Workshop 4
How to conduct correlation analysis and check scale reliability

Dr. Naser Alziyadat

Murdoch
UNIVERSITY

RELIABILITY AND VALIDITY

Murdoch UNIVERSITY

These important concepts relate to any measure/instrument

1. **Validity of a measure**: The degree to which a measure actually measures what we think it measures.

2. **Reliability of a measure:** The consistency of a measure over time or under similar situations.

In social, or business research we cannot expect any measure to be perfectly valid or perfectly reliable. The best we can do is to design our measures to be as valid and reliable as they can be and where feasible run some statistical tests to evaluate them.

.

Scale reliability

- It is important that your scales to measure the phenomena are reliable.
- Reliability concerns the internal consistency.
- It represents the degree to which items that make up the scale hang together (i.e. are they measuring the underlying construct?)
- Scale reliability is measured by Cronbach's alpha coefficient.

Cronbach's alpha

- Activity: Find the scale reliability for the ICSR scale that contains 20 items (D1-I4).
- Procedure- Analyze-Scale-Reliability analysis-choose the individual items that make up the scale-Model section, make sure alpha is selected-Type your scale name in the scale label box-Statistics, select: item, scale, and scale if item deleted- inter-item section click on correlations-in the Summaries section, click on correlations.

Cronbach's alpha: interpretation and reporting

- Check the number of cases is correct in the Case Processing Summary

Case Processing Summary

		N	%
Cases	Valid	27	100.0
	Excluded[a]	0	.0
	Total	27	100.0

a. Listwise deletion based on all variables in the procedure.

- Inter-item Correlation Matrix. Check for negative values.
- All values should be positive indicating that they measure the same underlying characteristic.
- The presence of negative values may indicate that some items have not been reverse scored.
- Check Cronbach's alpha value

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.787	.788	20

Cronbach’s alpha

- It is important that your scales to measure the Values of the Cronbach's alpha statistic range from 0 to 1 (Gliem & Gliem 2003).
- A value of 0.9 or over is considered to indicate an excellent level of internal consistency of a scale.
- A value between 0.8- 0.89 is deemed good, 0.7-0.79 acceptable, and 0.6-0.69 questionable (George & Mallery 2003, cited in Gliem & Gliem 2003).
- The value of Cronbach's alpha is partly dependent on the number of items in a scale and the mean inter-item correlation (Gliem & Gliem 2003; Pallant 2011); therefore, scales with fewer items than 10 may have small values. In such a case, it is important for the scale to have a mean inter-item correlation between 0.2 and 0.4 (Briggs & Cheek 1986, cited in Pallant 2011).

Cronbach's alpha: interpretation and reporting

- Corrected Item-Total Correlation in the Item-Total Statistics Table gives you an indication of the degree to which each item correlates to the total score.
- Alpha if item deleted gives you indications about the impact on the cronbach's alpha if the item is deleated.

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
Does your organisation integrate social and environmental masalih and mafasidin addition to the economic ones when conducting environmental scanning?	20.6296	42.550	.550	.864	.764
Does your organisation refine the strategic objectives in order to meet social and environmental masalih in addition to economic masalih?	20.4815	42.798	.567	.783	.764
Do the key performance indicators KPIs of your organisation include social and environmental masalih in addition to the economic ones?	20.7407	44.815	.383	.788	.776

Cronbach’s alpha: reporting Sample

- Using SPSS software, Cronbach's alpha statistic was estimated to test the reliability of the scale included in the questionnaire in the present research. The Crornbach's alpha for the scale of the process of implementing ICSR is 0.787 which is deemed to be acceptable (Pallant, 2016).
- Now, identify the items that when deleted alpha is going to increase.

Statistical techniques by purpose

Murdoch UNIVERSITY

- Exploring relationships (techniques covered in this unit).
  - Correlation:
    - Pearson or spearman correlations
    - To explore the strength and the direction of the relationship between two continuous variables.
  - Partial correlation (continuous variables):
    - An extension of spearman correlation
    - Allows you to control for the possible effects of another confounding variable.
  - Multiple regression (continuous variables):
    - A sophisticated extension of correlation
    - Allows you to explore the predictive ability of a set of independent variables on one continuous dependent variable.

CORRELATION

- Correlation examines the strength of a connection between two characteristics belonging to the same individual/event/equipment
- The concept of correlation **does not** include the proposition that one thing is the cause and the other the effect
- We merely say that two things are systematically connected

RELATIONSHIPS

- Two variables can be positively correlated - an increase in one variable coincides with an increase in another variable e.g. the more electricity used the higher the power bill.

- A negative correlation - when one variable increases as the other decreases e.g. as price increases, demand decreases.

- A zero or random correlation - when variations in two variables occur randomly e.g. number of accountants graduating per year with total annual attendance at national football league matches

THE CORRELATION COEFFICIENT (R)

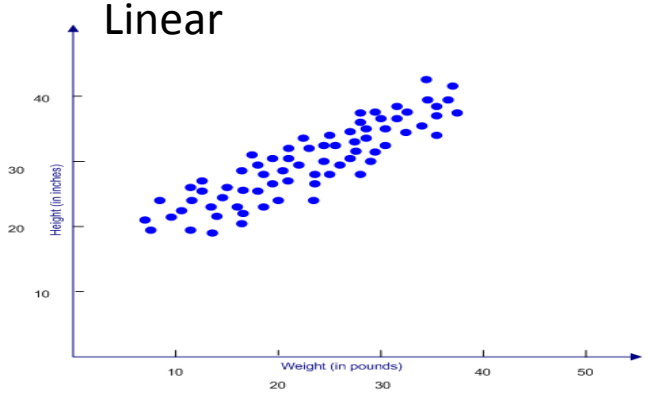
- The Correlation is measured by the correlation coefficient which is usually designated as ‘r’.

