CORRELATION Pearson Correlation & Spearman Correlation

Dr Asrul Anuar Zulkifli

M.D (UKM), MscMPH (MSU),

Micro credential Digital Health (Monash)

CONTENTS

01 Overview

02 Pearson Correlation

- ☐ Definition & Key Concepts
- □ Assumptions
- ☐ Pearson Correlation Coefficient, r
- □ Calculation Manual & SPSS
- □ Examples

03 Spearman Rank Correlation

- ☐ Definition & Key Concepts
- Applications
- ☐ Spearman Rank Correlation Coefficient
- ☐ Calculation Manual & SPSS



OVERVIEW

Correlation: Correlation describes the relationship between two variables.

Methods of studying correlation:

(depending on data type, assumptions, and objective)

- 1. Scatter Diagram (Visual Method)
- 2. Pearson's Correlation Coefficient (r)
- 3. Spearman's Rank Correlation (ρ or r_s)
- 4. Kendall's Tau
- 5. Concurrent Deviation Method (Less common)

1. Scatter Diagram (Visual Method)

- A graphical representation of data pairs on an XY plot.
- **Visual insight** into the type and strength of relationship.
- Pros: Simple and intuitive.
- Cons: No precise value for correlation.

2. Pearson's Correlation Coefficient (r)

- •Measures **linear correlation** between two continuous variables.
- •Assumes normal distribution, linearity, no outliers.
- •Value between -1 and +1.
- •Formula-based method.

3. Spearman's Rank Correlation (ρ or r_s)

- •Measures monotonic relationship using ranks.
- •Useful when data are **ordinal** or not normally distributed.
- •Less sensitive to outliers.

4. Kendall's Tau

- •Measures association based on the concordance of paired ranks.
- •Often used in small sample sizes or when tied ranks are present.
- •More robust than Spearman in some cases.

5. Concurrent Deviation Method (Less common)

- •Based on the direction (+ or −) of deviations from the mean.
- •Simple, but provides only a rough idea of correlation.

Summary Table

Method	Data Type	Type of Relationship	Notes
Scatter Plot	Any	Visual/Any	Exploratory
Pearson Correlation (r)	Continuous	Linear	Requires normality & linearity
Spearman Correlation (ρ)	Ordinal/Nonparam.	Monotonic	Non-parametric, based on ranks
Kendall's Tau	Ordinal	Monotonic	Robust, good for small samples
Concurrent Deviation	Any	Direction only	Simplified method

Pearson Correlation (r)

OBJECTIVES

01

Explain the concepts and principles of Pearson Correlation

02

Discuss the assumptions

03

Perform Pearson correlation test manually and using SPSS software, interpret and report the results.

Pearson Correlation (r)

- A statistical measure that assesses the strength and direction of a linear relationship between two continuous variables.
- Also called the Pearson Product-Moment Correlation Coefficient.
- Symbol: r

Historical Background

Developed by **Karl Pearson** in the early
20th century.

Used widely in fields like psychology, medicine, economics, and social sciences.

Assumptions of Pearson Correlation

- 1. Variables must be continuous (interval or ratio)
- 2.Linear relationship between variables
- 3. Normal distribution
- 4. Absence of Outliers
- 5. Homoscedasticity (equal variances)

1. Continuous (Interval or Ratio) Variables

- Both variables should be measured at the interval or ratio level (i.e., they are continuous).
- Examples of variables that meet this criterion include revision time (measured in hours), intelligence (measured using IQ score), exam performance (measured from 0 to 100), weight (measured in kg), and so forth.
- Ordinal or categorical variables are not appropriate for Pearson correlation. If the data is ordinal, consider using Spearman's rank correlation.

Examples of Interval Variables

These have equal intervals but **no true zero**:

Variable	Description
Temperature (°C or °F)	0°C does not mean "no temperature"
IQ Score	An IQ of 0 doesn't mean no intelligence
Dates (Calendar Years)	0 AD is arbitrary, not a true zero
Standardized Test Scores	e.g., SAT, GRE

Examples of Ratio Variables

These have a **true zero** and allow full range of mathematical operations (including ratios):

Variable	Description
Height (cm, m)	0 cm = absence of height
Weight (kg, lbs)	0 kg = no weight
Age (in years)	0 years = newborn
Income (\$)	0 = no income
Distance (km, miles)	0 = no distance
Time (seconds)	0 = no time elapsed

2. Linearity

Definition: The relationship between the two variables should be linear — i.e., changes in one variable are associated with proportional changes in the other.

Why it matters: Pearson correlation only captures linear relationships. A nonlinear pattern may result in a low r-value, even when a strong relationship exists.

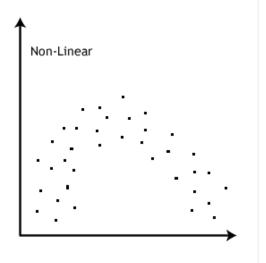
How to check: Use a scatterplot to visually inspect if the data follows a straight-line trend.

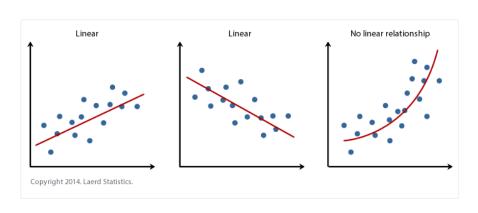
Quick Steps

- Click Graphs -> Legacy Dialogs -> Scatter/Dot (in older versions of SPSS) OR
 Click Graphs > Scatter/Dot (in newer versions of SPSS)
- 2. Select **Simple Scatter**
- 3. Click **Define**
- 4. Click **Reset** (recommended)
- 5. Select the predictor/independent variable and move it into the **X Axis** box
- 6. Select the criterion/dependent variable and move it into the **Y Axis** box
- 7. Select **Titles** to add a title (recommended), then click **Continue**
- 8. Select **OK**
- 9. Your scatter plot will appear in SPSS Output Viewer

Scatterplot using SPSS Statistics

Your scatterplot may look something like one of the following:

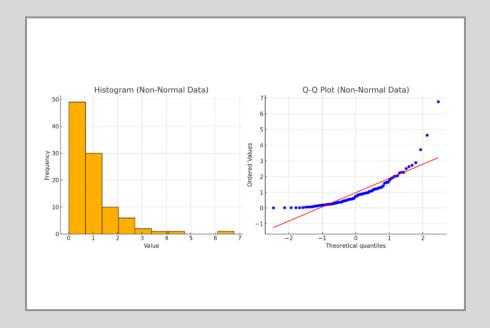


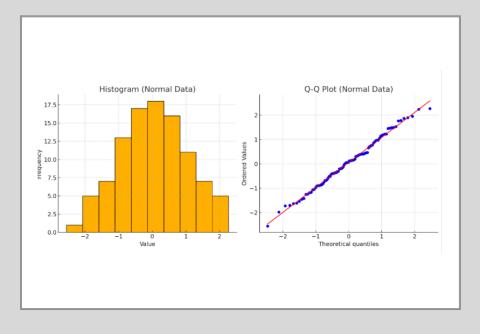


3. Normal Distribution (Bivariate Normality)

- Each variable should be approximately normally distributed.
- The joint distribution of both variables should be bivariate normal.
 - But this assumption is difficult to assess, so a simpler method is more commonly used.
 - This simpler method involves determining the normality of each variable separately.
- Especially important when testing significance (p-values).
- How to check:
 - Histogram or Q-Q plot for each variable
 - Use normality tests like Shapiro-Wilk or Kolmogorov-Smirnov.
 - ✓ If the p > 0.05, the data is normal.

