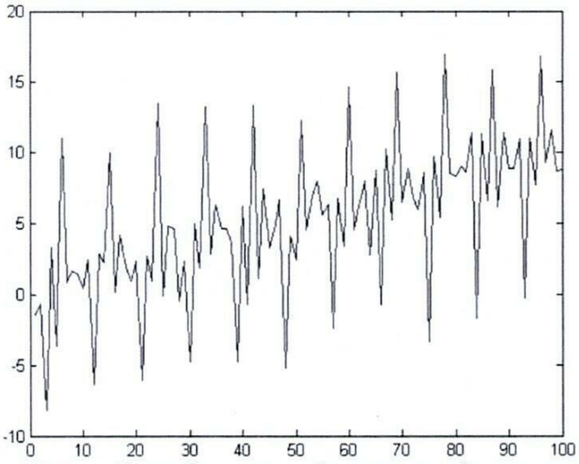
3.4.5. Pasos para descomposición Clásica

1. **Calcular la Tendencia**: Usualmente se realiza mediante el promedio móvil.
2. **Estimar la Estacionalidad**: Se obtiene el promedio de los valores estacionalesdespués de remover la tendencia.
3. **Calcular los Residuos**: Restar (para el modelo aditivo) o dividir (para el modelomultiplicativo) la tendencia y la estacionalidad del valor original

3.5.Componente Aleatoria

Son movimientos erráticos que no siguen un patrón específico y que obedecen a causas diversas. Este comportamiento representa todos los tipos de movimientos de una serie detiempo que no son tendencia, variaciones estacionales ni fluctuaciones cíclicas. Este componente es prácticamente impredecible. (Gonzalo, Rios , 2008)

Figura 1 *Componente aleatoria*



Nota. Correlograma de una serie no estacionaria. Tomado de (Gonzalo,Rios , 2008).

Este componente explica la variabilidad aleatoria de la serie, es impredecible, es decir,

63

nose puede esperar predecir su impacto sobre la serie de tiempo, esto se debe a factores a corto plazo, imprevisibles y no recurrentes que afectan a la serie de tiempo.

La mayoría de las componentes irregulares se conforman de variabilidad aleatoria. Sin embargo, ciertos sucesos a veces impredecibles como huelgas, cambio de clima, elecciones, conflictos armados, factores económicos, etc. pueden causar irregularidad enuna variable. Un claro ejemplo es el efecto del shock de precios de agosto de 1990 sobreel comportamiento de la inflación. (Sanabria, 2012).

3.6. Ejemplos Aplicativos

Ejemplo 6

Supongamos que una empresa de motos está evaluando la demanda de motores a lo largo de un año y ha registrado la siguiente demanda mensual (por unidad):

a)Pronosticar del mes 4 hasta el 13 usando un promedio móvil de 3 periodos.

Tabla 6:

Mes	DEMANDA(Unidad)	PRONOSTICO	E	l E l
1	977	[redacted]		
2	918			
3	**896**			
4	923	930.3	-7.3	7.3
5	959	912.3	46.7	46.7
6	926	926.0	0.0	0.0
7	921	936.0	-15.0	15.0
8	871	935.3	-64.3	64.3
9	840	906.0	-66.0	66.0
10	881	877.3	3.7	3.7
11	920	864.0	56.0	56.0
12	880	880.3	-0.3	0.3
13	[redacted]	893.7	[redacted]	

Fuentes: Elaboración propia

Medidas de desempeño	PROMEDIO MOVIL	
Error medio(ME)	-5	CONVIENE
Error Medio Absoluto(MAE)	29	

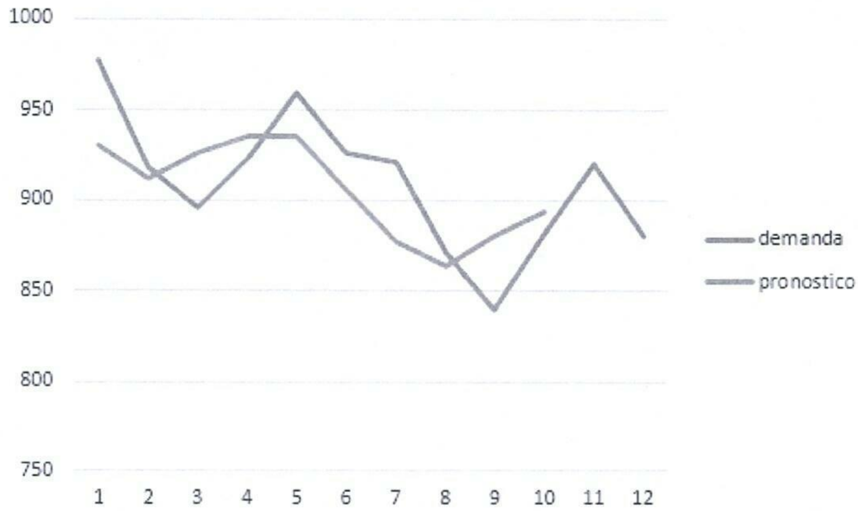
a) ¿Cuál será la demanda pronosticada para el mes 13?