- Correlations (r) range from +1.00 perfect positive to -1.00 perfect inverse with a midpoint 0.00 indicating absolute randomness

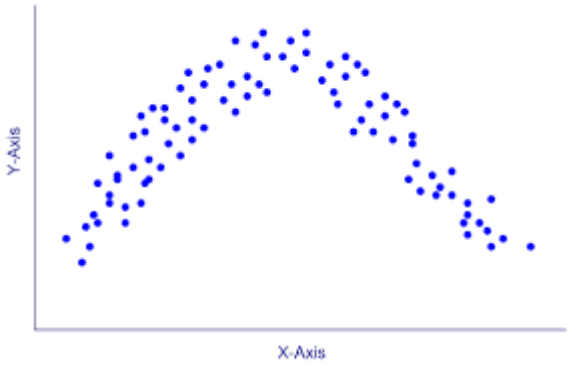
Factors to consider when interpreting correlation

- Non-linear relationship
  - The correlation coefficient (e.g. Pearson r) provides an indication of the linear (straight-line) relationship between variables.
  - In situations where the two variables are related in non-linear fashion (e.g. curvilinear), Pearson r will seriously underestimate the strength of the relationship.
  - Always check the scatterplot, particularly if you obtain low values of r.

Linear

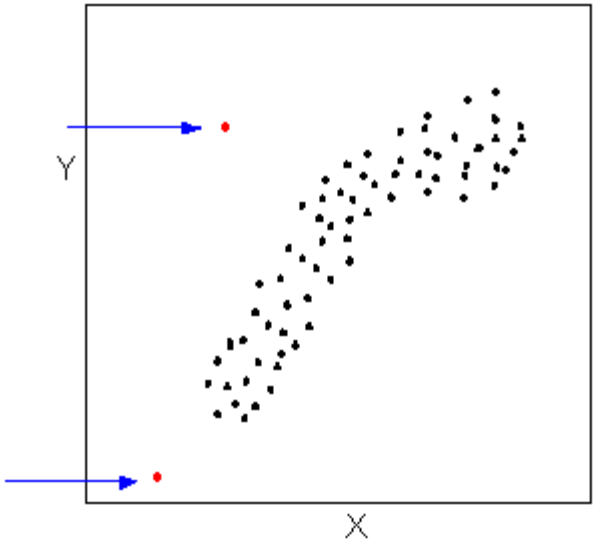


Non-linear



Factors to consider when interpreting correlation

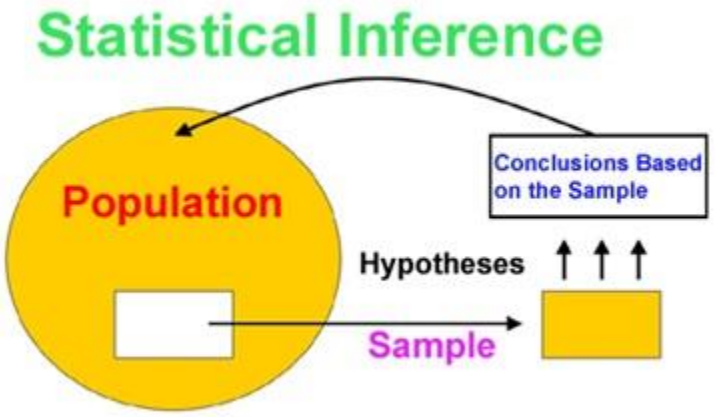
- Outliers
  - values that are substantially lower or higher than the other values in the data set.
  - can have a dramatic effect on the correlation coefficient, particularly in small samples.
  - they could be due to error:
    - a careless answer from a respondent>>>check for errors and correct if appropriate.
    - a true value from a rather strange individual >>>consider removing or recoding the offending value.



Note: in your assignment or final exam, do not remove the outliers due to time restrictions.

Factors to consider when interpreting correlation

- Restricted range of scores
  - Correlation coefficients from studies using a restricted range of cases are often different from studies where the full range of possible scores are sampled.
  - Overgeneralization.
  - Best sampling techniques.
  - Large data sets.



Factors to consider when interpreting correlation

- Correlation versus causality
  - Correlation does not indicate that one variable *causes* the other.
  - The correlation between two variables (A and B) could be due:
    - A causes B.
    - B causes A.
    - A & B complete each other.
    - There is another third variable (V) causes both A and B. (e.g. Ice-cream, violence, and hot weather)
  - watch out for the possible influence of a third, confounding variable when designing your own study.
  - If you suspect the possibility of other variables that might influence your result, you can use partial correlation.

Assumptions for correlation

- Level of measurement
  - Data must be interval or ratio (continuous) (Pearson)
  - The exception to this is if you have one dichotomous independent variable (with only two values: e.g. sex) and one continuous.
    - You should have roughly the same number of people or cases in each category of the dichotomous variable.
  - Spearman's rho correlation is suitable for ranked or ordinal data (e.g Likert Scale). It is a non-parametric alternative to Pearson as well.
- Related Pairs
  - Each subject (respondent) must provide a score on both variable X and variable Y.
  - Both pieces of information must be from the same subject (Respondent).

Assumptions for correlation

- Independence of observations
  - The observations that make up your data must be independent of one another.
  - Each observation or measurement must not be influenced by any other.
  - Any situation where the observations or measurements are collected in a group setting, or subjects are involved in some form of interaction with one another, should be considered suspect.

Assumptions for correlation

- Normality
  - Scores on each variable should be normally distributed. What if not?
- Linearity
  - The relationship between the two variables should be linear. What if not?