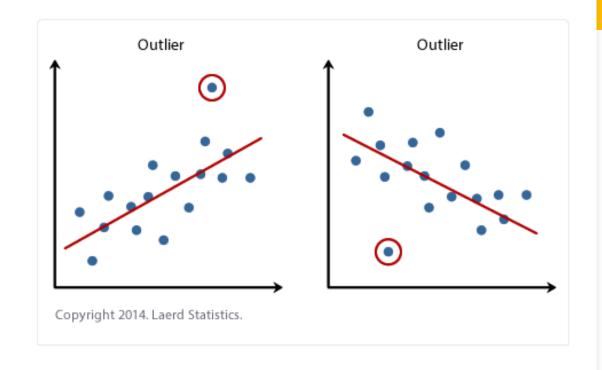
- 1. Open SPSS: Start the SPSS software.
- 2. **Navigate to Explore:** Go to the SPSS menu: Analyze > Descriptive Statistics > Explore.
- 3. **Select Variables:** In the "Explore" window, transfer the variable(s) you wish to test for normality to the "Dependent List".
- 4. Customize Plots: Click on the "Plots..." button.
- 5. Check Normality Plots: In the "Plots" window, select the "Normality plots with tests" option.
- 6. **Run the Test:** Click "Continue" to return to the main "Explore" window and then click "OK" to run the normality test.
- 7. Interpret Results: SPSS will generate output, including descriptive statistics, histograms, and the results of the Kolmogorov-Smirnov and Shapiro-Wilk tests of normality. The output will show the p-values for these tests. If the p-value is greater than 0.05 for both tests, the data is considered approximately normally distributed.





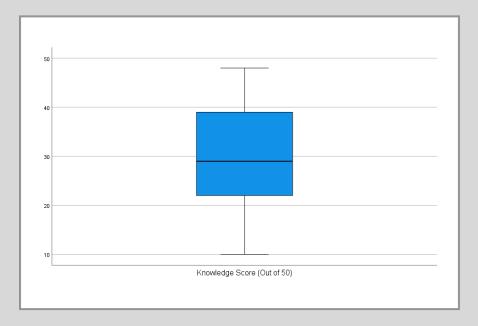
4. Absence of Outliers

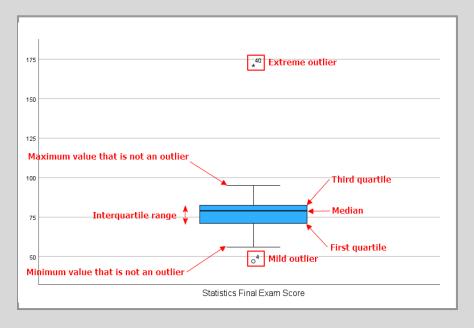
- Pearson correlation is very sensitive to outliers.
- Even a single extreme value can dramatically skew the correlation coefficient.
- How to check:
 - ✓ Boxplots
 - ✓ Explore -> Using outlier identification
 - ✓Z-scores (values > ±3 considered outliers)



1. Using Box Plots:

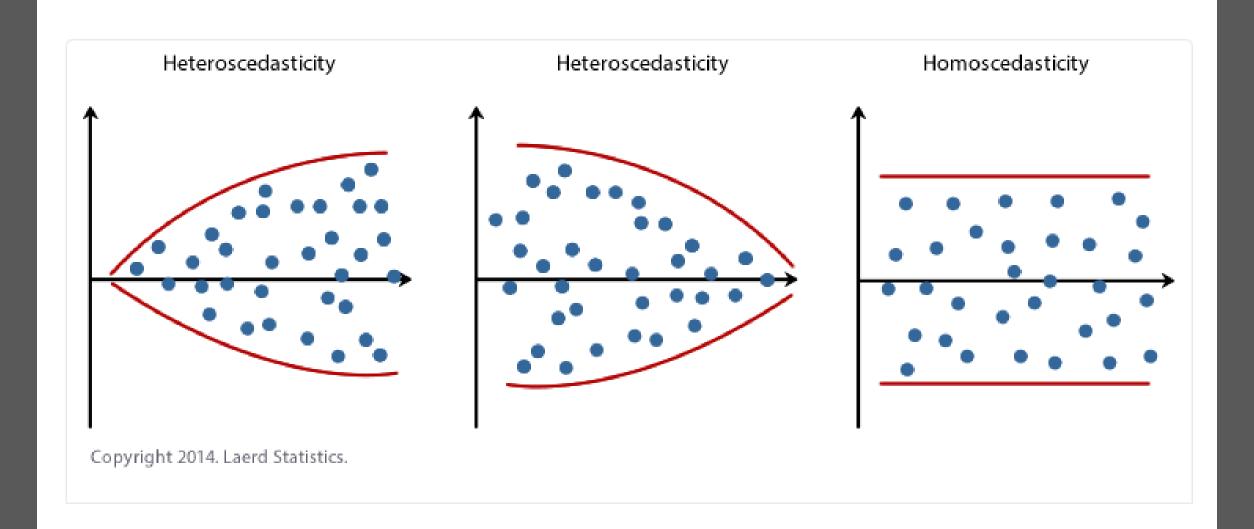
- Open SPSS: Launch the SPSS software and open your dataset.
- Navigate to Graphs: Go to "Graphs" > "Legacy Dialogs" > "Boxplot".
- **Select Variable:** Choose the variable you want to analyze for outliers and move it to the "Dependent" box.
- **Define Categories (Optional):** If you have grouping variables, move them to the "Categorical" box to create separate box plots for each group.
- Generate the Plot: Click "OK" to create the box plot.
- Examine Outliers: Look for points that fall outside the "whiskers" of the box plot. These points are potential outliers.

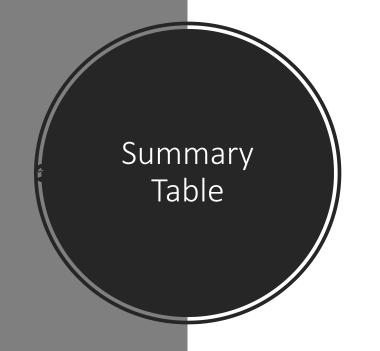




5. Homoscedasticity (Equal Variances)

- Homoscedasticity refers to the assumption that the variance of the errors (or residuals) is constant across all levels of the independent variable.
- Also called homogeneity of variances.
- Homoscedasticity is not a core assumption for Pearson correlation itself.
- Violations lead to heteroscedasticity, which can distort the strength of the correlation.
- How to check:
 - ✓ Plot residuals vs. fitted values
 - ✓ Look for a "fan shape" pattern indicates heteroscedasticity





Assumption	Description	How to Check
1. Continuous Variables	Both variables must be interval or ratio scale	Know your data scale
2. Linearity	Straight-line relationship Scatterplot	
3. Normality	Each variable is normally distributed	Histogram, Q-Q Plot, Normality Tests
4. No Outliers	No extreme values affecting results Boxplot, Z-scores	
5. Homoscedasticity	Equal spread of data along regression line	Residual plots

Pearson Correlation Coefficient, r

The calculations are based on the differences between the observations x_i , y_i and their means \overline{x} and \overline{y} as shown in the following formula:

Formula

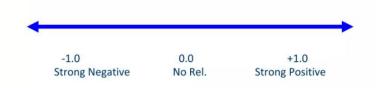
$$r = \frac{\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \overline{x})^2 \sum_{i=1}^{n} (y_i - \overline{y})^2}}$$

where:

- X_i, Y_i are individual data points
- $ar{X}, ar{Y}$ are the mean values of X and Y

where x_i , y_i are values of the n pairs of two variables.

Interpretation of r

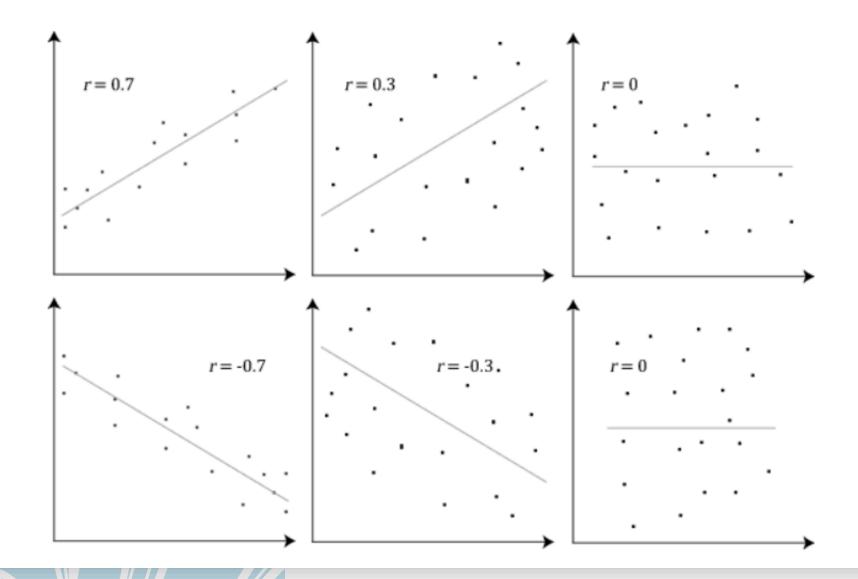


- R tells us how close is the linear relationship between the 2 variables
- R lies between -1 and +1
- Negative values indicate a negative linear relationship
 - ✓ 1 variable increases, the other decreases
- Positive values indicate a positive linear relationship
 - ✓ 1 variable increases, so does the other
- R = 0 indicates no linear relationship
 - ✓ The values of each variable are independent of each other

