INGENIERÍA INDUSTRIAL
Asignatura: Estadística II **Año: 3er.**

Tema 3. Análisis de Correlación
3.1. Introducción. Análisis de Correlación
3.2. Coeficiente de Correlación de Pearson
3.3. Prueba de hipótesis de independencia de variables
3.4. Coeficiente de Correlación de Spearman
3.5. Coeficiente de Correlación de Kendall.
Objetivos: Que el alumno se capaz de:
Calcular los coeficientes de correlación, de Pearson, Spearman, y Kendall, utilizando la forma algebraica o aplicaciones digitales estadísticas, interpretando sus resultados.
Método: Elaboracion Conjunta.
Medios: Tiza y pizarra, Libro de texto, MDD, telef. moviles.
Forma de enseñanza: Clase Encuentro

- **Probabilidades y Estadística para ingenieros. J.E Freund, I.R. Miller, R. Jonhnson.**
- **Materiales Didácticos Digitales entregados por el profesor**

3.1. Introducción. Análisis de Correlación

Hasta ahora nuestro estudio se centró en el análisis de una solo variable, de manera, que a este tipo de estudio se le conoce como **análisis univariado**, pero en la mayoría de los problemas que se presentan en la vida surge la necesidad de observar el comportamiento de 2 ó más variables y estudiar la relación que existe entre ellas.

Ejemplo:

Relación entre el rendimiento laboral de un empleado con el tiempo dedicado a la capacitación
Relación entre los gastos mensuales de una familia y los ingresos mensuales de ella.
Relación entre la demanda de un producto y la calidad él.

Para realizar este tipo de estudio existen **el análisis multivariado**, que cuenta con diversas técnicas estadísticas, una de ellas es:

Análisis de Correlación.

Este análisis mide la fuerza o grado de asociación que existe entre las variables. El **análisis de correlación** intenta medir la fuerza de tales relaciones entre dos variables por medio de un solo numero denominado **coeficiente de correlación** que mide tanto la fuerza como la dirección de la relación. En el estudio se abordan tres coeficientes de correlación: Pearson, Spearman, Kendall

3.2. Coeficiente de correlación de Pearsón (rho), evalúa la relación lineal entre dos variables continuas. Una relación es lineal cuando un cambio en una variable se asocia con un cambio proporcional en la otra variable. Por ejemplo, usted puede usar una correlación de Pearson para evaluar si las temperaturas en las instalaciones de producción están asociados con los accidentes laborales en ellas

Se parte de los supuestos siguientes:

- Los datos siguen una distribución Normal.
- Son independientes unos de otros.
- No existen datos anómalos

Expresión matemática:

$$r_{xy} = \frac{n\sum X * Y - \left(\sum X\right)\left(\sum Y\right)}{\left[\sqrt{n\sum X^2 - \left(\sum X\right)^2}\right]\left[\sqrt{n\sum Y^2 - \left(\sum Y\right)^2}\right]}$$

En el Análisis de Correlación, es oportuno interpretar sus valores. Para ello se analizará la siguiente tabla.

Parámetro (Población)	**Estadístico (Muestra)**	**Valores del coeficiente**	**Grado de la asociación**	**Ejemplo**
ρ	r	-1	Fuerte asociación negativa	Ahorro personales de salario mensual y gastos personales
		0	No hay asociación	Ahorro personales de salario mensual y color del pelo
		1	Fuerte Asociación positiva	Calidad de un producto y demanda del producto

Valor r	Fuerza de la correlación
0.0 < 0.1	no hay correlación
0.1 < 0.3	poca correlación
0.3 < 0.5	correlación media
0.5 < 0.7	correlación alta
0.7 < 1	correlación muy alta

Ejemplo 1

Suponga que quiere analizar si existe relación entre la velocidad de una línea de producción (X) y el número de defectos por lote (Y). Se recolectaron los siguientes datos:

