

ANÁLISIS ESTADÍSTICO

CORRELACIÓN LINEAL

Jorge Fallas
jfallas56@gmail.com

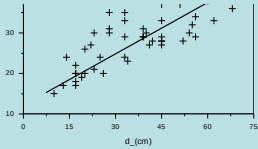
2010

Temario

- **Concepto de correlación**
- **Diagramas de dispersión**
- **Correlación: dirección, intensidad**
- **Coef. Correlación lineal de Pearson**
 - **Prueba de hipótesis: significancia de R**
- **Coef. Correlación Spearman**
 - **Prueba de hipótesis: significancia de R**
- **Software**
 - **XLStats**

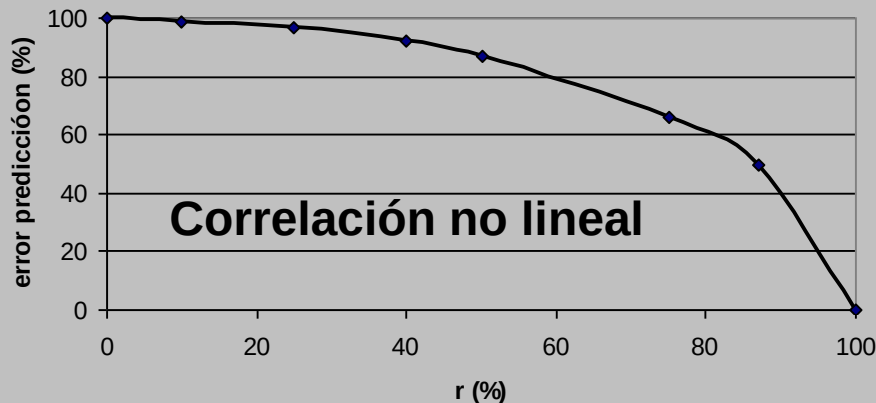
Concepto correlación

Correlación lineal



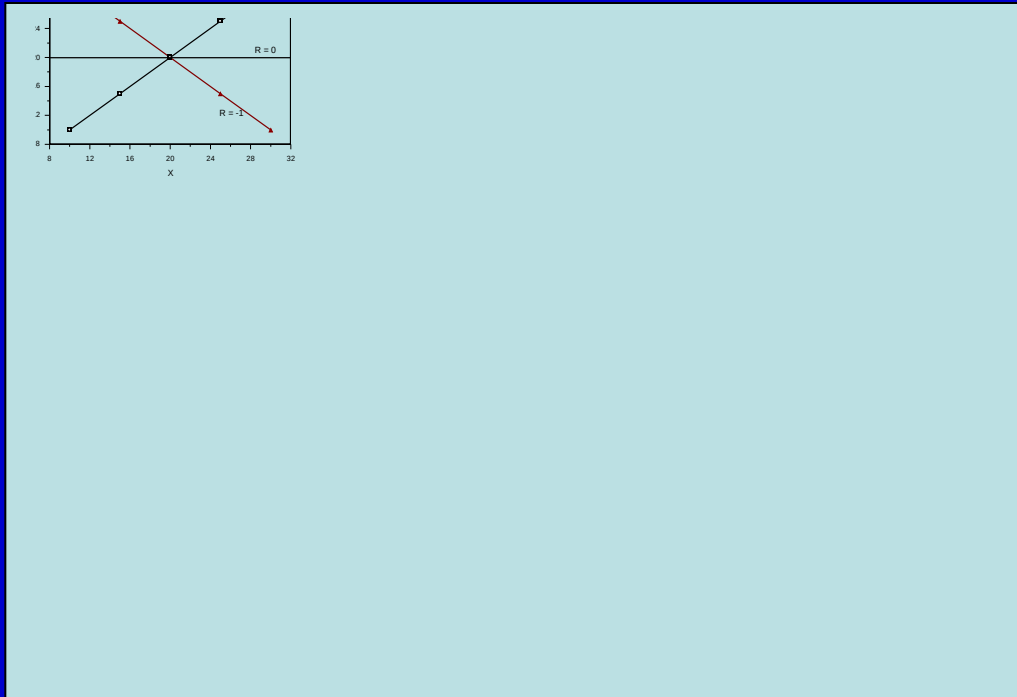
- Asociación entre dos variables
- La asociación puede ser lineal ó no lineal
- Correlación no es sinónimo de causa-efecto
- Explicación de asociación debe sustentarse en argumentos biológicos, físicos
- Puede medirse correlación a nivel paramétrico y no paramétrico

