

Ch 12.2 and 12.4 Scatter Plot and Correlation

Ch 12.2 and 12.4 Scatter plot and correlation

Correlation

Ex1. Given the matched pair sample data below. Can we conclude correlation between height and shoe size?

Adult	A	B	C	D	E
height	67"	70"	71"	65"	73"
shoe size	8	12	10.5	7.5	11

Ex2. Given the matched pair sample below, can we conclude correlation between shoe size and math scores?

children	A	B	C	D	E	F
shoe size	8	10	8.5	9	11	8
scores	65	92	71	82	95	80

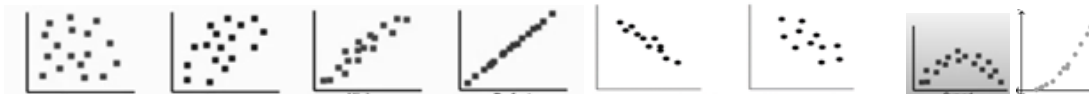
Terms:

Correlation: Correlation between matched pair data (x, y) exists when values of y are associated with the values of x. Note: **correlation does not imply causation.**

Tools to study correlation:

- 1) **Graphical:** scatter plot. Each pair of (x, y) is plotted as one point on a graph. If a systematic pattern exists, there is correlation between x and y. Note: the pattern can be linear or non-linear.
- 2) **Mathematical:** use (x, y) sample data to calculate a correlation coefficient (r) . Value of r is used to determine if linear correlation exists and the strength and type of linear correlation.

Scatter plot examples



no correlation weak positive strong positive perfect positive strong negative weak negative non-linear correlations

Scatter plot

Construct scatter plot: Enter x and y data to statdisk in two different columns. Data/scatter plot/

Select x and y columns. Uncheck show regression line.

copy the scatter plot by labeling the axis and axis title.

Correlation coefficient (r)

The value shows how strongly the matched pair data x, y related to each other linearly.

Use Statdisk, enter data to 2 different columns.

Analysis/Correlation and Regression. Enter significance, select x and y columns, Evaluate.

output under "correlation result"

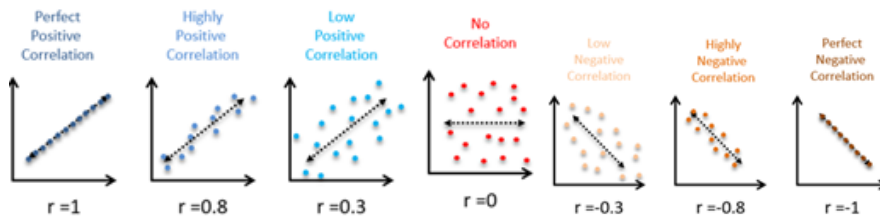
r = is the correlation coefficient, critical r is the critical threshold for evidence of linear correlation.

p-value is the probability of getting the sample under the H0 assumption of no linear correlation.

Properties of r:

- 1) between -1 and 1. r = 0 means no linear correlation. r = 1 means perfect linear correlation.
- 2) If |r| is close to 1, there is strong linear correlation. If |r| is close to 0, there is weak linear correlation.
- 3) r > 0, correlation is positive, x increase, y increase.
r < 0, correlation is negative, x increase, y decrease.

Relationship between scatter plot and correlation coefficient r .



Use Guess correlation game to understand relationship between r and scatter plot. <https://istics.net/Correlations/>

To determine if matched pair (x, y) has linear correlation:

Step 1: Check scatter plot, If non-linear pattern exists, conclude no linear correlation.

Step 2:

Method 1: Use Hypothesis test method with a given α .

ρ = correlation coefficient for population.

r = correlation coefficient for sample.

$H_0: \rho = 0$ (no linear correlation) $H_a: \rho \neq 0$

Use statdisk/Analysis/Correlation and Regression/ to find p-value.

P-value $\leq \alpha$ Reject H_0 , conclude linear correlation

p-value $> \alpha$ Fail to reject H_0 , conclude no linear correlation.

Method2: Compare r and critical value.

Use Analysis/Correlation and Regression to find r and critical r .

If $-\text{critical } r \leq r \leq +\text{critical } r$, conclude no linear correlation.

If $r < -\text{critical } r$ or $r > +\text{critical value of } n \text{ and } \alpha$, conclude linear correlation.

Note: Check scatter plot for non-linear correlation before deciding linear correlation. Do not depend on r only or p-value only.

Ex1. Determine if linear correlation exists between the following pairs of r and p-value given n and α . Assume scatter plots do not show any non-linear patterns.

a) $r = -0.823$, critical $r = \pm 0.754$

Since r is < -0.754 , conclude there is linear correlation.

b) $\alpha = 0.05$, p-value = 0.012

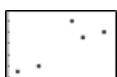
Since only p-value is given, use hypothesis testing method

$0.012 < 0.05$, Reject H_0 , conclude there is linear correlation.

Ex2. Determine if linear correlation exists between height and shoe size in the given matched pair data. Use $\alpha = 0.05$

Adult	A	B	C	D	E
height	67"	70"	71"	65"	73"
shoe size	8	12	10.5	7.5	11

Enter data Statdisk data columns. Statdisk/data/scatter plot/uncheck regression line.



Graph does not show non-linear pattern.

Analysis/Correlation and Regression/

Select height and shoe size columns.

Output: $r = 0.8485$, critical $r = \pm 0.8783$, $p\text{-value} = 0.0692$

Method 1: (p-value method)

$0.0691 > 0.05$, fail to reject H_0 , conclude no linear correlation.

Method 2: (critical value method)

$0.8485 < 0.878$, conclude no linear correlation.

Ex3. Determine if linear correlation exists between shoe size and math scores for 6 children. use $\alpha = 0.05$

children	A	B	C	D	E	F
shoe size	8	10	8.5	9	11	8
scores	65	92	71	82	95	80

Enter to statdisk data columns.



Data/scatter plot/ select data columns, uncheck show regression line.

Scatter plot does not show non-linear pattern.

Analysis/Correlation and Regression/enter significance.

Select data columns. Evaluate

Output: $r = 0.8758$, critical $r = \pm 0.8114$, $p\text{-value} = 0.0222$

Method 1: (p-value method)

$p\text{-value } 0.0222 < 0.05$ Reject H_0 , conclude linear correlation.

Method 2: (critical value method)

$0.876 > 0.811$, conclude linear correlation

Other properties of r :

- 1) r does not change if x and y switches.
- 2) r does not change when different units are used in x and/or y .

Ch 12.2 and 12.4 Scatter Plot and Correlation is shared under a [not declared](#) license and was authored, remixed, and/or curated by LibreTexts.