Interpreting Pearson Correlation Coefficient

Range of Correlation Coefficient Values	Level of Correlation	Range of Correlation Coefficient Values	Level of Correlation
0.80 to 1.00	Very Strong Positive	-1.00 to -0.80	Very Strong Negative
0.60 to 0.79	Strong Positive	-0.79 to -0.60	Strong Negative
0.40 to 0.59	Moderate Positive	-0.59 to -0.40	Moderate Negative
0.20 to 0.39	Weak Positive	-0.39 to -0.20	Weak Negative
0.00 to 0.19	Very Weak Positive	-0.19 to -0.01	Very Weak Negative

Different relationships and their correlation coefficients are shown in the diagram below:



Calculate a Pearson Correlation Coefficient (r) for the following dataset:

Υ
45
47
39
58
68
76
75
74
78
81

Performing Pearson Correlation (Manual Calculation)

Step 1: Calculate the Mean of X and Y

Mean

X	Y
6	45
12	47
13	39
17	58
22	68
25	76
27	75
29	74
30	78
32	81

Step 2: Calculate the Difference between Means

	x	у	x - x _{mean} y - y _{mea}
	6	45	,
	12	47	
	13	39	
	17	58	
	22	68	
	25	76	
	27	75	
	29	74	
	30	78	
	32	81	
Mean	21.3	64.1	

Step 3: Calculate the Remaining Values

	x	у	x - x _{mean}	y - y _{mean}	(x - x _{mean})*(y - y _{mean})	(x - x _{mean}) ²	(y - y _{mean}) ²
	6	45	-15.3	-19.1			
	12	47	-9.3	-17.1			
	13	39	-8.3	-25.1			
	17	58	-4.3	-6.1			
	22	68	0.7	3.9			
	25	76	3.7	11.9			
	27	75	5.7	10.9			
	29	74	7.7	9.9			
	30	78	8.7	13.9			
	32	81	10.7	16.9			
an	21.3	64.1					

Step 4: Calculate the Sums

×	у	x - x _{mean}	y - y _{mean}	(x - x _{mean})*(y - y _{mean})	(x - x _{mean}) ²	(y - y _{mean})
6	45	-15.3	-19.1	292.23	234.09	364.81
12	47	-9.3	-17.1	159.03	86.49	292.41
13	39	-8.3	-25.1	208.33	68.89	630.01
17	58	-4.3	-6.1	26.23	18.49	37.21
22	68	0.7	3.9	2.73	0.49	15.21
25	76	3.7	11.9	44.03	13.69	141.61
27	75	5.7	10.9	62.13	32.49	118.81
29	74	7.7	9.9	76.23	59.29	98.01
30	78	8.7	13.9	120.93	75.69	193.21
32	81	10.7	16.9	180.83	114.49	285.61
21.3	64.1		Sum			

Mean

Step 5: Calculate the Pearson Correlation Coefficient

Formula
$r = \frac{\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})}{(x_i - \overline{x})(y_i - \overline{y})}$
$\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 \sum_{i=1}^{n} (y_i - \bar{y})^2}$

	x	у	x - x _{mean}	y - y _{mean}	(x - x _{mean})*(y - y _{mean})	(x - x _{mean}) ²	(y - y _{mean}) ²
	6	45	-15.3	-19.1	292.23	234.09	364.81
	12	47	-9.3	-17.1	159.03	86.49	292.41
	13	39	-8.3	-25.1	208.33	68.89	630.01
ĺ	17	58	-4.3	-6.1	26.23	18.49	37.21
	22	68	0.7	3.9	2.73	0.49	15.21
	25	76	3.7	11.9	44.03	13.69	141.61
	27	75	5.7	10.9	62.13	32.49	118.81
ľ	29	74	7.7	9.9	76.23	59.29	98.01
	30	78	8.7	13.9	120.93	75.69	193.21
	32	81	10.7	16.9	180.83	114.49	285.61
1	21.3	64.1		Sum	1172.7	704.1	2176.9

$$r = 1172.7 / \sqrt{(704.1)*(2176.9)} = 0.947$$

The Pearson Correlation Coefficient, r = 0.947.

Mean

Since this value is close to 1, this is an indication that X and Y are strongly positively correlated.

Step 6: Determine p-value Using Pearson Correlation Table

Step 1: Calculate r manually

$$\sqrt{r}=0.947$$

Step 2: Determine your sample size n

Step 3: Calculate the degrees of freedom

- df = n-2
- df = 10 2 = 8

Step 4: Choose your significance level (α)

Common levels:

- $\alpha = 0.05 \text{ alpha} = 0.05 \alpha = 0.05 \text{ (95\% confidence)}$
- α =0.01\alpha = 0.01 α =0.01 (99% confidence)

Step 5: Refer to a Pearson critical value table

• Look up the **critical r values** for your **df = n - 2** at the desired significance level.