Missing data

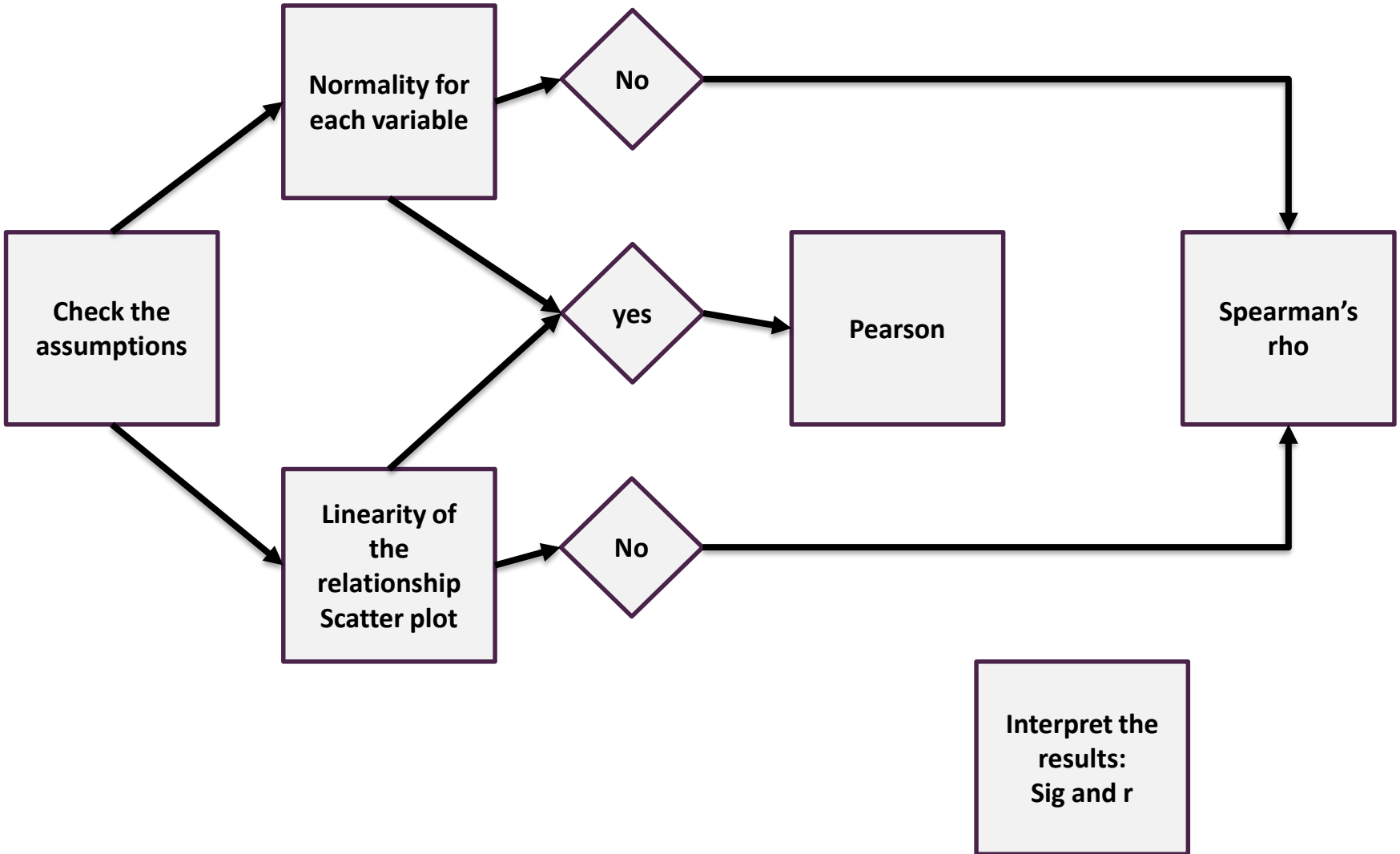
- Obtaining complete data is very rare in human sciences. Thus, it is important to check for missing data.
- If you have missing data, then you can do one of the following from the menu of the specified analysis:
  - Exclude cases listwise option: A case will be totally excluded from all the analyses if it is missing even one piece of information. This may affect your sample size and test results.
  - Exclude cases pairwise option: excludes the cases only if they are missing the data required for the specific analysis. They will still be included in any of the analyses for which they have the necessary information.
  - Use the mean (in some tests). But it should be used with caution.

Correlation

- Correlation analysis is used to describe the strength and direction of the linear relationship between two variables.
- We are going to cover two types of correlation:

  - Pearson product-moment correlation (r).
    - It can be used with two continuous variables; and one continuous variable with a dichotomous variable (M/F) (Parametric).
  - Spearman Rank Order Correlation (rho):
    - Particularly useful when your data does not meat the requirements of Pearson correlation (Non-parametric).

How to decide on correlation



Practical example

- Access the data file survey.sav
  - The survey was designed to explore the factors that affect respondents' physiological adjustment and wellbeing.
  - We are going to assess the correlation between respondents' feelings of control and their level of perceived stress.
- Example of research question:
  - Is there a relationship between the amount of control people have over their internal states and their levels of perceived stress?
  - Do people with high levels of perceived control experience lower levels of perceived stress?

Practical example

- Preliminary analysis for correlation
  - Check the assumptions by generating a scatter plot. They are both continuous variables (scale)>>> this means that we have to go for the parametric Pearson correlation after checking the assumptions of linearity and normality.
    - Dependent variable: total perceived stress (total perceived stress)
    - Independent variable : amount of control over internal state (total PCOISS).