X: velocidad de una línea de producción

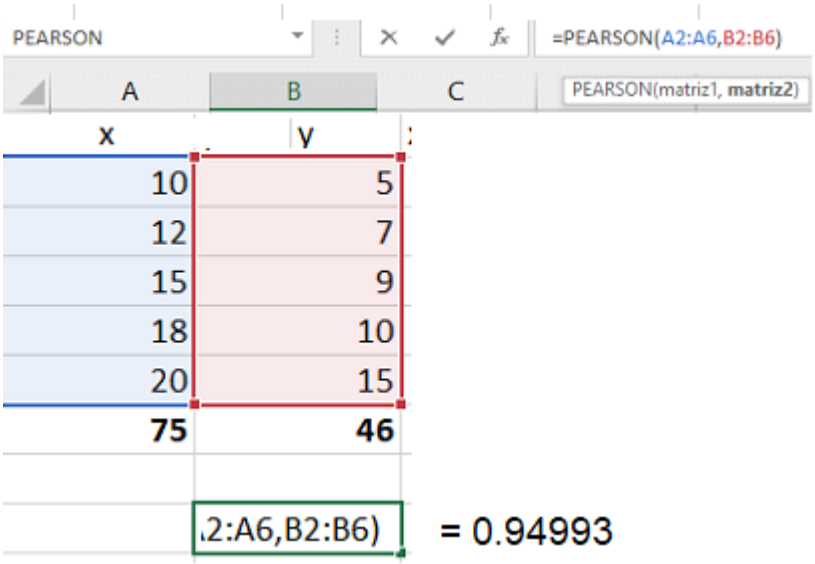
Y: número de defectos por lotes

Velocidad (m/min	Defectos (Y)	XY	X^2	Y^2
10	5	50	100	25
12	7	84	144	49
15	9	141	225	81
18	10	180	324	100
20	15	300	400	225
75	**46**	**749**	**1193**	**480**

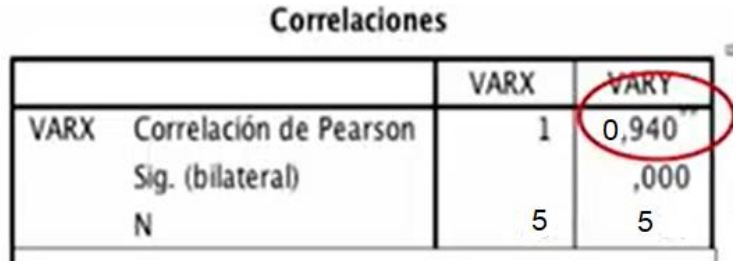
Solución: Algebraica

$$r = \frac{n\sum xy - \left(\sum x\right)\left(\sum y\right)}{\sqrt{n\sum x^2 - \left(\sum x\right)^2} * \sqrt{n\sum y^2 - \left(\sum y\right)^2}} \quad r = \frac{5\sum 749 - \left(\sum 75\right)\left(\sum 46\right)}{\sqrt{5\sum 1193 - \left(\sum 75\right)^2} * \sqrt{5\sum 480 - \left(\sum 46\right)^2}} = 0.94993$$

Solución empleando Excel



Solución con otros paquetes estadísticos Ejemplo con SPSS



Correlaciones

		VARX	VARY
VARX	Correlación de Pearson	1	0,940**
	Sig. (bilateral)		,000
	N	5	5

R/ Existe una fuerte correlación positiva entre la velocidad de la línea de producción y el número de defectos por lotes.

Nota: es importante señalar que la correlación entre variables no significa causalidad. Ejemplo cada año se gasta más en ciencia y esto no es la causa que también aumenta la temperatura promedio del planeta.

Pero esa relación se da para la muestra de datos seleccionada, pero si quisiéramos ver si sucede igual a toda la población tendríamos que hacer una prueba de hipótesis de correlación

3.3. Prueba de Hipótesis de independencia (no relación de variables)

1er Paso: Planteamiento de la Hipótesis.

H_o: ρ = 0 las variables son independientes (no hay relación entre las variables)
H_i: $\rho \neq 0$ las variables son dependientes (hay relación entre las variables)

2Do Paso: Fijar el Nivel de significación α

α= 0.05

3er Paso: Regla de decisión

Si pvalor < α Rechazo Ho

4to Paso: Toma de decisión.

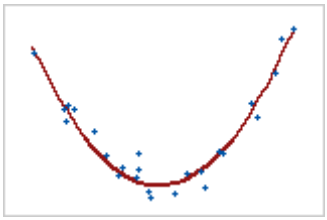
0.000 < 0.05 entonces se rechaza Ho.

R/ Existe una correlación entre las variables velocidad de la línea de producción y el número de defectos por lotes, con una confianza del 95%

3.4. Coeficiente de correlación de Spearman (r_s).

Características de su empleo

- Mientras que el coeficiente de correlación de Pearson solo miden relaciones lineales en cambio el coeficiente de correlación de Spearman (r_s) mide la relación monótona (no necesariamente lineal)



- Puede medir la relación entre dos variables cuantitativa y también puede medir esa relación entre una variable cuantitativa y otra cualitativa.
- Los dato no tiene que seguir una distribución Normal.
- Ideal para datos ordinales o no normales.
- No se afecta ante datos anómalos.

Expresión Matemática

$$r_s = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

Pasos para el cálculo del coeficiente de correlación de Spearman (s_r).

1. Recoger los datos.
2. Asignar rangos.
3. Calcular la diferencia de rangos (d).
4. Elevar la diferencia (d) al cuadrado.