2 0.800 3 0.687 4 0.608 5 0.550 6 0.506 7 0.471 8 0.442 9 0.418 10 0.398 11 0.380 12 0.364 13 0.350 14 0.338 15 0.327 16 0.316	7049 0.80538-8400 0.729299 0863 0.66943-6727 0.62148-1589 0.582200 02796 0.54935-8662 0.52140-8062 0.49726-1564562 0.45750	0.996917 0.950000 0.878339 0.811401 0.754492 0.706734 0.666384 0.631897 0.602069 0.575983 0.552943	0.934333 0.882194 0.832874 0.788720 0.749776 0.715459 0.685095 0.658070	0.01 0.999877 0.990000 0.958735 0.917200 0.874526 0.834342 0.797681 0.764592 0.734786 0.707888	0.999000 0.991139 0.974068 0.950883 0.924904 0.898260 0.872115 0.847047	df\ 35 40 45 50 60 70 80 90		0.242859 0.230620 0.210832 0.195394 0.182916		0.02 0.380976 0.357787 0.338367 0.321796 0.294846 0.273695 0.256525	0.418211 0.393174 0.372142	0.518898 0.489570
2 0.800 3 0.687 4 0.608 5 0.550 6 0.506 7 0.471 8 0.442 9 0.418 10 0.398 11 0.380 12 0.364 13 0.350 14 0.338 15 0.327 16 0.316	0000 0.90000 7049 0.80538 8400 0.72929 0863 0.66943 6727 0.62148 1589 0.58220 2796 0.54935 8662 0.49726 0216 0.47615 4562 0.45750	0.950000 0.878339 0.811401 0.754492 0.706734 0.666384 0.631897 0.602069 0.575983 0.552943	0.980000 0.934333 0.882194 0.832874 0.788720 0.749776 0.715459 0.685095 0.658070	0.990000 0.958735 0.917200 0.874526 0.834342 0.797681 0.764592 0.734786	0.999000 0.991139 0.974068 0.950883 0.924904 0.898260 0.872115 0.847047	40 45 50 60 70 80 90	0.201796 0.190345 0.180644 0.164997 0.152818 0.142990	0.257278 0.242859 0.230620 0.210832 0.195394 0.182916	0.304396 0.287563 0.273243 0.250035 0.231883 0.217185	0.357787 0.338367 0.321796 0.294846 0.273695	0.393174 0.372142 0.354153 0.324818 0.301734	0.489570 0.464673 0.443201 0.407865 0.379799
3 0.687 4 0.608 5 0.550 6 0.506 7 0.471 8 0.442 9 0.418 10 0.398 11 0.380 12 0.364 13 0.350 14 0.338 15 0.327 16 0.316	7049 0.80538-8400 0.729299 0863 0.66943-6727 0.62148-1589 0.582200 02796 0.54935-8662 0.52140-8062 0.49726-1564562 0.45750	0.878339 0.811401 0.754492 0.706734 0.666384 0.631897 0.602069 0.575983 0.552943	0.934333 0.882194 0.832874 0.788720 0.749776 0.715459 0.685095 0.658070	0.958735 0.917200 0.874526 0.834342 0.797681 0.764592 0.734786	0.991139 0.974068 0.950883 0.924904 0.898260 0.872115 0.847047	45 50 60 70 80 90	0.190345 0.180644 0.164997 0.152818 0.142990	0.242859 0.230620 0.210832 0.195394 0.182916	0.287563 0.273243 0.250035 0.231883 0.217185	0.338367 0.321796 0.294846 0.273695	0.372142 0.354153 0.324818 0.301734	0.464673 0.443201 0.407865 0.379799
4 0.608 5 0.550 6 0.506 7 0.471 8 0.442 9 0.418 10 0.398 11 0.380 12 0.364 13 0.350 14 0.338 15 0.327 16 0.316	8400 0.729299 0863 0.669439 6727 0.621489 1589 0.582200 2796 0.549359 8662 0.497269 0216 0.476150 4562 0.45750	0.811401 0.754492 0.706734 0.666384 0.631897 0.602069 0.575983 0.552943	0.882194 0.832874 0.788720 0.749776 0.715459 0.685095 0.658070	0.917200 0.874526 0.834342 0.797681 0.764592 0.734786	0.974068 0.950883 0.924904 0.898260 0.872115 0.847047	50 60 70 80 90	0.180644 0.164997 0.152818 0.142990	0.230620 0.210832 0.195394 0.182916	0.273243 0.250035 0.231883 0.217185	0.321796 0.294846 0.273695	0.354153 0.324818 0.301734	0.443201 0.407865 0.379799
5 0.550 6 0.506 7 0.471 8 0.442 9 0.418 10 0.398 11 0.380 12 0.364 13 0.350 14 0.338 15 0.327 16 0.316	0.66943: 0.62148: 1589	0.754492 0.706734 0.666384 0.631897 0.602069 0.575983 0.552943	0.832874 0.788720 0.749776 0.715459 0.685095 0.658070	0.874526 0.834342 0.797681 0.764592 0.734786	0.950883 0.924904 0.898260 0.872115 0.847047	60 70 80 90	0.164997 0.152818 0.142990	0.210832 0.195394 0.182916	0.250035 0.231883 0.217185	0.294846 0.273695	0.324818 0.301734	0.407865 0.379799
6 0.506 7 0.471 8 0.442 9 0.418 10 0.398 11 0.380 12 0.364 13 0.350 14 0.338 15 0.327 16 0.316	6727 0.62148 1589 0.58220 2796 0.54935 8662 0.52140 8062 0.49726 0216 0.47615 4562 0.45750	0.706734 0.666384 0.631897 0.602069 0.575983 0.552943	0.788720 0.749776 0.715459 0.685095 0.658070	0.834342 0.797681 0.764592 0.734786	0.924904 0.898260 0.872115 0.847047	70 80 90	0.152818 0.142990	0.195394 0.182916	0.231883 0.217185	0.273695	0.301734	0.379799
7 0.471 8 0.442 9 0.418 10 0.398 11 0.380 12 0.364 13 0.350 14 0.338 15 0.327 16 0.316	1589 0.58220 2796 0.54935 8662 0.52140 8062 0.49726 0216 0.47615 4562 0.45750	0.666384 0.631897 0.602069 0.575983 0.552943	0.749776 0.715459 0.685095 0.658070	0.797681 0.764592 0.734786	0.898260 0.872115 0.847047	80 90	0.142990	0.182916	0.217185			
8 0.442 9 0.418 10 0.398 11 0.380 12 0.364 13 0.350 14 0.338 15 0.327 16 0.316	2796 0.54935 8662 0.52140 8062 0.49726 0216 0.47615 4562 0.45750	0.631897 0.602069 0.575983 0.552943	0.715459 0.685095 0.658070	0.764592 0.734786	0.872115 0.847047	90				0.256525	0.282958	0.356816
9 0.418 10 0.398 11 0.380 12 0.364 13 0.350 14 0.338 15 0.327 16 0.316	8662 0.52140 8062 0.49726 0216 0.47615 4562 0.45750	0.602069 0.575983 0.552943	0.685095 0.658070	0.734786	0.847047		0.134844	0.172558	0.204968			
10 0.398 11 0.380 12 0.364 13 0.350 14 0.338 15 0.327 16 0.316	0.49726 0216 0.47615 4562 0.45750	0.575983 0.552943	0.658070						0.201500	0.242227	0.267298	0.337549
11 0.380 12 0.364 13 0.350 14 0.338 15 0.327 16 0.316	0216 0.47615 4562 0.45750	0.552943		0.707888	0.000000	100	0.127947	0.163782	0.194604	0.230079	0.253979	0.321095
12 0.364 13 0.350 14 0.338 15 0.327 16 0.316	4562 0.45750		0.633863		0.823305	125	0.114477	0.146617	0.174308	0.206245	0.227807	0.288602
13 0.350 14 0.338 15 0.327 16 0.316		0.532413		0.683528	0.800962	150	0.104525	0.133919	0.159273	0.188552	0.208349	0.264316
14 0.338 15 0.327 16 0.316	0688 0.44086		0.612047	0.661376	0.779998	175	0.096787	0.124036	0.147558	0.174749	0.193153	0.245280
15 0.327 16 0.316		0.513977	0.592270	0.641145	0.760351	200	0.090546	0.116060	0.138098	0.163592	0.180860	0.229840
16 0.316	8282 0.42590	0.497309	0.574245	0.622591	0.741934	250	0.081000	0.103852	0.123607	0.146483	0.161994	0.206079
	7101 0.41236	0.482146	0.557737	0.605506	0.724657	300	0.073951	0.094831	0.112891	0.133819	0.148019	0.188431
17 0 307	6958 0.40002	0.468277	0.542548	0.589714	0.708429	350	0.068470	0.087814	0.104552	0.123957	0.137131	0.174657
17 0.507	7702 0.38873	0.455531	0.528517	0.575067	0.693163	400	0.064052	0.082155	0.097824	0.115997	0.128339	0.163520
18 0.299	9210 0.37834	0.443763	0.515505	0.561435	0.678781	450	0.060391	0.077466	0.092248	0.109397	0.121046	0.154273
19 0.291	1384 0.36873	0.432858	0.503397	0.548711	0.665208	500	0.057294	0.073497	0.087528	0.103808	0.114870	0.146436
20 0.284	4140 0.35982	0.422714	0.492094	0.536800	0.652378	600	0.052305	0.067103	0.079920	0.094798	0.104911	0.133787
21 0.277	7411 0.35153	0.413247	0.481512	0.525620	0.640230	700	0.048427	0.062132	0.074004	0.087789	0.097161	0.123935
22 0.271	1137 0.34378	0.404386	0.471579	0.515101	0.628710	800	0.045301	0.058123	0.069234	0.082135	0.090909	0.115981
23 0.265	5270 0.33652	0.396070	0.462231	0.505182	0.617768	900	0.042711	0.054802	0.065281	0.077450	0.085727	0.109385
24 0.259	9768 0.32970	0.388244	0.453413	0.495808	0.607360	1000	0.040520	0.051993	0.061935	0.073484	0.081340	0.103800
25 0.254	4594 0.32328	0.380863	0.445078	0.486932	0.597446	1500	0.033086	0.042458	0.050582	0.060022	0.066445	0.084822
26 0.249	9717 0.31722	0.373886	0.437184	0.478511	0.587988	2000	0.028654	0.036772	0.043811	0.051990	0.057557	0.073488
27 0.245	5110 0.31149	0.367278	0.429693	0.470509	0.578956	3000	0.023397	0.030027	0.035775	0.042457	0.047006	0.060027
28 0.240	0749 0.30605	0.361007	0.422572	0.462892	0.570317	4000	0.020262	0.026005	0.030984	0.036773	0.040713	0.051996
29 0.236	6612 0.30089	0.355046	0.415792	0.455631	0.562047	5000	0.018123	0.023260	0.027714	0.032892	0.036417	0.046512
30 0.232	0.50005	0.349370	0.409327	0.448699	0.554119	•						

Step 6: Compare your r to the critical values

- If r > r critical, the result is statistically significant.
- If r < r critical, the result is **not statistically significant**.
- Your r=0.947 is greater than all critical values.

