NS5353 Week 5 Workbook

SPSS and Inferential Statistics

Preface:

There are a wide range of inferential statistics used in quantitative research designs. However, due to the limited time constraints in this short intensive 6 week subject, we will cover only a small range of inferential statistics, providing you with an introduction to performing basic inferential statistics using SPSS.

Completing this workbook is essential to understanding the task required for Assessment Task 2(a) and 2(b). When deciding on research questions and statistical tests to use to answer your research questions, you must choose the appropriate statistical test from those presented in this workbook.

Table of Contents:

Background Information about the data:

The data set provided for use alongside this workbook is available for download in your Week 5, Topic 3 content. The data set refers to a fictitious study that involves testing two interventions that were developed to help students overcome their anxiety concerning a statistics exam. Students were randomly divided into two groups (**maths skills** and **confidence building**) and asked to complete a number of scales. These were the Fear of Statistics Test (**fear of stats time 1)**, Confidence in coping with Statistics Scale (**confidence scale score time 1**) and Anxiety Scale (**anxiety scale score time 1**). One group (**group 1 – maths skills**) was given sessions designed to improve maths skills. The other group (**group 2 – confidence building**) was given sessions designed to build confidence in the ability to cope with statistics. After the program, the subjects were asked to complete the scales again (**fear of stats test score time2, confidence scale score time2** and **anxiety scale score time2**). Their performance on the stats exam was also measured.

Scales used to collect data:

Fear of Statistics test:
- 10 questions measured on a 5-point Likert scale.
- Scores range between 10 and 50
- High scores indicate high fear

Confidence in coping with Statistics Scale:
- 5 questions measured on a 5-point Likert scale
- Scores range between 5 and 25
- High scores indicate high confidence

Anxiety Scale:
- 10 questions measures on a 5-point Likert scale.
- Scores range between 10 and 50
- High scores indicate high anxiety

Exam:
- Score out of 100

Grade:
- 70 or above is considered a pass grade

Effort
- Students were given a rating of "poor, average or excellent" at the end of the intervention. This was how much effort the subject coordinator thought the students gave throughout the class.

Some supplementary content before beginning to use SPSS:

Before beginning to use SPSS to undertake inferential statistical tests, it is vital that you understand the content provide in Box 1 below.

Decisions…..decisions….Which statistical test/analysis should I use?

It is essential that you choose the right test/analysis to answer your research question. Your choice of test depends on:

1. Your research question
2. Type of variables (numerical/categorical)
3. If numerical data (parametric/non-parametric)
4. Type of data (paired/unpaired)

Type of variables: We have already learnt about research questions in previous units, so let's talk about the type of variable. You need to decide if you have numerical or categorical variables. To run inferential statistical analysis, you need 2 or more variables, this means you will have either…

- 1 numerical, 1 categorical
- 2 categorical
- 2 numerical

For numerical variables: When numerical variables are involved, you will need to determine if the test is ***parametric or non-parametric*** *and also determine if you have* ***paired or unpaired data.***

Parametric and non-parametric tests:

Parametric tests assume that the data is **normally distributed**.

Non-parametric tests do not rely on any normal distribution of data **(data NOT normally distributed)**.

So, how do I know if my data is normally distributed or not?

This can be determined by:

1. Application of the Central Limit Theorem
2. Visual inspection and statistical tests

1. The Central Limit Theorem (CLT)

- The CLT states that if you have a population with a mean and standard deviation and take sufficiently large samples from the population, the distribution of the sample means will be approximately normally distributed.
- This will hold true when the population is initially normal distributed or skewed, providing that the sample size is 30 or more ($n \geq 30$).
- Therefore, we can assume our data is normally distributed if you have 30 or more cases/participants/respondents etc.

