

Correlación

Xochitl Cárdenas

Puedes encontrar esta presentación y materiales adicionales en mi perfil de Github:
<https://github.com/xochitlcardenas>



Objetivos de aprendizaje

1. Definir relación, covarianza y coeficiente de correlación.
2. Identificar estrategias para determinar si existe una relación, exponiendo ventajas y desventajas.
3. Calcule el valor de Pearson r .
4. Interpretar el coeficiente de pearson a partir de 3 criterios: fuerza, dirección y significancia.
5. Explicar por qué correlación no necesariamente implica causalidad.
6. Reforzar lo aprendido con un ejercicio.

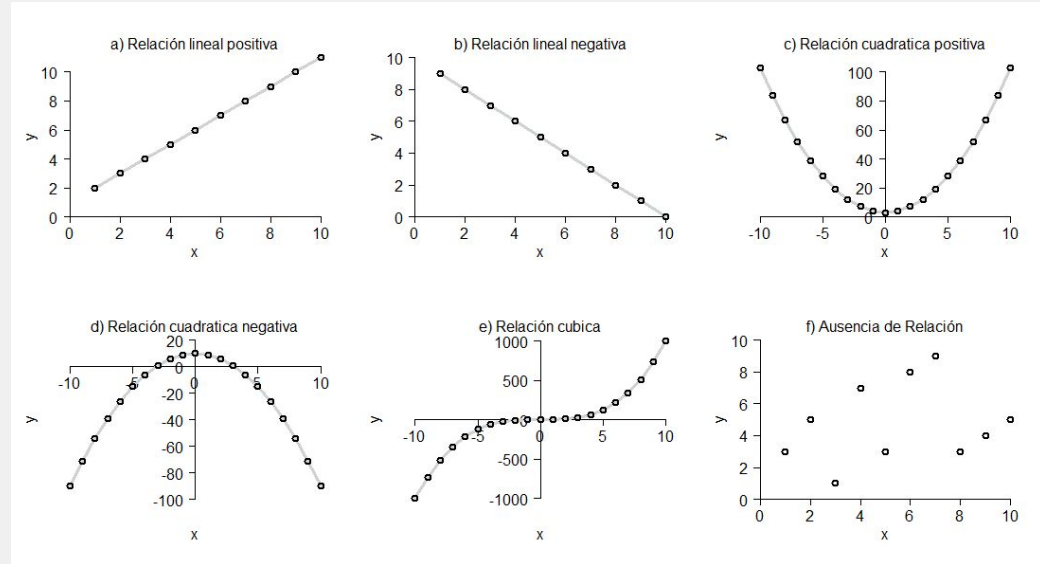
Estadística Inferencial

		Tipos de pruebas					
		Relación o asociación	Comparación / diferencias				Predicción o estimación
				Muestras independientes		Muestras relacionadas	
	Nivel de medición de la variable dependiente			1 o 2 gpos	Más de 2 gpos	2 mediciones	Más de 2 mediciones
No paramétrica	Nominal		X²	X²	Mc Nemar	Q de Cochran	
	Ordinal	Spearman	U de Mann Whitney	Kruskall Wallis	Wilcoxon	Friedman	
Paramétrica	Intervalar	Pearson	t de Student	ANOVA	t de Student	ANOVA	Regresión lineal
	De razón						

Coeficiente de Correlación de Pearson

Diagrama de dispersión

- Identificar el tipo de relación (forma)
- Punto (X_i, Y_i)
- No refleja el grado de la relación