Test the assumption of linearity

Procedure for generating a scatterplot

1. From the menu at the top of the screen click on: **Graphs**, then click on **Scatter**.

2. Click on **Simple**. Click on the **Define** button.

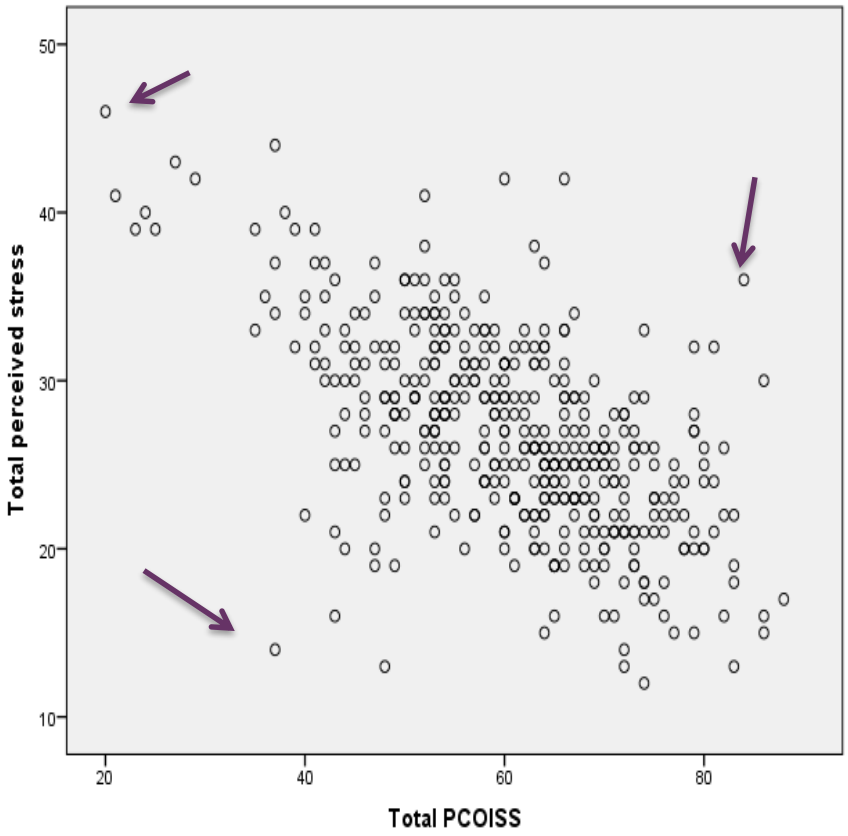
3. Click on the first variable and move it into the **Y-axis** box (this will run vertically).

By convention, the dependent variable is usually placed along the Y-axis (e.g. total perceived stress).

4. Click on the second variable and move to the **X-axis** box (this will run across the page). This is usually the independent variable (e.g. total PCOISS).

5. If you would like to add a title, click on **Titles**. Type in a title. Click on **Continue** and then **OK**.

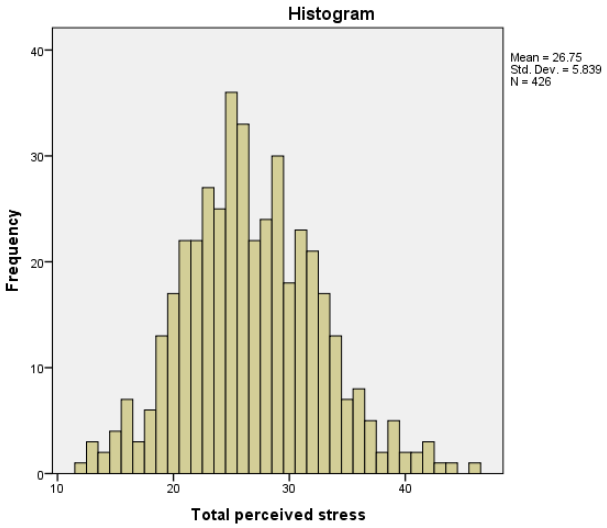
- Check for outliers
  - Outliers can seriously influence some analyses, so this is worth investigating.
  - were the data entered correctly?
  - could these values be errors?
- Inspecting the distribution of data points
  - Are the data points spread all over the place? This suggests a very low correlation.
  - Are all the points neatly arranged in a narrow cigar shape? This suggests quite a strong correlation.
  - Could you draw a straight line through the main cluster of points, or would a curved line better represent the points? Pearson assumes linear relationship.
  - Is the shape even from one end to the other? Or does it start off narrow and then get fatter? If this is the case, your data may be violating the assumption of homoscedasticity.
- Determine the direction of the relationship between the variables