 - \bigcirc 0.947 > 0.872 \rightarrow Significant at α = 0.001

Conclusion

Your correlation is very strong and highly statistically significant — p < 0.001. This means there is very strong evidence that the relationship between your two variables is not due to chance.

Pearson Correlation Coefficient using SPSS

SPSS Statistics

Example

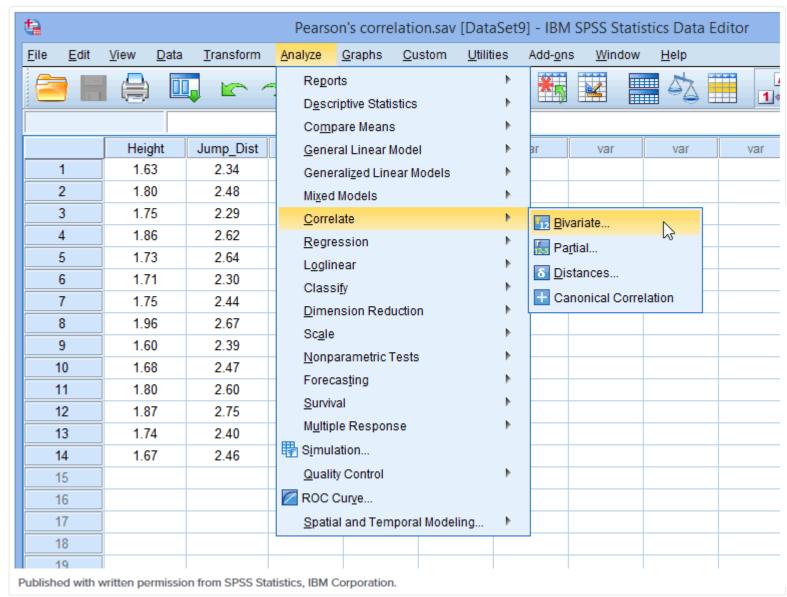
A researcher wants to know whether a person's height is related to how well they perform in a long jump. The researcher recruited untrained individuals from the general population, measured their height and had them perform a long jump. The researcher then investigated whether there was an association between height and long jump performance by running a Pearson's correlation.

Dataset:

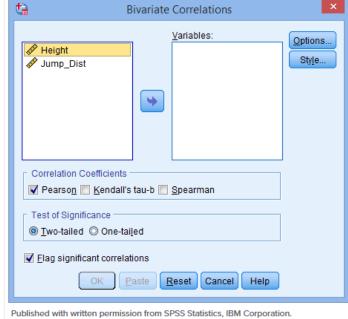
In SPSS Statistics, we created two variables so that we could enter our data: Height (i.e., participants' height) and Jump_Dist (i.e., distance jumped in a long jump).

Dataset: Height_Jump_Pearson.xlsx

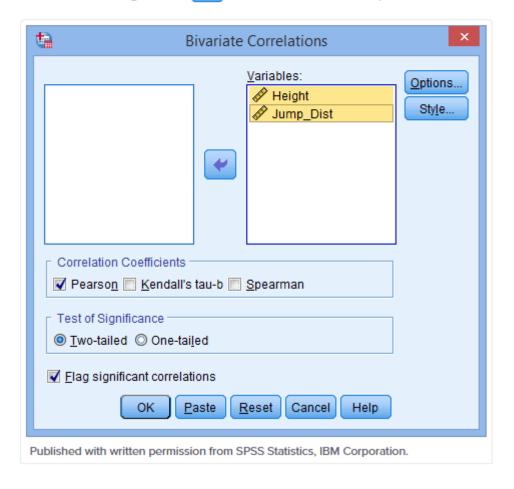
Click Analyze > Correlate > Bivariate... on the main menu, as shown below:



You will be presented with the **Bivariate Correlations** dialogue box:

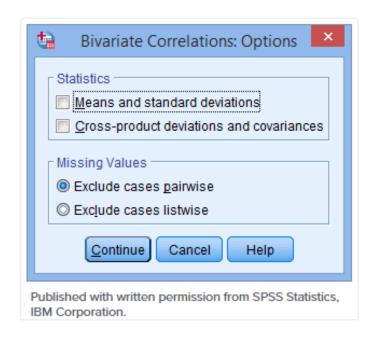


Transfer the variables Height and Jump_Dist into the Variables: box by dragging-and-dropping them or by clicking on them and then clicking on the button. You will end up with a screen similar to the one below:



Make sure that the
☐ Pearson checkbox is selected under the —Correlation Coefficients— area (although it is selected by default in SPSS Statistics).

Click on the Options... button and you will be presented with the **Bivariate Correlations: Options** dialogue box. If you wish to generate some descriptives, you can do it here by clicking on the relevant checkbox in the -Statistics- area.



Click on the Continue button. You will be returned to the Bivariate Correlations dialogue box.

6 Click on the OK button. This will generate the results of Pearson's correlation.

Output for Pearson's correlation

Correlations

		Height	Jump_Dist
Height	Pearson Correlation	1	.706**
	Sig. (2-tailed)		.005
	N	14	14
Jump_Dist	Pearson Correlation	.706**	1
	Sig. (2-tailed)	.005	
	Ν	14	14

^{**.} Correlation is significant at the 0.01 level (2-tailed).

Published with written permission from SPSS Statistics, IBM Corporation.

In this example, we can see that the Pearson correlation coefficient, r, is 0.706, and that it is statistically significant (p = 0.005).

Reporting the Output

A Pearson product-moment correlation was run to determine the relationship between height and distance jumped in a long jump. There was a strong, positive correlation between height and distance jumped, which was statistically significant (r = .706, n = 14, p = .005).

Example 1

Topic:

Relationship between Physical Activity and Body Mass Index (BMI) among Adults: A Pearson Correlation Analysis

Background:

Public health studies often examine how lifestyle behaviors, like physical activity, relate to health outcomes such as obesity (measured by BMI).

Physical activity is known to reduce the risk of overweight and obesity. In this study, we randomly surveyed **40 adults** to record:

- Their **physical activity** (total minutes of exercise per week).
- Their BMI (Body Mass Index, kg/m²).

Dataset: PA_BMI.xlsx

Hypothesis:

- Null hypothesis (H_0): There is no correlation between physical activity and BMI.
- Alternative hypothesis (H₁): There is a significant correlation between physical activity and BMI.



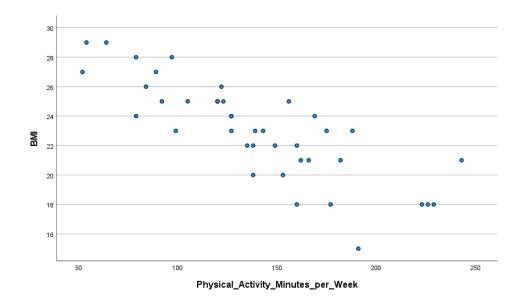
Assumptions

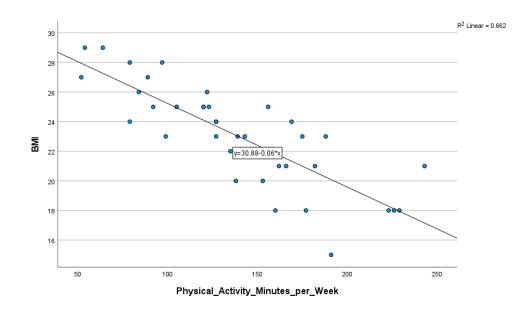
- 1. Variables?
- 2. Linearity?
- 3. Normality?
- 4. Outliers?
- 5. Homoscedasticity (optional)

Dataset

Variable	Туре	Role
Physical Activity (minutes per week)	Continuous	Independent variable
Body Mass Index (BMI)	Continuous	Dependent variable

Physical_Activity_Minutes_per_Week	вмі
175	23
143	23
182	21
226	18
138	20
138	22
229	18
188	23
127	24
177	18
127	24
127	23
162	21
54	29
64	29
122	26
99	23
166	21
105	25
79	28
223	18
139	23
153	20
79	24
123	25
156	25
92	25
169	24
120	25
135	22
120	25
243	21
149	22
97	28
191	15
89	27
160	22
52	27
84	26
160	18