5. Aplicar la expresión matemática o fórmula.

Ejemplo 2

Se quiere evaluar si el nivel de experiencia de operadores (X) está relacionado con el tiempo de ensamblaje (Y). Los datos son:

Experiencia	Tiempo	Rango X	Rango Y	Diferencia de rango (d)	d^2
Bajo	25	1	3	-2	4
Media	20	2	2	0	0
Alto	15	3	1	2	4
				total	8

$$r_s = 1 - \frac{6\sum d_i^2}{n(n^2-1)} \qquad r_s = 1 - \frac{6\sum 8}{3(3^2-1)} = -1$$

R/ Existe una correlación negativa perfecta entre la experiencia de los operadores y el tiempo de ensamblaje, a medida que aumenta la experiencia disminuye el tiempo de ensamblaje

Ejemplo 3.

Analizar si existe una relación entre el número de horas que un grupo de estudiantes universitarios dedica al estudio y sus notas finales en un examen.

1. Recoger los datos:

Estudiante	Horas de Estudio (X)	Nota Final (Y)
A	4	78
B	6	85
C	2	72
D	8	88
E	5	80

2. Convertir los Datos en Rangos:

Estudiante	Horas de Estudio (X)	Rango de X	Nota Final (Y)	Rango de Y
A	4	2	78	2
B	6	4	85	4
C	2	1	72	1
D	8	5	88	5
E	5	3	80	3

3. Calcular las Diferencias de Rangos:
4. Elevar al cuadrado las Diferencias de Rango:

Estudiante	Rango de X	Rango de Y	d_i	$(d_i)^2$
A	2	2	0	0
B	4	4	0	0
C	1	1	0	0
D	5	5	0	0
E	3	3	0	0

5. Aplicar la formula: $r_s = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$

$$r_s = 1 - \frac{6 \times 0}{5(5^2 - 1)}$$

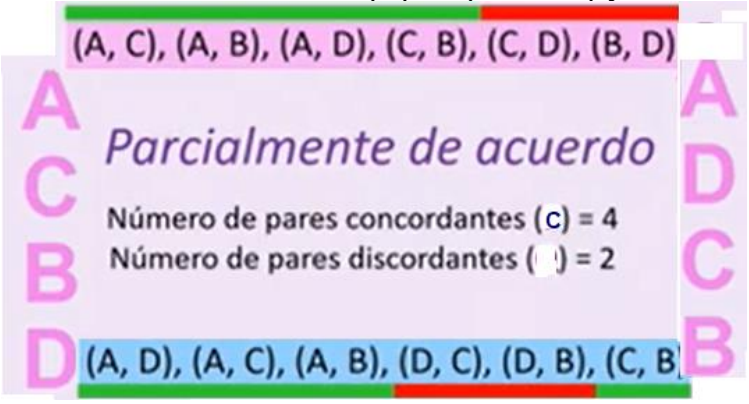
$$r_s = 1 - \frac{0}{120} = 1$$

R/ Existe una relación perfecta positiva entre las horas de estudio de los estudiantes y las notas que obtienen

3.5 Coeficiente de correlación de Kendall (tau), comúnmente conocido como coeficiente τ de Kendalll (con la letra griega τ, tau), es utilizado para medir **la asociación ordinal entre dos cantidades medidas**. Por lo que los datos no tienen que estar normalmente distribuidos.

Por ejemplo, dos inspectores diferentes examinan 45 Tv para determinar los defectos encontrados en las pantallas. ¿Concuerdan consistentemente los inspectores en su clasificación?

Ideal para medir la similitud entre dos conjuntos de ranking ejemplo, dos evaluadores determinan la calidad de 4 marcas de equipos (A,B,C,D) y la ordenan según su ranking.



Podemos calcular la **tau de Kendall** con esta **fórmula**:

$$\tau = \frac{C - D}{C + D}$$

Donde **C** es el número de pares concordantes

y **D** es el número de pares discordantes

τ = (4-2) / (4+2) = 2 / 6 = 0.33

R/ Existe una cierta concordancia entre las evaluaciones de ambos evaluadores.

El Coeficiente de Kendal es muy similar que el de Spearman, pero se deberia preferir en el caso que se disponen de muy pocos datos y hay muchos empates de rangos.

CONCLUSIONES

- *Pearson:* Ideal para relaciones lineales en datos paramétricos (ej: control de calidad).

- *Spearman:* Para tendencias no lineales (ej: rankings de eficiencia).

- *Kendall:* Útil con muestras pequeñas o empates (ej: logística).

ESTUDIO INDEPENDIENTE.

En un proceso de fabricación, se cree que a mayor temperatura (X), mayor es la resistencia (Y) del material. Calcule e interprete el coeficiente de correlación de Pearson y de Spearman con los datos:

(X): 100, 120, 150, 180 °C

(Y): 50, 55, 60, 65 MPa