Note: *If you have multiple groups (for example, splitting the sample into male and female), data for* ***EACH group*** *must fit these criteria.*

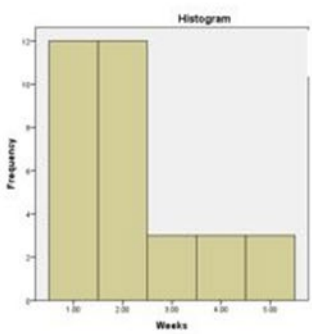
2. Visual inspection and statistical tests can be viewed together as follows:

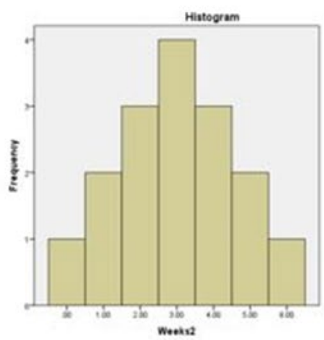
- 2A The histogram of your data represents an approximately symmetrical distribution *(this is a visual inspection only and cannot infer normal distribution alone)*
- 2B We can use skewness value to compare to 0 and to describe our data *(this is for descriptive purposes only and cannot infer normal distribution alone)*
    - A value of 0 = perfect normal distribution
    - A value of >0 = positive skew
    - A value of <0 = negative skew
- 2C We can look at the Shapiro-Wilk test of normality *(this is our value of interest when determining our data distribution)*
    - For normal distributions, the significance value (Sig) will be **> 0.05 (non-significant).**
    - For non-normally distributed data, the significance value (Sig) will be **< 0.05 (significant).**

Therefore, data is considered to follow a normal distribution (and therefore be parametric) if the following applies:

- The histogram looks symmetrical AND the statistical tests are non-significant.

For example…





Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation	Skewness	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error
Weeks	33	1.00	5.00	2.1818	1.28585	1.047	.409
Valid N (listwise)	33						

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation	Skewness	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error
Weeks2	16	.00	6.00	3.0000	1.63299	.000	.564
Valid N (listwise)	16						

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Weeks	.283	33	.000	.807	33	.000

a. Lilliefors Significance Correction

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Weeks2	.125	16	.200*	.972	16	.872

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

> .05 = normally distributed

Paired or unpaired data

When numerical variables are involved, you also need to determine if you have ***paired or unpaired data.***

So, how do we determine if data is paired or unpaired?

- **Unpaired**
  - Two different groups of individuals or items
    - **An intervention group and a placebo group**
    - **Standard medication vs. trial medication**
    - **Case vs. control**

- **Paired**
  - The same individual or items are measured twice (or more often)
  - More than one measurement per individual or item

  For example…
    - **Measurement before and after an intervention**
    - **Heart rate measured before an exam and then during an exam**

Common Statistical Tests

1 numerical, 1 categorical

- **Independent samples t-test**
- **Paired samples t-test**
- **Paired Wilcoxon's test**
- **One Way Analysis of variance (ANOVA)**
- **Mann-Whitney U test**
- **Kruskal-Wallis test**

2 categorical

- **Chi-square test of independence**
- **McNemar's test**
- **Chi-square test for trend**

2 numerical

- **Correlation**
- **Regression**

The tests underlined in the above image will be the focus for learning activities in this workbook.

Each parametric test has a non-parametric equivalent. Refer to the last page of this workbook to see a "Decision Tree" which can be used to determine the appropriate test.

Box 1: Deciding which test to use

Table 1 displays a range of common parametric and non-parametric statistical tests used by health researchers.

Note: This table does not present every statistical test available, but instead provides a range of common statistical tests used by health researchers.