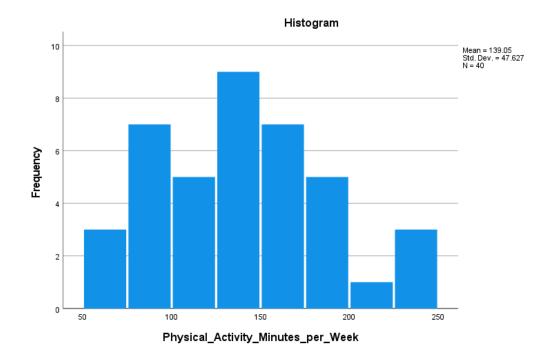


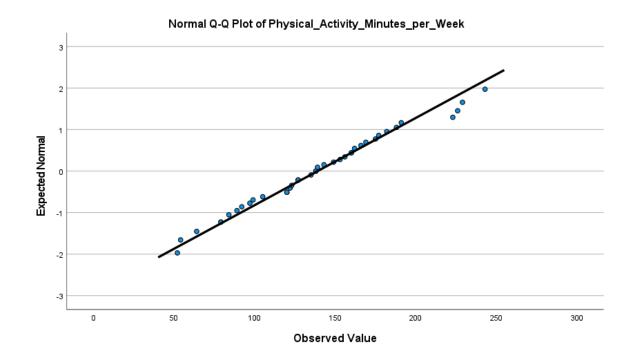


Tests of Normality

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Physical_Activity_Minutes _per_Week	.070	40	.200*	.979	40	.666

- *. This is a lower bound of the true significance.
- a. Lilliefors Significance Correction

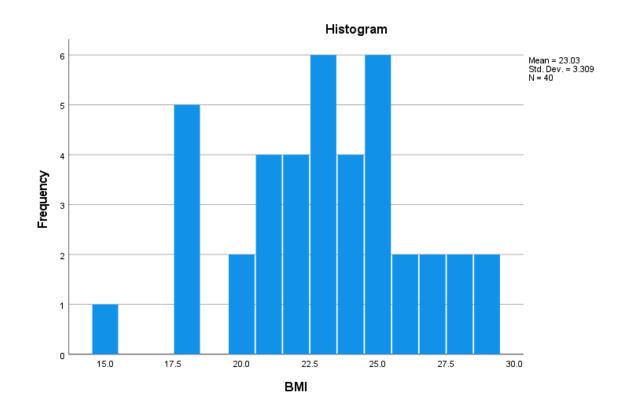


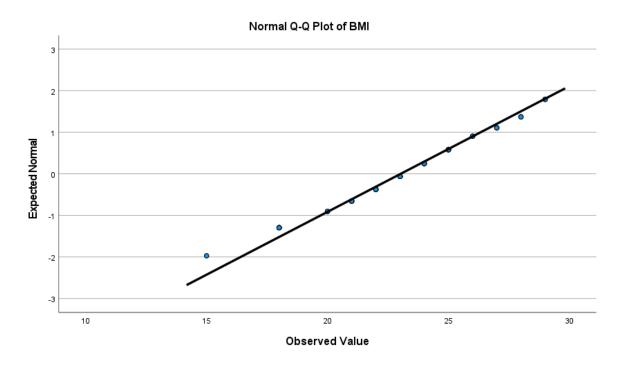


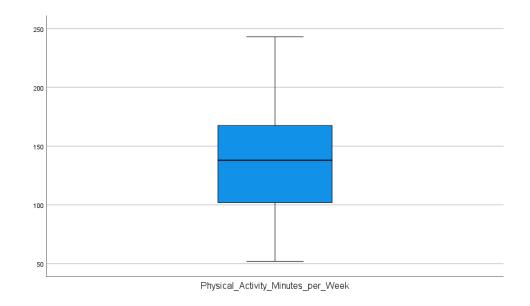
Tests of Normality

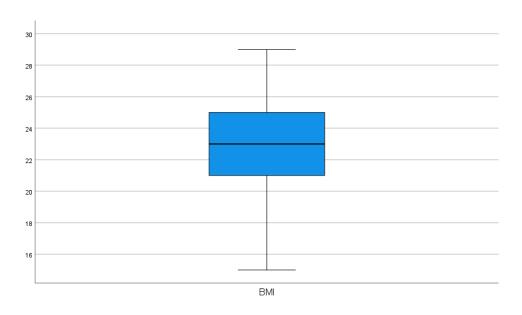
Kolmogorov-Smirnov ^a				Shapiro-Wilk			
	Statistic df Sig.		Sig.	Statistic	Sig.		
ВМІ	.097	40	.200*	.973	40	.441	

- *. This is a lower bound of the true significance.
- a. Lilliefors Significance Correction



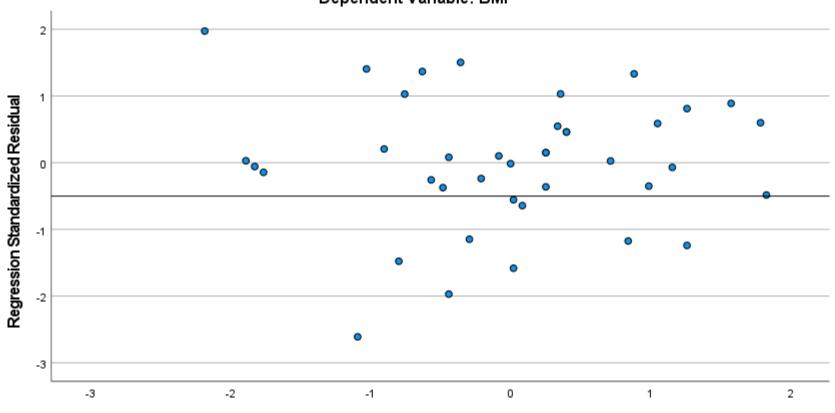






Scatterplot

Dependent Variable: BMI



Regression Standardized Predicted Value

The Pearson correlation test was conducted to assess the strength and direction of the relationship between physical activity and BMI.

Correlations		
	Physical_Activ ity_Minutes_p er_Week	ВМІ
Pearson Correlation	1	813**
Sig. (2-tailed)		<.001
N	40	40
Pearson Correlation	813**	1
Sig. (2-tailed)	<.001	
N	40	40
	Pearson Correlation Sig. (2-tailed) N Pearson Correlation Sig. (2-tailed)	Physical_Activ ity_Minutes_p er_Week Pearson Correlation 1 Sig. (2-tailed) N 40 Pearson Correlation813** Sig. (2-tailed) <.001

Interpretation:

- The Pearson correlation coefficient (r = -0.813)
 indicates a strong negative correlation between
 Physical Activity Minutes per Week and BMI.
- The p-value (< 0.001) shows that the correlation is statistically significant at the 0.01 level.

Meaning:

- → As physical activity increases, BMI tends to **decrease**.
- → Adults who are more physically active tend to have a lower BMI.

Conclusion:

There is a **strong, statistically significant negative relationship** between physical activity levels and BMI among the 40 adults surveyed.

This supports public health strategies promoting regular physical activity to help manage or lower BMI.

Example 2

Title:

Relationship between Age and Systolic Blood Pressure among Hypertensive Patients: A Pearson Correlation Analysis.

Background:

Hypertension (high blood pressure) is a major risk factor for cardiovascular diseases.

Blood pressure tends to increase with age due to changes in blood vessel elasticity and other physiological factors. This study aims to assess the relationship between **age** and **systolic blood pressure (SBP)** among patients diagnosed with hypertension.

Dataset: AGE_BP.xlxs

Hypotheses:

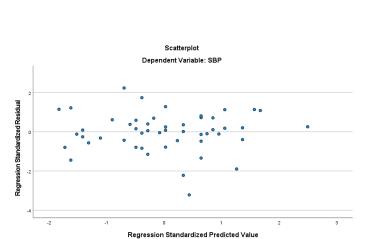
- •Null Hypothesis (H_0): There is no significant relationship between age and systolic blood pressure among hypertensive patients.
- •Alternative Hypothesis (H_1): There is a significant relationship between age and systolic blood pressure among hypertensive patients.