La demanda pronostica para el mes 13 es de 893.7 o 894 motores

62

Gráfico 7

Relación Demanda-Pronostico



Fuentes: Elaboración propia

Ejemplo 7

Supongamos que una empresa constructora está evaluando la demanda de concreto a lolargo de un año y ha registrado la siguiente demanda mensual (en toneladas):

a) Pronosticar del mes 4 hasta el 13 usando un promedio móvil de 2 periodos.

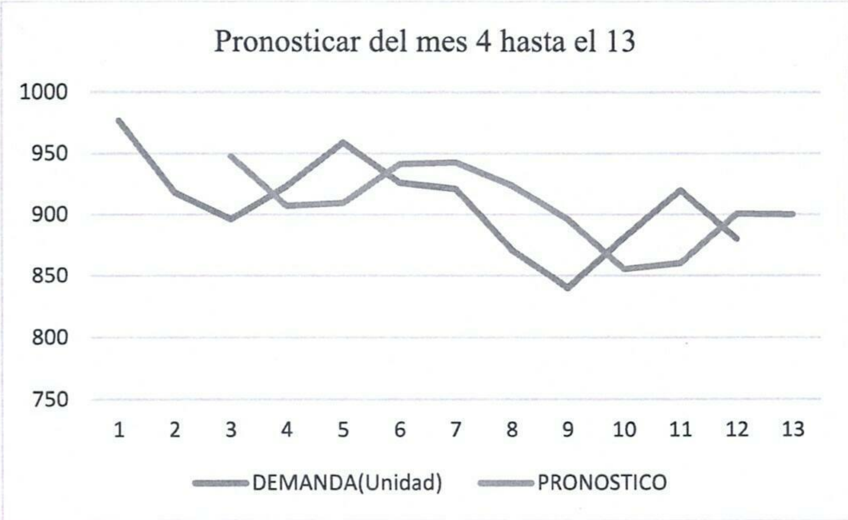
Tabla

Mes	Demanda(Unidad)	Pronóstico	E	$\|E\|$
1	977			
2	918			
3	896	947.5	-51.5	51.5
4	923	907	16	16
5	959	909.5	49.5	49.5
6	926	941	-15	15
7	921	942.5	-21.5	21.5
8	871	923.5	-52.5	52.5
9	840	896	-56	56
10	881	855.5	25.5	25.5
11	920	860.5	59.5	59.5
12	880	900.5	-20.5	20.5
13		900		

Resumen :

Medidas de desempeño	Promedio Móvil
Error medio(ME)	-7
Error Medio Absoluto(MAE)	37

Gráfico 8



Fuentes: Elaboración propia

3.7.Ciclicidad en Series de tiempo

Definición: es una serie de tiempo se refiere a las fluctuaciones que ocurren a intervalos no fijos y que tienen una duración mayor que las variaciones estacionales.Estas fluctuaciones están relacionadas con fenómenos que se repiten a lo largo del tiempo, pero no de manera regular o predecible. Los ciclos son influenciados por factores económicos, políticos, sociales, y naturales, y pueden abarcar varios años.

3.7.1. Características de Ciclicidad

1. **Duración Variable:** A diferencia de las variaciones estacionales, que ocurren enintervalos fijos y predecibles, los ciclos tienen duraciones que pueden variar significativamente.
2. **Magnitud Variable:** La amplitud de los ciclos también puede variar. Algunos ciclos pueden tener efectos profundos en la serie de tiempo, mientras que otros

pueden ser menos pronunciados

3. **No Regularidad:** Los ciclos no ocurren a intervalos regulares. Pueden ser influenciados por una variedad de factores que no son constantes en el tiempo.
4. **Relación con el Entorno Económico:** Los ciclos a menudo están asociados confases de expansión y contracción en la economía, conocidas como ciclos económicos. Estas fases incluyen períodos de crecimiento económico seguidos por recesiones.

Ejemplo 8 :Ciclicidad

1. Ciclos Económicos:

✓ Ciclo de Negocios: La economía tiende a experimentar ciclos de expansión y contracción. Un ciclo económico típico incluye una fase de expansión, un pico, una recesión y un valle. Por ejemplo, la Gran Depresión de los años 1930 y la crisis financiera global de 2008 son ejemplos de recesiones significativas en el ciclo económico.

✓ Ciclos Kondratiev: Estos son ciclos económicos de largo plazo de aproximadamente 50 a 60 años. Según la teoría de Kondratiev, las economías pasan por períodos de innovación tecnológica y crecimiento seguidos de estancamiento y recesión.