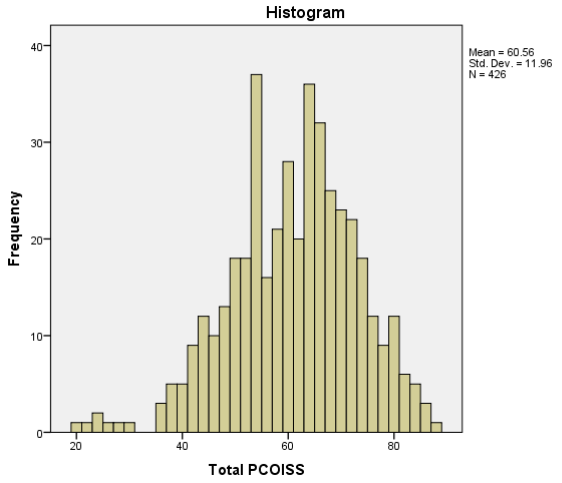


Test the assumption of normality

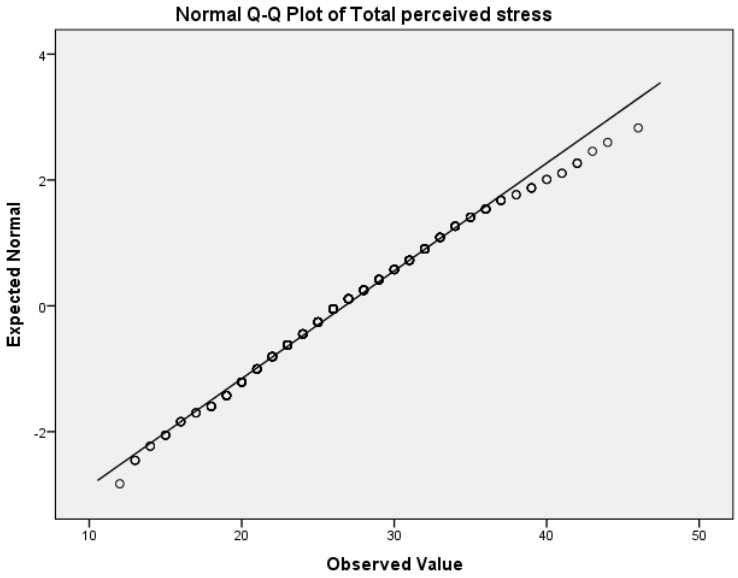
Descriptives

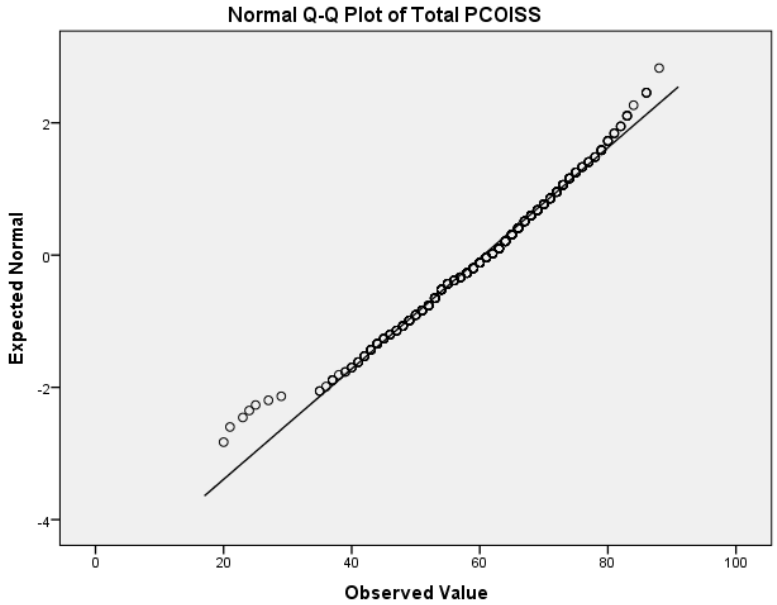
			Statistic	Std. Error
Total perceived stress	Mean		26.75	.283
	95% Confidence Interval for Mean	Lower Bound	26.20	
		Upper Bound	27.31	
	5% Trimmed Mean		26.65	
	Median		26.00	
	Variance		34.098	
	Std. Deviation		5.839	
	Minimum		12	
	Maximum		46	
	Range		34	
	Interquartile Range		8	
	Skewness		.274	.118
	Kurtosis		.149	.236
Total PCOISS	Mean		60.56	.579
	95% Confidence Interval for Mean	Lower Bound	59.42	
		Upper Bound	61.70	
	5% Trimmed Mean		60.86	
	Median		62.00	
	Variance		143.033	
	Std. Deviation		11.960	
	Minimum		20	
	Maximum		88	
	Range		68	
	Interquartile Range		16	
	Skewness		-.408	.118
	Kurtosis		.275	.236





Test the assumption of normality





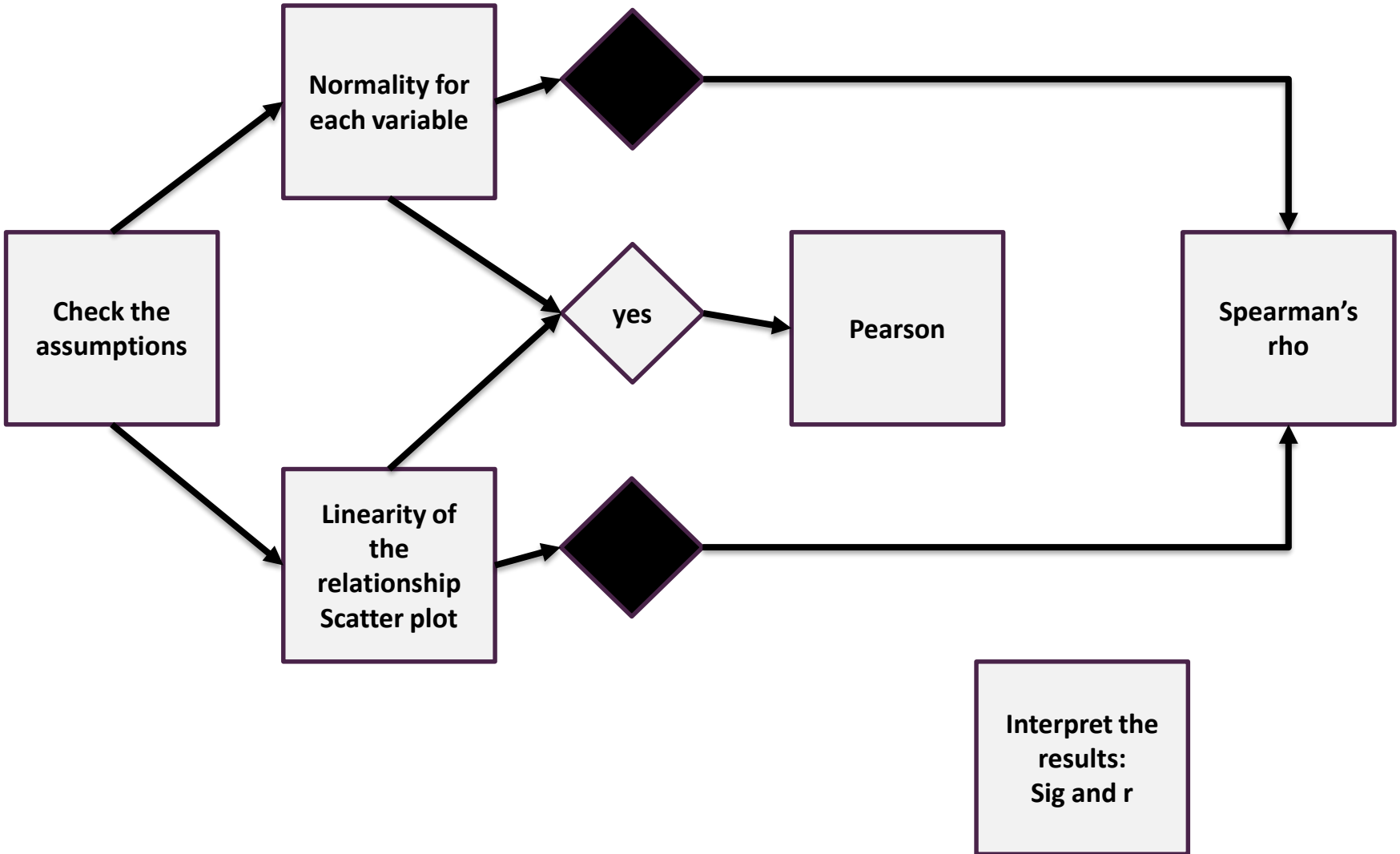
Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Total perceived stress	.070	426	.000	.992	426	.016
Total PCOISS	.060	426	.001	.987	426	.001