Coeficientes de correlación

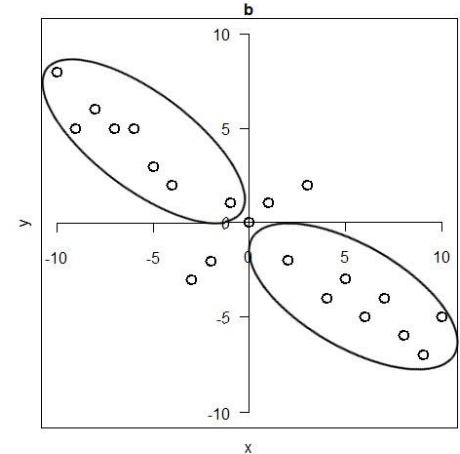
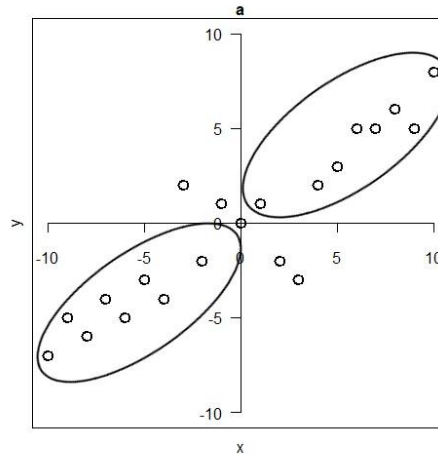
-Expresa cuantitativamente la magnitud y dirección de la relación.

Estrategías:

- Proporción de puntos
- Puntuaciones diferenciales $x = (X - X')$.

Igualmente fallan al tratar de reflejar la magnitud exacta, por lo que se proponen las siguientes:

- $\sum x^2$ y covarianza
- **Coeficiente de correlación de Pearson**



Covarianza

Sujetos	X(Sueño)	Y(Estrés)	x	y	xy
1	1	6	-1.6	-22.9	36.64
2	1	7	-1.6	-21.9	35.04
3	2	19	-0.6	-9.9	5.94
4	2	25	-0.6	-3.9	2.34
5	2	23	-0.6	-5.9	3.54
6	3	37	0-4	8.1	3.24
7	3	30	0-4	1.1	0.44
8	4	45	1.4	16.1	22.54
9	4	47	1.4	18.1	25.34
10	4	50	1.4	21.1	29.54
Medias	2.6	28.9		Σ	164.6

‘La relación lineal entre 2 variables es tanto más intensa cuanto mayor es la suma de esos productos (en valor absoluto)’

Incorporar el tamaño de la muestra: covarianza

$$S_{XY} = \frac{\Sigma x_i y_i}{n-1} = \frac{164.6}{9} = 18.28$$

Un valor de covarianza igual a cero indica ausencia de relación, sin embargo, este método no cuenta con un límite superior que indique una relación perfecta.

Solución: relativizar la covarianza

Coeficiente de correlación de Pearson

Sujetos	X(Sueño)	Y(Estrés)	XY	X ²	Y ²
1	1	6	6	1	36
2	1	7	7	1	49
3	2	19	38	4	361
4	2	25	50	4	625
5	2	23	46	4	529
6	3	37	111	9	1369
7	3	30	90	9	900
8	4	45	180	16	2025
9	4	47	188	16	2209
10	4	50	200	16	2500
SUMATORIAS	26	289	916	80	10603
PRODUCTOS	ΣX=	7514			
CUADRADOS	676	83521			

$$R_{XY} = \frac{S_{XY}}{S_X S_Y} \quad S_X = \sqrt{\frac{\Sigma(X_i - \bar{X})^2}{n-1}}$$

La ecuación se interpreta como el grado en que la covarianza alcanza su máximo.

$$R_{XY} = \frac{n\Sigma X_i Y_i - \Sigma X_i \Sigma Y_i}{\sqrt{n\Sigma X_i^2 - (\Sigma X_i)^2} \sqrt{n\Sigma Y_i^2 - (\Sigma Y_i)^2}}$$

$$R_{XY} = \frac{10(916) - 7514}{\sqrt{10(80) - 676} \sqrt{10(10603) - 83521}} = \frac{1646}{1669.72} = 0.985$$

Coeficiente de correlación de Pearson

Valor	Interpretación
0	Relación nula
0 - 0.2	Relación muy baja
0.2 - 0.4	Relación baja
0.4 - 0.6	Relación moderada
0.6 - 0.8	Relación alta
0.8 - 1	Relación muy alta
1	Relación perfecta

Signo	Dirección
+	Positiva
-	Negativa

Signo : positivo o negativo

Valor entre -1 y 1

$$R_{XY} = 0.985$$

Relación muy alta

Hipótesis nula de independencia lineal

Probabilidad de aceptar o rechazar la hipótesis nula. De acuerdo a ello se debe planear la regla de decisión.