2. Ciclos naturales:

✓ Ciclos Climáticos: Fenómenos como El Niño y La Niña afectan el clima global y ocurren en ciclos irregulares de varios años. Estos ciclos pueden tener un impacto significativo en sectores como la agricultura y la pesca.

3. Ciclos en Mercados Financieros

✓ Ciclos Bursátiles: Los mercados financieros a menudo experimentan ciclosde "bull" (mercados alcistas) y "bear" (mercados bajistas). Estos ciclos reflejan el optimismo y pesimismo de los inversores y pueden durar varios años

4. Análisis de Calidad

Para analizar la ciclicidad en series de tiempo, los métodos más comunes incluyen:

1. **Descomposición de Series de Tiempo:** Separar la serie de tiempo en sus componentes básicos: tendencia, estacionalidad, ciclicidad e irregularidad. Estopermite un análisis más claro de cada componente.
2. **Modelos Econométricos:** Modelos como ARIMA (Autoregressive

59

Integrated Moving Average) pueden ser utilizados para capturar la ciclicidad en datos deseries de tiempo. El componente auto regresivo y de media móvil ayuda a modelar la correlación temporal.

3. **Análisis de Fourier:** Una técnica matemática que descompone una serie detiempo en componentes sinusoidales de diferentes frecuencias, lo que puedeayudar a identificar ciclos de diferentes duraciones.
4. **Filtros:** Filtros como el filtro Hodrick-Prescott se utilizan para extraer lacomponente cíclica de una serie de tiempo económica.

Ejemplo 9:

Ciclo de Vida de Cojinetes en Motores (Ingeniería Mecánica)

Un motor utiliza cojinetes que muestran un patrón de fallas cada cierto tiempo debido al uso continuo. Las fallas ocurren en el siguiente orden (en horas de operación):

Horas de Operación	Falla Acumulada
100	1
200	2
300	3
400	4
500	5

Objetivo:

Determinar cuántas fallas ocurrirán a las 10000 horas de operación si el patrón continúa

Solución:

1. **Tendencia:**
   Las fallas aumentan en forma lineal, con un promedio de una falla adicional cada 100 horas.
2. Proyección a 10000 horas
3. Siguiendo este patrón, se espera que las fallas acumuladas a las 1000 horas sean:

$$\text{Fallas acumuladas}= \frac{10000}{100} = 100$$

Conclusión:

Se espera que ocurran 100 fallas en total después de 10000 horas de operación, por lo que se deben proveer repuestos suficientes para esa cantidad de fallas

4. Conclusión

El análisis de series de tiempo en el contexto de probabilidad es una herramienta fundamental en estadística y análisis de datos, que permite descomponer y entender el comportamiento de datos temporales complejos. Asimismo, identificando los componentes clave de una serie de tiempo, es esencial para realizar predicciones precisas y tomar decisiones informadas.

Los principales componentes de una serie de tiempo incluyen la tendencia, la estacionalidad, los ciclos y componentes aleatorios.

Estos componentes proporcionan una visión clara y estructurada de los datos, facilitando una interpretación más precisa y una mejor toma de decisiones. La capacidad de descomponer una serie de tiempo en sus componentes fundamentales permite a los analistas identificar tendencias, ajustar por estacionalidad, entender ciclos y minimizar el impacto del ruido, mejorando así la eficacia y precisión de las predicciones.

El estudio de los componentes de una serie de tiempo en probabilidad es indispensable para cualquier análisis riguroso de datos temporales. Esta metodología no solo proporciona una comprensión más profunda de los datos, sino que también habilita la implementación de modelos predictivos más robustos y estrategias operativas más eficaces, conduciendo a decisiones más informadas y a una mejor planificación a futuro.

57

5. Bibliografía

Estadística, P. Y. (2024). *Series de tiempo*. Probabilidad y Estadística. https://www.probabilidadyestadistica.net/series-de-tiempo/#google_vignette

World Bank Open Data. (s. f.). World Bank Open Data.

Peña, D. (2005). *Análisis de series temporales*. Alianza Editorial.

Hernández, R. (2008). "Ciclos económicos y series de tiempo: una revisión crítica." *Revista de Economía Institucional*, 10(18), 215-240

Sanabria, E. M. (2012, 29 octubre). *Componentes de una serie de tiempo.*

Bowerman, B. L., & O'Connell, R. T. (1993). Pronósticos, series de tiempo y regresión: Un enfoque aplicado. Editorial Reverté

Hyndman, R. J., & Athanasopoulos, G. (2014). Análisis de series de tiempo y pronósticos. Editorial Springer

Montgomery, D. C., Jennings, C. L., & Kulahci, M. (2015). Análisis y pronóstico de series de tiempo. Editorial Wiley.

De Livera, A. M., Hyndman, R. J., & Snyder, R. D. (2011). Análisis de series de tiempo con estacionalidad y tendencia. Journal of Statistical Software, 41(1), 1-18.

Brockwell(2007) ,*Introducción a las series temporales y la previsión*

56