a. Lilliefors Significance Correction

Remember to compromise the 5 criteria of normality assessment…

Remember


Normality for each variable
Check the assumptions
yes
Pearson
Spearman's rho
Linearity of the relationship Scatter plot
Interpret the results: Sig and r

Practical example

Procedure for calculating Pearson product-moment correlation

1. From the menu at the top of the screen click on: **Analyze**, then click on **Correlate**, then on **Bivariate**.

2. Select your two variables and move them into the box marked **Variables** (e.g. total perceived stress, total PCOISS).You can list a whole range of variables here, not just two. In the resulting matrix, the correlation between all possible pairs of variables will be listed. This can be quite large if you list more than just a few variables.

3. Check that the **Pearson** box and the **2 tail** box have a cross in them. The two-tail test of significance means that you are not making any specific prediction concerning the direction of the relationship between the variables (positive/negative).You can choose a **one-tail** test of significance if you have reasons to support a specific direction.

4. Click on the **Options** button.

For **Missing Values**, click on the **Exclude cases pairwise** box.

Under **Options** you can also obtain means, standard deviations if you wish. Click on **Continue**.

5. Click **OK**.

The direction of the relationship.
The *more* control people feel they have, the *less* stress they experience

Information about the sample

The strength of the relationship.
r=.10 to .29 or r=–.10 to –.29 small
r=.30 to .49 or r=–.30 to –.4.9 medium
r=.50 to 1.0 or r=–.50 to –1.0 large

Correlations

		Total perceived stress	Total PCOISS
Total perceived stress	Pearson Correlation	1	-.581**
	Sig. (2-tailed)		.000
	N	433	426
Total PCOISS	Pearson Correlation	-.581**	1
	Sig. (2-tailed)	.000	
	N	426	430

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

Significance level
indicates a (%) risk of concluding that a difference exists when there is no actual difference. Most of the times it is pre-determined as 5%.

Reporting the results

Correlations

		Total perceived stress	Total PCOISS
Total perceived stress	Pearson Correlation	1	-.581**
	Sig. (2-tailed)		.000
	N	433	426
Total PCOISS	Pearson Correlation	-.581**	1
	Sig. (2-tailed)	.000	
	N	426	430

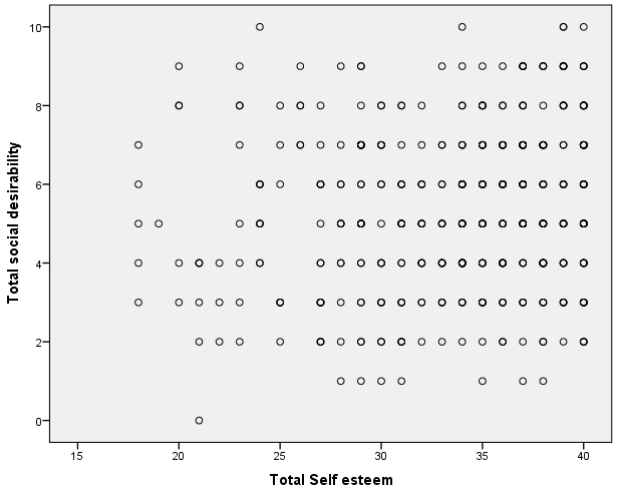
**. Correlation is significant at the 0.01 level (2-tailed).

The relationship between perceived control of internal states (as measured by the PCOISS) and perceived stress (as measured by the Perceived Stress scale) was investigated using Pearson product-moment correlation coefficient. Preliminary analyses were performed to ensure no violation of the assumptions of normality, linearity and homoscedasticity. There was a strong, negative correlation between the two variables [r=–.58, n=426, p<.0005], with high levels of perceived control associated with lower levels of perceived stress.

Another example

- We are interested in measuring the association between the Total Self-esteem variable and the Total Desirability.
- Both variables are ‘continuous’ (scale)>>> We target the parametric technique (Pearson).
- Test the assumption of linearity>>>generate a scatter plot.
- Test the assumption of normality