Assumptions

- 1. Variables?
- 2. Linearity?
- 3. Normality?
- 4. Outliers?
- 5. Homoscedasticity (optional)

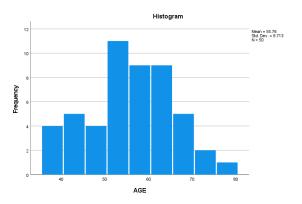
R² Linear = 0.194 AGE

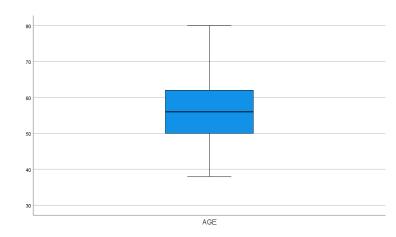


Tests of Normality

Kolmogorov-Smirnov ^a				Shapiro-Wilk		
	Statistic	tic df Sig. Statistic df		Sig.		
AGE	.080	50	.200*	.981	50	.575

- *. This is a lower bound of the true significance.
- a. Lilliefors Significance Correction

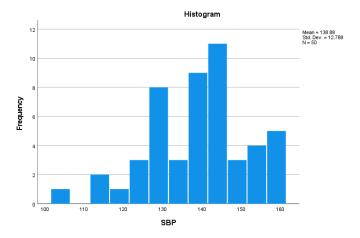


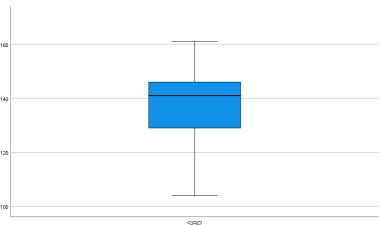


Tests of Normality

Kolmogorov-Smirnov ^a				Shapiro-Wilk			
Statistic		df	df Sig. Statis		df	Sig.	
SBP	.095	50	.200*	.975	50	.362	

- *. This is a lower bound of the true significance.
- a. Lilliefors Significance Correction





Correlations								
	AGE SBP							
AGE	Pearson Correlation	1	.441**					
	Sig. (2-tailed)		.001					
	N	50	50					
SBP	Pearson Correlation	.441**	1					
	Sig. (2-tailed)	.001						
	N	50	50					
**. Correlation is significant at the 0.01 level (2-tailed).								

△ Interpretation:

- The Pearson r = 0.441 indicates a moderate positive correlation between Age and Systolic Blood Pressure.
- The **p-value = 0.001** is **statistically significant** at the 0.01 level.

Meaning:

→ As age increases, systolic blood pressure tends to increase as well among hypertensive patients.



Conclusion (Public Health Context):

This study found a moderate, statistically significant positive relationship between age and systolic blood pressure among hypertensive patients.

From a public health perspective, these findings highlight the importance of **early screening**, **regular monitoring**, **and proactive blood pressure management**, especially in older adults.

References on Sample Size for Pearson Correlation

1. Bonett, D. G., & Wright, T. A. (2000).

Title: Sample size requirements for estimating Pearson, Kendall, and Spearman correlations with fixed precision.

Journal: *Psychological Methods, 5(2), 241–250.*

This study discusses how small samples ($n \approx 20-30$) can sometimes be acceptable for strong correlations (r > 0.5), but to accurately estimate correlations (especially moderate ones), larger samples (r > 50) are recommended.

Source: Bonett, D. G., & Wright, T. A. (2000). Sample size requirements for estimating Pearson, Kendall and Spearman correlations with fixed precision. *Psychological Methods*, 5(2), 241–250. https://doi.org/10.1037/1082-989X.5.2.241

2. Hulley, S. B., Cummings, S. R., Browner, W. S., Grady, D., & Newman, T. B. (2013).

Title: Designing Clinical Research (4th ed.).

Publisher: Lippincott Williams & Wilkins.

A classic book on study design. They suggest that **for a correlation study**, **30 participants** is a **minimum rough rule of thumb**, but **larger samples** are needed if detecting **weak to moderate correlations**.

□ Source: Hulley, S. B., Cummings, S. R., Browner, W. S., Grady, D. G., & Newman, T. B. (2013). *Designing clinical research* (4th ed.). Lippincott Williams & Wilkins.

3. Cohen, J. (1988).

Title: Statistical Power Analysis for the Behavioral Sciences (2nd ed.).

Publisher: Routledge Academic.

- Cohen provided guidelines for effect sizes and sample size calculations:
- ✓ To detect a small correlation ($r \approx 0.1$), you need about 783 samples.
- ✓ For a **medium correlation** ($r \approx 0.3$), you need about **84** samples.
- ✓ For a large correlation ($r \approx 0.5$), around 28 samples are sufficient.
- ✓ So, if you're expecting a large correlation, your sample size of n = 20 is nearly acceptable, but slightly underpowered for detecting medium to small correlations.
- □ Source: Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

4. Mukaka, M. M. (2012).

Title: A guide to appropriate use of Correlation coefficient in medical research.

Journal: Malawi Medical Journal, 24(3), 69-71.

Mukaka mentions that for reliable Pearson correlation studies, larger sample sizes are preferable but does not give a strict cutoff — just that $n \ge 30$ is commonly aimed for.

[□] Source: Mukaka, M. M. (2012). A guide to appropriate use of correlation coefficient in medical research. *Malawi Medical Journal, 24*(3), 69–71. https://www.ajol.info/index.php/mmj/article/view/81588

EXAMPLE SCIENTIFIC RESEARCH USING PEARSON CORRELATION

Healthscope 2021, Vol 4(1): 31-37

RESEARCH ARTICLE

Spatio-Temporal analysis of Dengue distribution pattern in Kuantan from 2015-2019

Mohamad Hizham Mohamad Hanapi, Farah Ayuni Shafie*

Centre of Environmental Health and Safety, Faculty of Health Sciences, Universiti Teknologi MARA Cawangan Selangor Kampus Puncak Alam, 42300 Bandar Puncak Alam, Selangor, Malaysia;

This study wants to prove the linear relationship between the high number of dengue cases to the high amount of rainfall and vice versa.

Table 3. Pearson correlation analysis of dengue cases per month and rainfall amount per month.

Pearson Correlation Results	Dengue case per month	Rainfall amount per month
R ² linear		0.003
Pearson Correlation, r		0.053
Significance 2-Tailed		0.686

Furthermore, the Pearson correlation (r) value in this study is not even close to 1. The Pearson correlation coefficient, r, can take a range of values from +1 to -1. A value of 0 indicates that there is no relationship between the two variables. Values greater than 0 indicate a positive relationship; that is, as the value of one variable increases, so does the value of the other variable. These analyses prove that the amount of rainfall does not affect the increase in the number of dengue cases in Kuantan for 2015 to 2019.

SPEARMAN RANK CORRELATION

OBJECTIVES



Explain the concepts and principles of Spearman rank correlation



Discuss the requirements and application of Spearman rank correlation test



Perform Spearman rank correlation test manually and using SPSS software, interpret and report the results.

What is Spearman Rank Correlation?

- Spearman Rank Correlation, is a nonparametric measure of the strength and direction of the association between two ranked variables.
- Often denoted as ρ (rho) or r_s ,
- Is the Non-parametric version of the Pearson correlation.
- Determines the strength and direction of the monotonic relationship between 2 variables.
- It does not assume a linear relationship or normally distributed data like Pearson correlation.

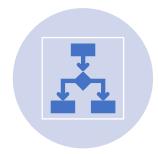
A monotonic relationship is a relationship that does one of the following:



As the value of one variable increases, so does the value of the other variable; OR



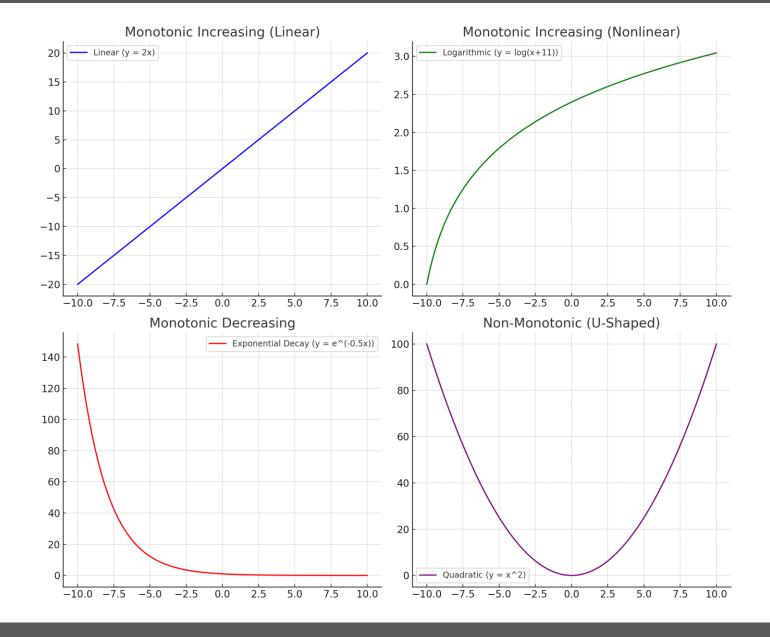
As the value of one variable increases, the other variable value decreases.