Correlación no lineal



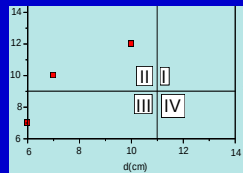
- Datos intervalo y razón. R de Pearson
- Datos ordinal Ro de Spearman
- Datos nominal: Chi cuadrado

Diagrama de dispersión



- Grado de correlación entre dos variables
- La relación es imperfecta. No siempre valores grandes de X corresponden a valores grandes de Y
- La naturaleza de la asociación puede describirse adecuadamente por una recta.
- El valor de “ r ” oscila entre 1 y -1. Cuando la asociación entre X y Y es perfecta “ r ” es igual a 1 ó -1 ; cuando no existe correlación lineal su valor es 0.

Intensidad de correlación y ubicación de pares ordenados con respecto a media



Si los datos se ubican en los cuadrantes I y III el valor de r será positivo y alto

Si los datos se ubican en los cuadrantes II y IV el valor de r será negativo y alto

¿De qué depende la intensidad de la correlación?

- **Principio de linealidad**

Cuanto mayor sea el número de observaciones que se ubiquen alrededor de la línea de mejor ajuste mayor será el valor de r .

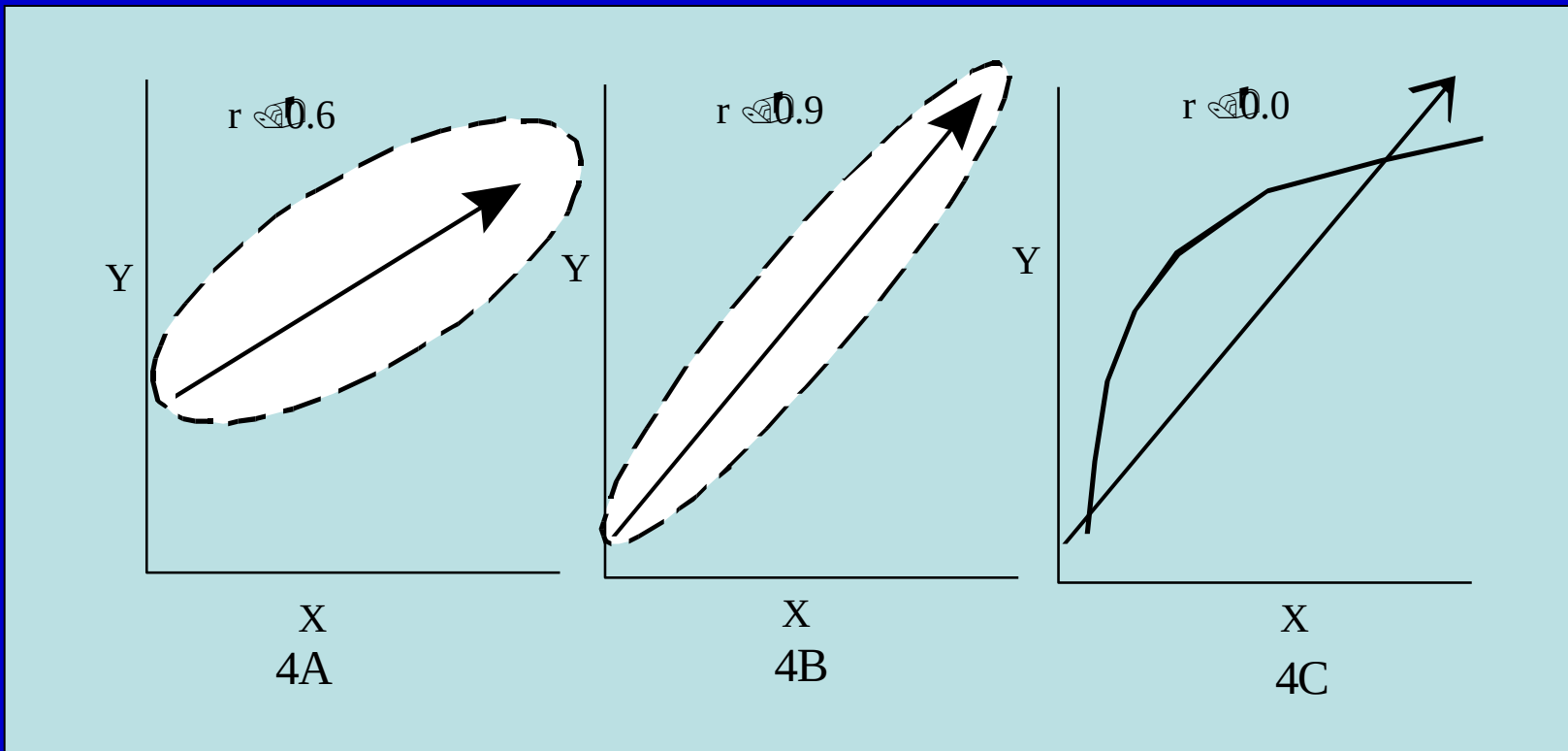
- **Ámbito de las observaciones**

El valor del coeficiente de correlación depende del grado de variabilidad de los datos (X,Y) . Si mantenemos constantes los otros factores que afectan el valor de “ r ”, restringiendo el ámbito de X y Y reduciremos el valor del coeficiente de correlación.

- **Unidad de medición y codificación de los datos**

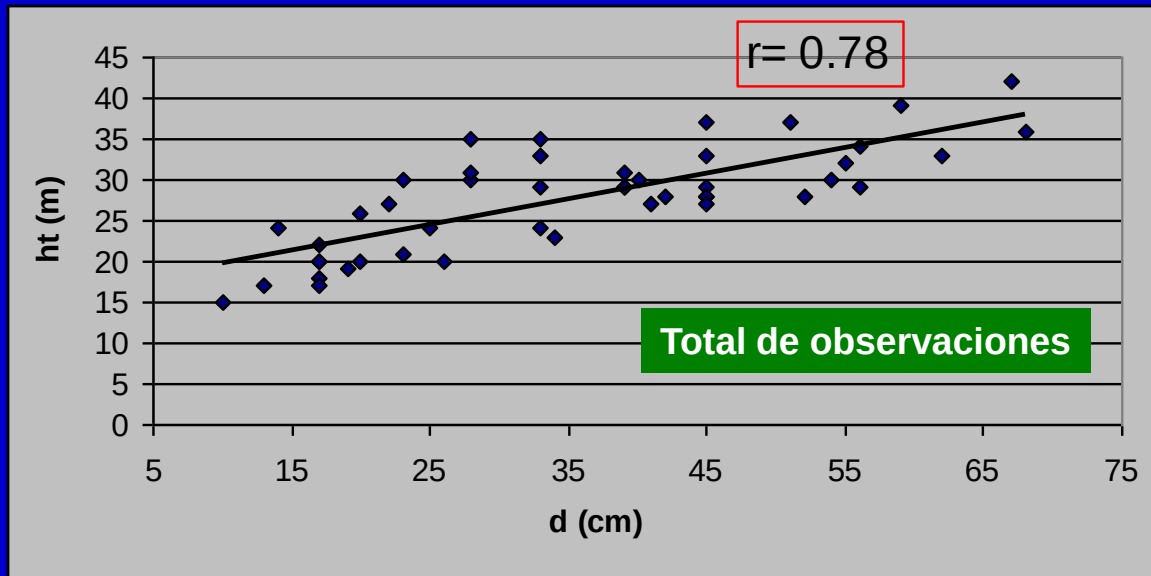
El valor de r refleja el grado de correspondencia entre las posiciones relativas de los pares X,Y y por tanto su valor no es afectado por transformaciones que impliquen el uso de una constante ó de operaciones algebraicas tales como suma, resta, multiplicación y división.