Another example



Descriptives

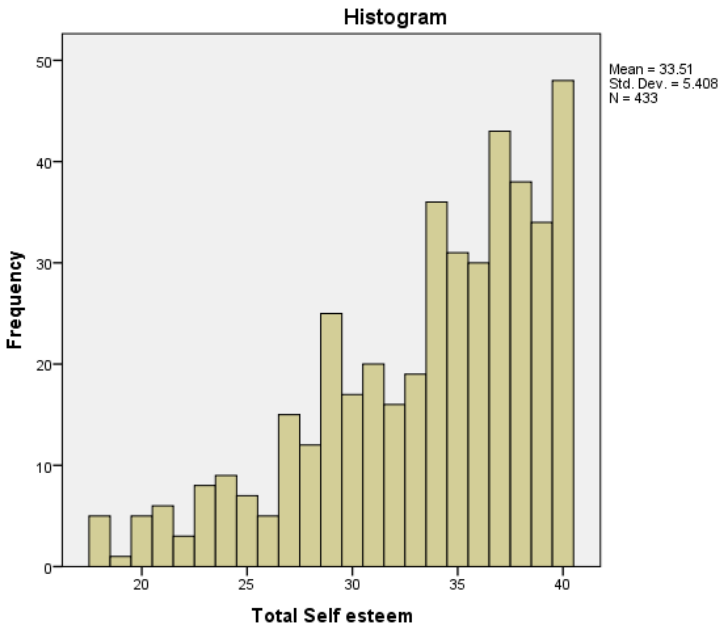
			Statistic	Std. Error
Total Self esteem	Mean		33.51	.260
	95% Confidence Interval for Mean	Lower Bound	33.00	
		Upper Bound	34.02	
	5% Trimmed Mean		33.89	
	Median		35.00	
	Variance		29.246	
	Std. Deviation		5.408	
	Minimum		18	
	Maximum		40	
	Range		22	
	Interquartile Range		8	
	Skewness		-.866	.117
	Kurtosis		.040	.234
Total social desirability	Mean		5.30	.098
	95% Confidence Interval for Mean	Lower Bound	5.11	
		Upper Bound	5.50	
	5% Trimmed Mean		5.29	
	Median		5.00	
	Variance		4.171	
	Std. Deviation		2.042	
	Minimum		0	
	Maximum		10	
	Range		10	
	Interquartile Range		3	
	Skewness		.089	.117
	Kurtosis		-.622	.234

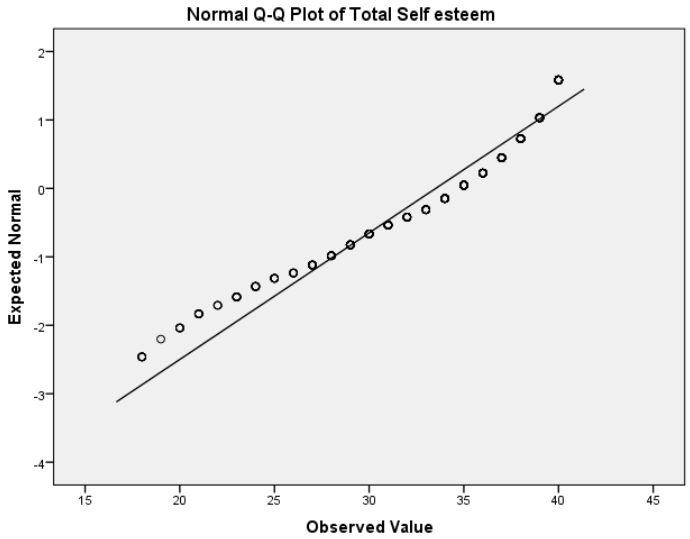
Tests of Normality

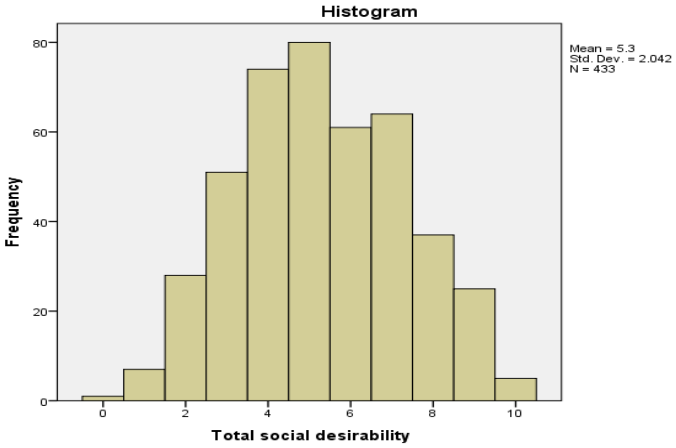
	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Total Self esteem	.136	433	.000	.916	433	.000
Total social desirability	.116	433	.000	.971	433	.000