- 1. Plantear la hipótesis y determinar el nivel de significancia

H_0 : No existe correlación significativa

H_1 : Existe correlación significativa

- 2. Comparar el valor R contra el valor crítico R de tablas.
- 3. Regla de decisión

$R_t \geq R$, se acepta la H_0 y se rechaza H_1 , las variables no están relacionadas

$R_t < R$, se rechaza la H_0 y se acepta H_1 , las variables están relacionadas

Hipótesis nula de independencia lineal

$$N = 10$$

$$gl = N - 2 = 10 - 2 = 8$$

$$R_{XY} = 0.985 > R_t = 0.6319$$

Por lo tanto, rechazamos H_0
y aceptamos H_1 .

Existe correlación entre las
variables.

$df = N - 2$	Level of Significance for One-Tailed Test, α_1 tail				
	.05	.025	.01	.005	.0005
	Level of Significance for Two-Tailed Test, α_2 tail				
	.10	.05	.02	.01	.001
1	.9877	.9969	.9995	.9999	1.0000
2	.9000	.9500	.9800	.9900	.9990
3	.8054	.8783	.9343	.9587	.9912
4	.7293	.8114	.8822	.9172	.9741
5	.6694	.7545	.8329	.8745	.9507
6	.6215	.7067	.7887	.8343	.9249
7	.5822	.6664	.7498	.7977	.8982
8	.5494	.6319	.7155	.7646	.8721
9	.5214	.6021	.6851	.7348	.8471
10	.4973	.5760	.6581	.7079	.8233
11	.4762	.5529	.6339	.6835	.8010
12	.4575	.5324	.6120	.6614	.7800
13	.4409	.5139	.5923	.6411	.7603
14	.4259	.4973	.5742	.6226	.7420
15	.4124	.4821	.5577	.6055	.7246
16	.4000	.4683	.5425	.5897	.7084
17	.3887	.4555	.5285	.5751	.6932
18	.3783	.4438	.5155	.5614	.6787
19	.3687	.4329	.5034	.5487	.6652
20	.3598	.4227	.4921	.5368	.6524

Correlación

Para identificar el nivel de relación entre dos variables podemos hacer uso de un estadístico que cuantifica el grado de relación, y este puede interpretarse de acuerdo con 3 puntos:

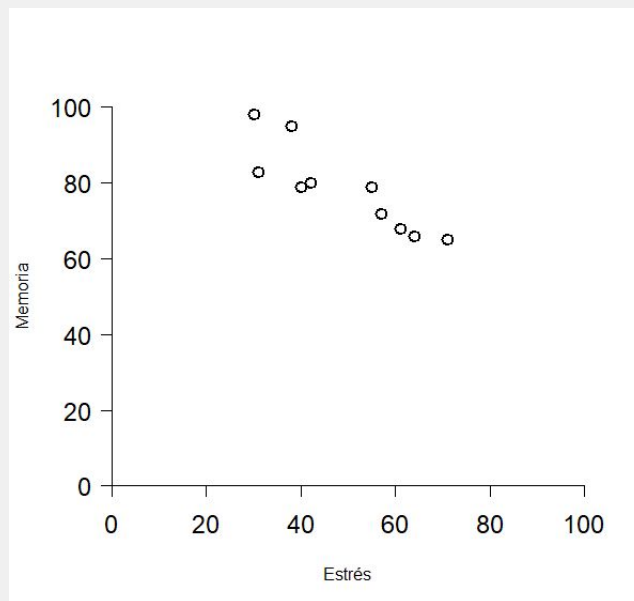
1. Fuerza (valor arrojado por coeficiente de Pearson)
2. Dirección (signo del coeficiente)
3. Significancia (prueba de hipótesis nula)

Correlación no implica causalidad.

Ejercicio 1

Un grupo de investigadores desea conocer si existe relación entre el estrés percibido y la memoria. Para esto, comparan la puntuación obtenida en una escala de estrés percibido y la puntuación obtenida en una prueba de memoria de 10 sujetos.