Principio linealidad



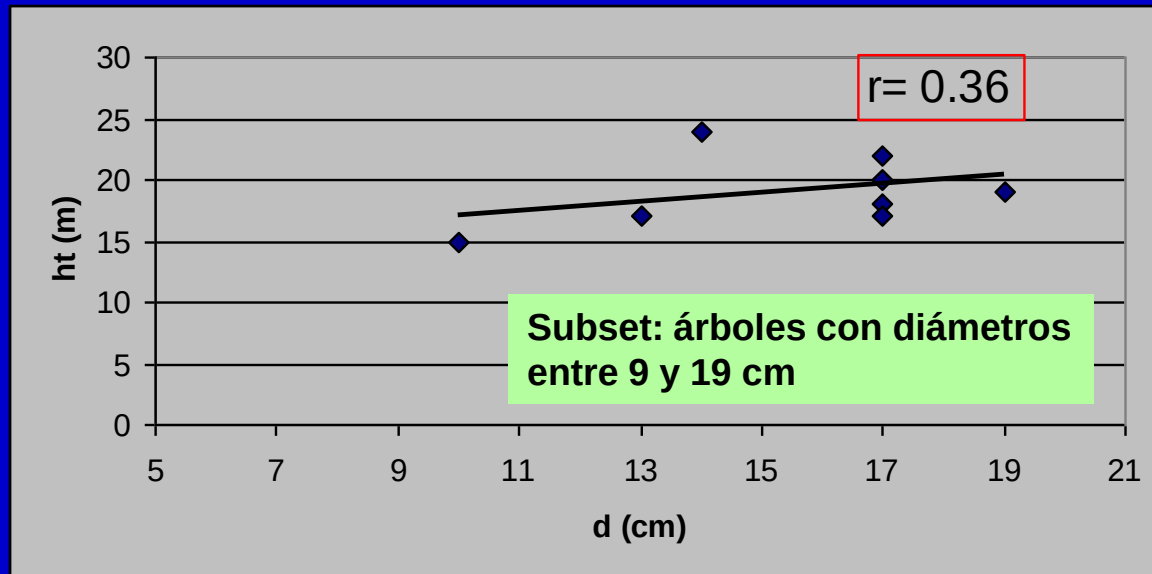
4A: Correlación moderada. 4B: Correlación alta. 4C: Correlación no lineal

Ámbito de observaciones



El coeficiente de correlación para la totalidad de las observaciones es 0.78, en tanto que para el subset es 0,36.

Esto nos indica que, dada una relación lineal, la magnitud del coeficiente de correlación depende del grado de variabilidad de los datos.



Si mantenemos constantes los otros factores que afectan el valor de " r ", restringiendo el ámbito de X y Y reduciremos el valor del coeficiente de correlación.

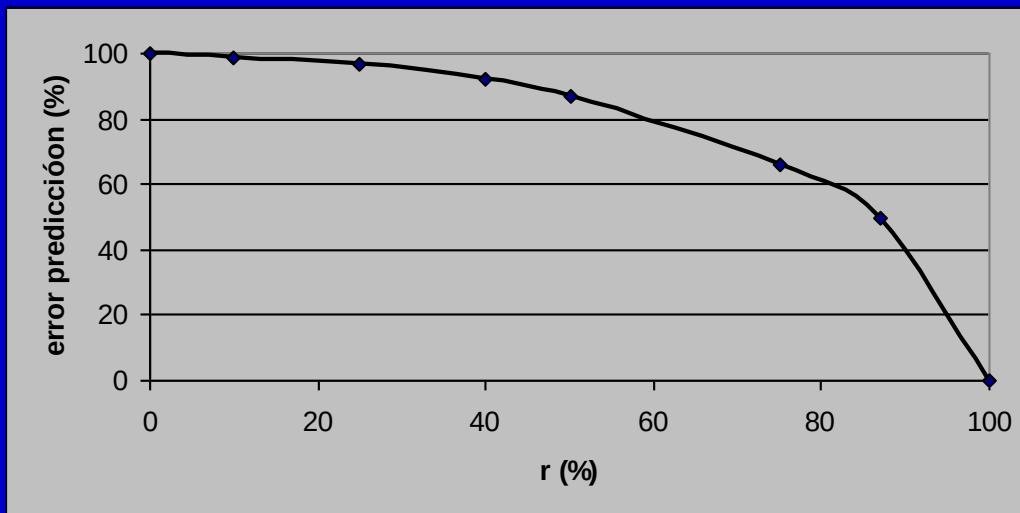
Interpretación r

- **Valor de r con respecto a 0, 1 y -1**

0: ausencia de correlación

1: Correlación positiva y perfecta

-1: Correlación negativa y perfecta



Relación entre r y error de predicción es no lineal

Incremento en valor de r es más importante cuanto más cercano se encuentre de 1

Proporción de la variabilidad explicada

Coeficiente correlación %	Error de predicción % *	Variabilidad explicada %
100	0	100
87	50	50
75	66	34
50	87	13
40	92	8
25	97	3
10	99	1
0	100	0