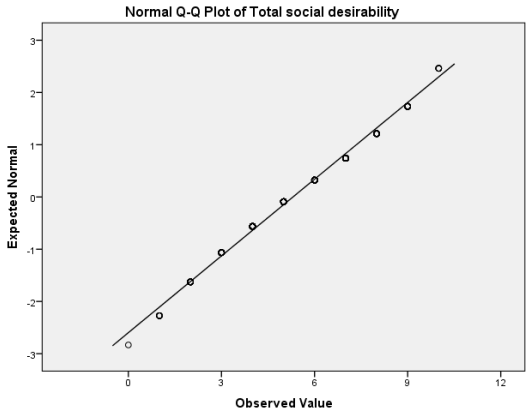
a. Lilliefors Significance Correction

Another example

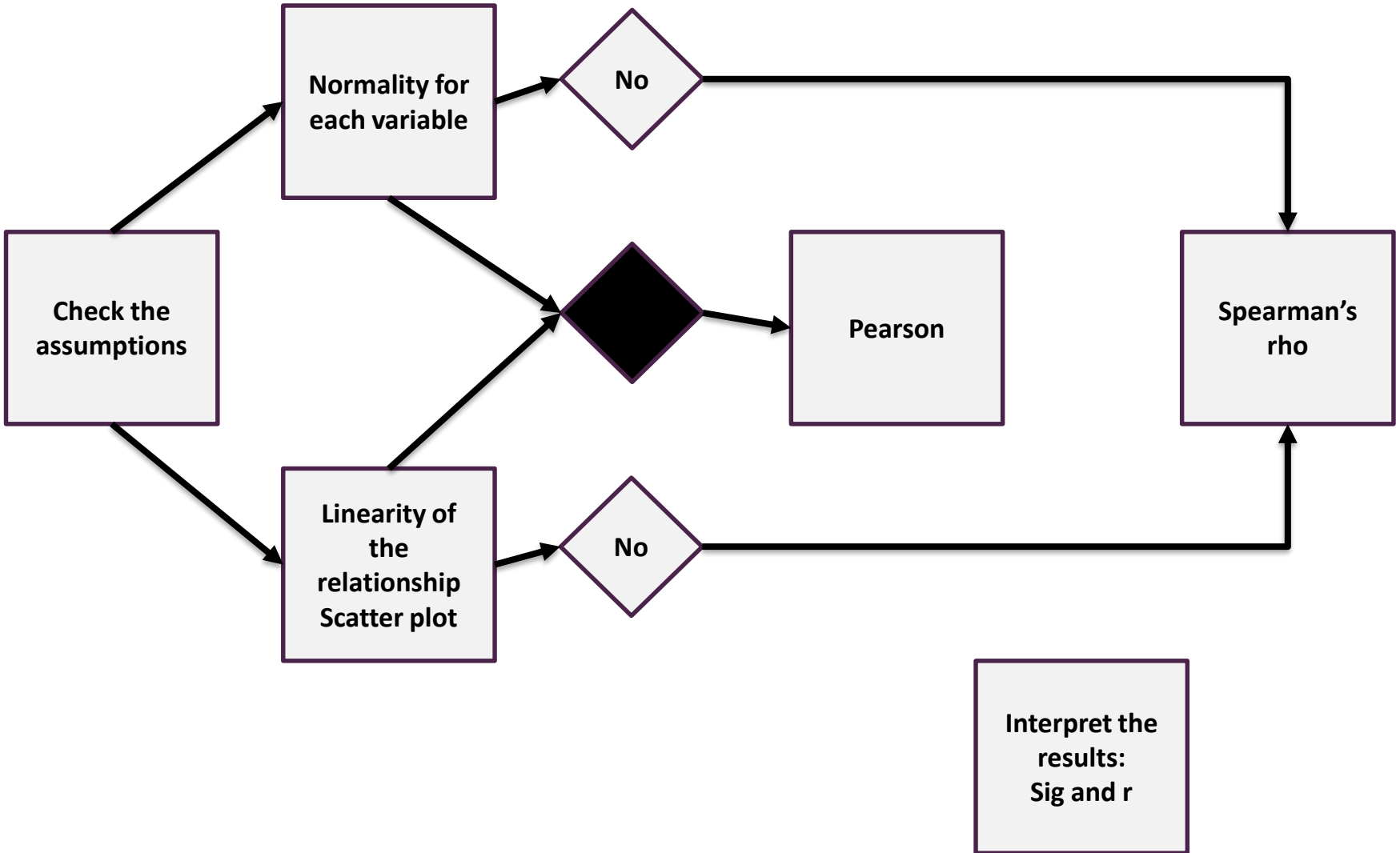








Remember


Check the assumptions
Normality for each variable
No
Linearity of the relationship Scatter plot
No
Pearson
Spearman's rho
Interpret the results: Sig and r

Procedure for calculating Spearman

1. From the menu at the top of the screen click on: **Analyze**, then click on **Correlate**, then on **Bivariate**.

2. Select your two variables and move them into the box marked

3. Check that the **Sperman** box and the **2 tail** box have a cross in them.
4. Click on the **Options** button. For **Missing Values**, click on the **Exclude cases pairwise** box. Under **Options** you can also obtain means, standard deviations if you wish. Click on **Continue**, then Click **OK**.

The direction of the relationship.
The *more* control people feel they have, the *less* stress they experience

Information about the sample

The strength of the relationship.
r=.10 to .29 or r=–.10 to –.29 small
r=.30 to .49 or r=–.30 to –.4.9 medium
r=.50 to 1.0 or r=–.50 to –1.0 large

Correlations

			Total social desirability	Total Self esteem
Spearman's rho	Total social desirability	Correlation Coefficient	1.000	.157**
		Sig. (2-tailed)	.	.001
		N	433	433
	Total Self esteem	Correlation Coefficient	.157**	1.000
		Sig. (2-tailed)	.001	.
		N	433	436

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

Significance level
indicates a (%) risk of concluding that a difference exists when there is no actual difference. Most of the times it is pre-determined as 5%.

Reporting the results

Correlations

			Total social desirability	Total Self esteem
Spearman's rho	Total social desirability	Correlation Coefficient	1.000	.157**
		Sig. (2-tailed)	.	.001
		N	433	433
	Total Self esteem	Correlation Coefficient	.157**	1.000
		Sig. (2-tailed)	.001	.
		N	433	436

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

The relationship between Total self-esteem and Total social desirability was investigated using Spearman's rho correlation after the assumptions of linearity and normality had been violated. There was a small, positive correlation between the two variables [r=.157, n=433, p<.001], with high levels of self esteem associated with highest levels of social desirability.

A third example

- We are interested in measuring the association between the variables agegp5 and edurec.
- Both variables are ‘ordinal’ >>> We target the non-parametric technique (Spearman).
- NO NEED TO test the assumption of linearityand normality.

Reporting the results

Correlations

			age 5 groups	educat recoded
Spearman's rho	age 5 groups	Correlation Coefficient	1.000	-.056
		Sig. (2-tailed)	.	.237
		N	439	439
	educat recoded	Correlation Coefficient	-.056	1.000
		Sig. (2-tailed)	.237	.
		N	439	439

The relationship between the two ordinal variables (Age 5 groups and Educat recorded) investigated using Spearman's rho correlation There was no statistically significant correlation [r=–.056, n=439, p>.05].

Murdoch
UNIVERSITY

Thank you
Please refer to the LMS and complete the tasks after the workshop