X (estrés)	Y (Memoria)
64	66
40	79
30	98
71	65
55	79
31	83
61	68
42	80
57	72
38	95



Ejercicio 1

1. Ubicar los puntos en un diagram de dispersión
2. Obtener el valor R
 - a. Elevar al cuadrado cada valor de X y Y
 - b. Obtener los productos de X y Y
 - c. Efectuar las sumatorias $\sum X$, $\sum Y$, $\sum X^2$, $\sum Y^2$, $\sum XY$
 - d. Aplicar la ecuación

$$R_{XY} = \frac{n\sum X_i Y_i - \sum X_i \sum Y_i}{\sqrt{n\sum X_i^2 - (\sum X_i)^2} \sqrt{n\sum Y_i^2 - (\sum Y_i)^2}}$$

- e. Interpretar valor de R obtenido
3. Comprobar relación por hipótesis nula
 - a. Calcular grado de libertad $gl = N-2$
 - b. Identificar valor de R en tablas y comparar con valor obtenido
 - c. Aplicar regla de decisión

$R_t \geq R$, se acepta la H_0 y se rechaza H_1 , las variables no están relacionadas

$R_t < R$, se rechaza la H_0 y se acepta H_1 , las variables están relacionadas

X (estrés)	Y (Memoria)	X^2	Y^2	XY
64	66	4096	4356	4224
40	79	1600	6241	3160
30	98	900	9604	2940
71	65	5041	4225	4615
55	79	3025	6241	4345
31	83	961	6889	2573
61	68	3721	4624	4148
42	80	1764	6400	3360
57	72	3249	5184	4104
38	95	1444	9025	3610
$\Sigma = 489$	$\Sigma = 785$	$\Sigma = 25801$	$\Sigma = 62789$	$\Sigma = 37079$

$$R_{XY} = \frac{n\sum X_i Y_i - \sum X_i \sum Y_i}{\sqrt{n\sum X_i^2 - (\sum X_i)^2} \sqrt{n\sum Y_i^2 - (\sum Y_i)^2}}$$

$$R_{XY} = \frac{10(37079) - (489 \cdot 785)}{\sqrt{10(25801) - (489)^2} \sqrt{10(62789) - (785)^2}} = \frac{-13075}{14843.07} = -0.88$$

Valor	Interpretación
0	Relación nula
0 - 0.2	Relación muy baja
0.2 - 0.4	Relación baja
0.4 - 0.6	Relación moderada
0.6 - 0.8	Relación alta
0.8 - 1	Relación muy alta
1	Relación perfecta

Signo	Dirección
+	Positiva
-	Negativa

$$R_{XY} = -0.88$$

Interpretación:
Relación negativa
muy alta

Tabla de valores críticos para R de Pearson, nivel de significancia para dos colas

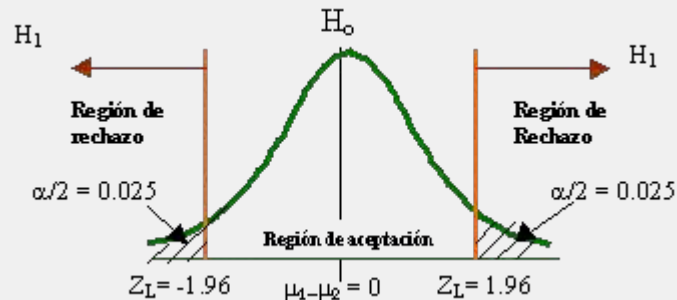
$df = N - 2$	Level of Significan	
	.05	.025
	Level of Significan	
	.10	.05
1	.9877	.9969
2	.9000	.9500
3	.8054	.8783
4	.7293	.8114
5	.6694	.7545
6	.6215	.7067
7	.5822	.6664
8	.5494	.6319
9	.5214	.6021
10	.4973	.5760

$$gl = 10 - 2 = 8$$

$$R_{XY} = 0.88 > R_t = 0.63$$

$R_t \geq R$, se acepta la H_0 y se rechaza H_1 , las variables no están relacionadas

$R_t < R$, se rechaza la H_0 y se acepta H_1 , las variables están relacionadas



Por lo tanto, podemos afirmar que existe una relación negativa muy alta entre las variables, la cual es significativa.

JASP

Programa para análisis estadístico.

- Análisis frecuentistas
- Análisis bayesianos

Descargar JASP:

<https://jasp-stats.org/>

Video Introductorio:

<https://youtu.be/HxqB7CUA-XI>