* $Sy_x = S_y (1 - r^2)^{1/2}$

Incrementos en valores de r son más importantes cuanto más cercanos se encuentren de 1 ó -1

Ej. peso y tamaño huella

Inferencia sobre R

Los siguientes estadísticos pueden utilizarse para probar por la significancia de r:

$$F_{1,n-2} = \frac{r^2 (n-2)}{1-r^2}$$

ó

$$t_{n-2} = \sqrt{\frac{r^2 (n-2)}{1-r^2}}$$

RECUERDE QUE EL TAMAÑO DE MUESTRA ES MUY IMPORTANTE EN PRUEBAS DE SIGNIFICANCA!

PROCEDIMIENTO

- Seleccione una muestra aleatoria de tamaño n
- Calcular coeficiente de correlación
- Inferencia sobre r
 - Intervalo de confianza
 - Prueba de hipótesis
 - Elegir en ambos casos nivel de confianza antes de realizar las pruebas ($1 - \alpha$)
- Aplicación práctica

RECORDAR CORRELACIÓN NO ES SINÓNIMO DE CAUSA- EFECTO

Correlación en XLSTats

1. Seleccione el *Menú Principal*

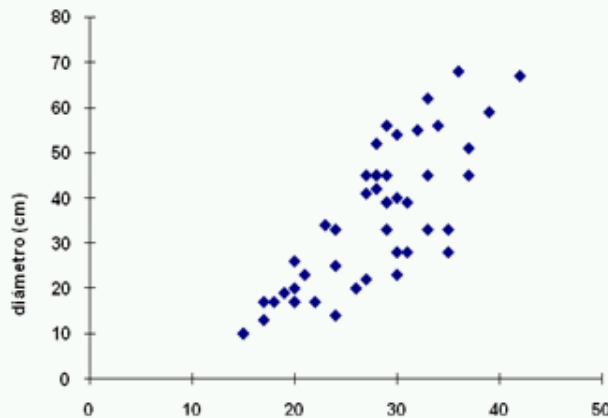
2 Numerical

2Num

Data	
diámetro (cm)	altura total (m)
40	30
10	15

2. Copie los datos en las columnas de datos

3. El programa le muestra el diagrama de dispersión para las variables diámetro-altura



4. Seleccione la hoja **Corr & Linear Regress** y configure la ventana de diálogo como se muestra a continuación:

Correlation Analysis

Correlation Coeff

Correlation 0.781847

Hypothesis Tests for ρ

$H_0: \rho = 0$

Alternative

☒ $\neq$

☐ $>$

☐ $<$

$H_1: \rho \neq 0$

p-value = 1.41E-10

$H_0: \rho = 0$ (No existe correlación lineal entre las variables diámetro-altura total)

$H_a: \rho \neq 0$ (Existe correlación lineal entre las variables diámetro-altura total)

5. **Conclusión:** El valor de “p” calculado (0.000) es inferior al valor del “p” crítico (0.05) y por tanto se rechaza H_0 y se concluye que las variables están linealmente correlacionadas.

Tamaño de muestra y significancia de r

Pares (X,Y)	g.l.	R	Ho: R = 0 (2 colas alfa=0,05) Ha: R <>0
12	10	0,55	<i>no rechazar</i>
18	16	- 0,45	<i>no rechazar</i>
22	20	0,45	rechazar Ho
102	100	0,21	rechazar Ho
500	498	- 0,15	rechazar Ho
1002	1000	0,10	rechazar Ho

En muestras muy grandes (Ej. 500 y 1000) coeficientes de correlación relativamente pequeños serán declarados como significativos en tanto que con muestras pequeñas sucederá lo opuesto.