Monotonicity is "less restrictive" than that of a linear relationship.

Key Point:

A monotonic relationship does **not require a straight-line (linear) relationship** but ensures that the direction of change is consistent.



Key Principles for Spearman Rank Correlation

- **1. Rank-based:** Rather than using the actual data values, Spearman works by ranking the data first.
- 2. Monotonic relationship:

If one variable increases, the other also increases (positive monotonic).

If one variable increases, the other decreases (negative monotonic).

Range:

+1 = perfect positive association

0 = no association

-1 = perfect negative association

Aspect	Requirement
Type of Variables	Ordinal, Interval, or Ratio data (but must be converted to ranks)
Relationship Type	Monotonic (not necessarily linear)
Sample Size	Ideally n ≥ 10; larger samples increase reliability
Assumptions	Observations are independentVariables should be at least ordinalMonotonic relationship expected

Requirements



When to Use Spearman Instead of Pearson?

- When your data is ordinal (ranked but not measured).
- When the relationship between variables is non-linear but monotonic.
- When your data violates normality assumptions (skewed distributions, outliers).
- When you want a robust correlation that's less sensitive to outliers.
 - □Spearman = robust because it only cares about rank order, not actual size differences.

Applications

- □Social Sciences: Rank relationship between educational level and income satisfaction.
- ☐ Medical Research: Rank association between severity of disease symptoms and quality of life.
- ☐ Environmental Studies: Rank pollution levels and biodiversity loss.
- ☐ Marketing: Rank customer satisfaction scores and product loyalty.

Spearman's rank correlation coefficient, r_s

$$ho = 1 - rac{6\sum d_i^2}{n(n^2-1)}$$

P = Spearman's rank correlation coefficient

 d_i = difference between the two ranks of each observation

n = number of observations

Performing Spearman Rank Correlation (Manual Calculation)

Let us consider the following example data regarding the marks achieved in a maths and English exam:

Exam	Marks									
English	56	75	45	71	61	64	58	80	76	61
Maths	66	70	40	60	65	56	59	77	67	63

Dataset: english_maths_dataset_for_spearman.xlsx

Step 1: Rank the English and Math marks separately.

Higher marks = Higher rank (1 = highest)

English Marks	Rank (English)
80	1
76	2
75	3
71	4
64	5
61	6.5 (tie)
61	6.5 (tie)
58	8
56	9
45	10

Maths Marks	Rank (Maths)
77	1
70	2
67	3
66	4
65	5
63	6
60	7
59	8
56	9
40	10

Step 2: Create a Full Table

Student	English Marks	Rank E	Maths Marks	Rank M	d = (Rank E - Rank M)	d²
1	56	9	66	4	5	25
2	75	3	70	2	1	1
3	45	10	40	10	0	0
4	71	4	60	7	-3	9
5	61	6.5	65	5	1.5	2.25
6	64	5	56	9	-4	16
7	58	8	59	8	0	0
8	80	1	77	1	0	0
9	76	2	67	3	-1	1
10	61	6.5	63	6	0.5	0.25

Step 3: Sum d^2

$$\sum d^2 = 25 + 1 + 0 + 9 + 2.25 + 16 + 0 + 0 + 1 + 0.25 = 54.5$$

Step 4: Apply the Spearman Rank Correlation Formula

The formula:

$$r_s = 1 - rac{6 \sum d^2}{n(n^2 - 1)}$$

Where:

n = 10 (number of students)

Substituting:

$$r_s = 1 - \frac{6(54.5)}{10(10^2 - 1)} = 1 - \frac{327}{10(99)} = 1 - \frac{327}{990} = 1 - 0.33 = 0.67$$

ightharpoonup Spearman's rank correlation coefficient $r_s=0.67$

Interpretation:

- r_s =0.67 means a moderate to strong positive correlation between English and Math marks.
- Students who perform well in English tend to also perform well in Math, although not perfectly.

Spearman rank correlation coefficient is r_s =0.67, indicating a moderate positive relationship between Math and English exam marks.

```
modifier_ob.
  mirror object to mirror
mirror_object
 peration == "MIRROR_X":
irror_mod.use_x = True
"Irror_mod.use_y = False
__mod.use_z = False
 _operation == "MIRROR_Y"
 irror_mod.use_x = False
 lrror_mod.use_y = True
 lrror_mod.use_z = False
 operation == "MIRROR_z"
  rror_mod.use_x = False
  lrror_mod.use_y = False
  lrror_mod.use_z = True
  welection at the end -add
   ob.select= 1
   er ob.select=1
   ntext.scene.objects.action
  "Selected" + str(modified
   rror ob.select = 0
  bpy.context.selected_obj
  ata.objects[one.name].sel
  int("please select exactle
  --- OPERATOR CLASSES ----
      mirror to the selected
   ject.mirror_mirror_x"
 ext.active_object is not
```

Steps to Enter and Analyze in SPSS

Step 1: Open SPSS

• Open a new dataset.

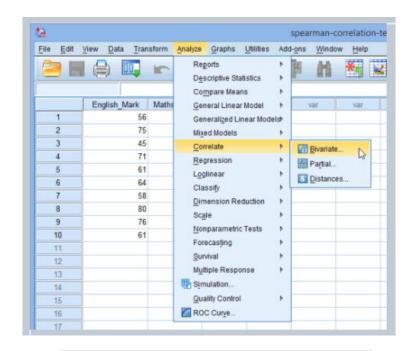
Step 2: Define Variables

- In Variable View:
- First variable name: English (Type: Numeric)
- Second variable name: Math (Type: Numeric)

Step 3: Enter / Import Data

Step 4: Perform Spearman Correlation

- 1. Click Analyze → Correlate → Bivariate.
- 2. Move both variables (English and Math) into the Variables box.
- 3. In the Correlation Coefficients section:
 - Tick Spearman.
 - Untick Pearson.
- 4. Ensure "Two-tailed" is selected under Test of Significance.
- 5. Click OK.
- SPSS will run the test and give you an output table.





SPSS Output

		Correlations			
			English	Maths	
Spearman's rho	English	Correlation Coefficient	1.000	.669*	
		Sig. (2-tailed)		.035	
		N	10	10	
	Maths	Correlation Coefficient	.669*	1.000	
		Sig. (2-tailed)	.035		
		N	10	10	
*. Correlation is significant at the 0.05 level (2-tailed).					

 r_s =0.670 (matches our manual calculation) SPSS will also show the p-value (e.g., if p < 0.05, the correlation is significant).

How to Report?

A Spearman rank-order correlation was conducted to assess the relationship between English and Math exam marks.

There was a moderate positive correlation between English and math marks, which was statistically significant, r_s =0.669,p<0.05.

Interpretation: The higher you ranked in math, the higher you ranked in English also, and vice versa.

SUMMARY

Pearson (r)

Measures the strength and direction of a linear relationship between two continuous variables.

- Direction:
 - Positive (个个): both variables increase together
 - Negative (↑↓): one increases as the other decreases

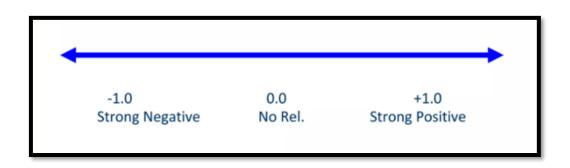
Nearson Assumptions

- Continuous variables (interval/ratio scale)
- Linear relationship (check scatterplot)
- Normal distribution
- No extreme outliers
- Homoscedasticity (equal spread of residuals)

Spearman (ρ or r_s)

Measures the **strength and direction of a monotonic relationship** using **ranked data**.

- ✓ Non-parametric, robust to outliers
- * Suitable when data is **ordinal**, **non-linear**, or **non-normal**



THANK YOU