Coeficiente de correlación de órdenes de Spearman

- El coeficiente de correlación de Spearman (Rho) se utiliza para cuantificar la intensidad y dirección de la correlación cuando las variables se miden o se transforman a un nivel de medición ordinal.
- Existen varias fórmulas
- Interpretación similar a r de Pearson : 0,1 y -1

Cálculo de Rho utilizando XLSTats

1. Seleccione del *Menú 2Num* la hoja

Extra Tools

Ordinal Variables

- Spearman's rho

- Kendall's tau-b



2. Seleccione la opción y configure la ventana para una prueba de dos colas y un alfa de 0.05.

3. El resultado es el siguiente:

Measures of Association for Ordinal Variables	
Spearman's Rank Coefficient (rho)	
Correlation Coeff ρ 0.74987615	Large sample Hypothesis Test for ρ $H_0: \rho = 0$ Alternative ☒ $\neq$ ☐ $>$ ☐ $<$ $H_1: \rho \neq 0$ T 7.51856 DF 44 p-value = 2E-09
Kendall's Rank Correlation Coefficient (tau-b)	
Correlation Coeff τ_b 0.58019112	Large sample Hypothesis Test for τ_b $H_0: \tau_b = 0$ Alternative ☒ $\neq$ ☐ $>$ ☐ $<$ $H_1: \tau_b \neq 0$ Z 5.663047 p-value = 1.49E-08

Ho: Rho = 0

Ha: Rho $\neq$ 0

Rho de Spearman

Rechazar Ho si P calculado es menor que P crítico

Tau-b de Kendall

4. **Conclusión:** El valor de "p" calculado (0.000) es inferior al valor del "p" crítico (0.05) y por tanto se rechaza Ho y se concluye que las variables están linealmente correlacionadas.

File Analysis Help

Bootstrapping Procedures ▶

Randomization Tests ▶

Bootstrapped Mean

Bootstrap Correlation ←

Bootstrapped Median

Bootstrapping Correlations

File Analysis Help

NReps

☒ 1,000☐ 2,000☐ 5,000☐ 10,000☐ 50,000

CI %

☐ 90 %☒ 95 %☐ 99 %

Obtained r

-0.206

Lower Conf. Limit

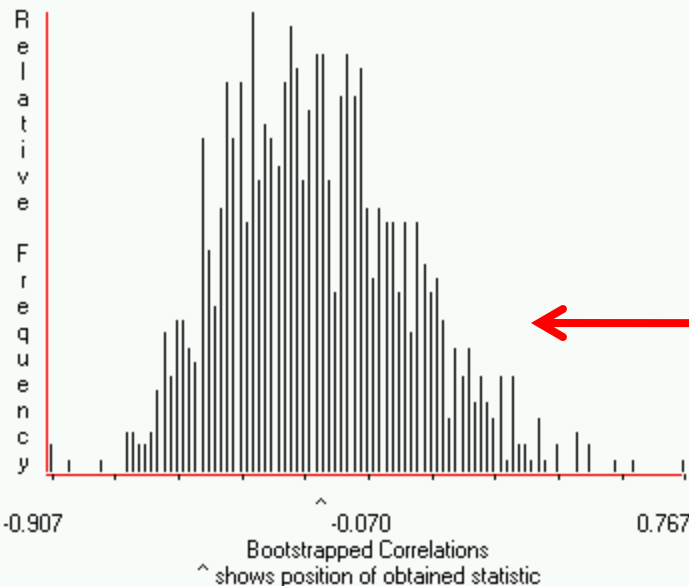
-0.612

Upper Conf. Limit

0.314

St. Error

0.242



Resamples

Sorting

Run

Main Menu

Exit

Análisis de correlación para datos de lluvia anual de la estación Bolivia.

\\uned_estad_2007\\software\\Resampling\\datos_ejemplos

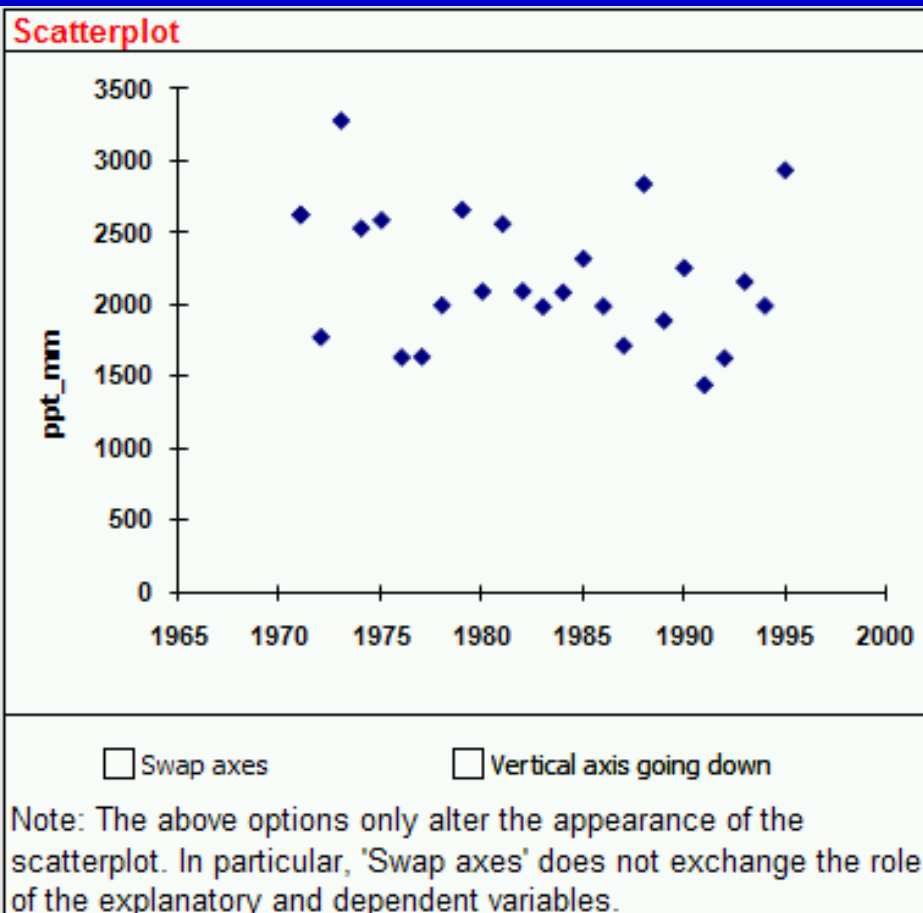
El programa le ofrece una estimación del intervalo de confianza para "r".

Distribución muestral de "r"

Correlación en XLSTats

Two-variable	
Number	25
Covariance	-698.88
Correlation	-0.2053
R ²	0.0421

Single-variable	
Año	
Number	25
Mean	1983
St Dev	7.3598
Min	1971
Max	1995
Median	1983
ppt_mm	
Number	25
Mean	2191.3
St Dev	462.62
Min	1448
Max	3279
Median	2095



Análisis de correlación para datos de lluvia anual de la estación Bolivia.

\\uned_estad_2007\software\Resampling\datos_ejemplos

Correlation Analysis	
Correlation Coeff	
Correlation	-0.20526
Hypothesis Tests for ρ	
$H_0: \rho = 0$	
Alternative	
☒ $\neq$ ☐ $>$ ☐ $<$	
$H_1: \rho \neq 0$	
p-value = 0.324973	

Conclusión estadística: se rechaza H_0 .

Conclusión practica: no existe evidencia para afirmar que ha existido un cambio en el patrón anual de lluvia de la estación.

FLUJOGRAMA

- 1. Seleccionar muestra
- 2. Calcular coeficiente de correlación
- 3. Probar por normalidad de los datos
- 4. Si los datos no cumplen con el supuesto de normalidad debe transformarlos
- 5. Seleccione nivel de confianza
- 6. Realice prueba de hipótesis
- 7. Calcule intervalos de confianza para “r”
- 8. Interprete los resultados. ¿Cuáles son los argumentos estadísticos y biológicos o de otra índole que le permiten explicar los resultados obtenidos? Ej. tamaño de muestra, muestreo sesgado, nivel de confianza utilizado, condiciones ambientales no normales (Ej. El Niño, La Niña)
- ¿Cual sería su recomendación final (acción)? ¿Otro estudio?

Recuerde la estadística es solo una herramienta!!!!