Common bivariate statistical tests used by health researchers				
Name	**Statistic**	**Purpose**	**Measurement Level**	
			Independent variable	**Dependent variable**
Parametric Tests				
t-test for independent groups	*t*	To test the different between two independent group means	Nominal	Interval, ratio
t-test for dependent groups	*t*	To test the different between two dependent group means	Nominal	Interval, ratio
Analysis of variance (ANOVA)	*F*	To test the difference among the means of three or more independent groups, or more than one independent variable	Nominal	Interval, ratio
Repeated measures of ANOVA	*F*	To test the difference among means of three or more related groups or sets of scores	Nominal	Interval, ratio
Pearson's r	*r*	To test the existence of a relationship between two variables	Interval, ratio	Interval, ratio
Non-parametric Tests				
Chi-squared test	X^2	To test the difference in proportions in two or more independent groups	Nominal	Nominal
Mann-Whitney U-test	*U*	To test the difference in ranks of scores on two independent groups	Nominal	Ordinal
Kriskal-Wallis test	*H*	To test the difference in ranks of scores of three or more independent groups	Nominal	Ordinal
Wilcoxon signed ranks test	*T (Z)*	To test the difference in ranks of scores of two related groups	Nominal	Ordinal
Spearman's rank order correlation	r_s	To test the existence of a relationship between two variables	Ordinal	Ordinal

Table 1: Common bivariate statistical tests used in healthcare research

(Adapted from Polit, Becke and Hungler, 2001 p.359)

Statistical Significance

When conducting statistical tests, often researchers are interested in determining if there are statistically significant findings. To determine the significance of findings, it is important that you understand p-values. The content in Box 2 below provides information to assist you to understand how to interpret p-values.

Understanding *p-values*

Definition
The *p*-value or probability value is used in statistical hypothesis testing to find the probability that the null hypothesis is true.

The significance level (or alpha level)
Most commonly, researchers set a significance level (or alpha level) at 0.05. The smaller the *p*-value, the higher the significance.

For example:

- If the *p*-value is below 0.05, this indicates a statistically significant finding. Eg p=0.001 or p=0.03
- If the p-value is more than 0.05, this indicates a non-significant finding. Eg p=0.56 or p=0.73

To assist understanding this further:
Click the link below to watch a You Tube clip to help consolidate your understanding of *p*-values:

P-Value explained – What is a *p*-value

Box 2: Understanding p-values

Important points to remember about reporting statistics:

Reporting Mean and Standard Deviation	The mean (M) and standard deviation (SD) need to be reported every time you mention a variable.
General rule for reporting Mean and Standard Deviation	When reporting the mean and standard deviation, you must ***use italics*** for *M* and *SD*.
Correlation coefficient – determining strength of a correlation	Between -1.00 and +1.00. General rule of thumb: 0 - 0.3 = weak relationship 0.4 - 0.7 = moderate relationship 0.8 - 1 = strong relationship
Statistical significance	Must be reported. This indicates whether the result is statistically significant. a *p*-value $\leq$ 0.05 = significant a *p*-value > 0.05 = non-significant

Section 1: One numeric and one categorical variable

1. Unpaired t-test or Independent t-test (for equal or unequal variances)

Purpose of t-test:
To test for a statistically significant difference between two independent sample means.

When should the t-test be used?
This test should be used for normally distributed data with an approximately equal amount of variability in each set of scores. The aim is to test difference between two separate (or independent) groups.

Sample Research Questions to use this statistical test:

- Are people aged 65+ with pets happier than those without pets?
- Do internal students achieve higher grades to external students?
- Is there a difference between the salaries of male academics and female academics?

Using SPSS to run an Independent t-test

The research question (example 1):
Is there a significant difference in fear of statistics scores between maths skills and confidence building groups before the interventions?

To find your answer to this research question, you will need to follow these steps:

STEP 1: We need to identify the variables and then classify the variables.
Variable 1: Type of class = categorical, two categories (maths skills and confidence building)
Variable 2: Fear of stats test scores = numerical

STEP 2: Decide whether the numerical variable is parametric or non-parametric.
To do this we need to test the criteria for normality for each level of the categorical variable (training type) by following these steps:

To obtain skewness values and examine histograms:
Click **Data → Split File**
Select **Compare groups**
Select and move ⇨ 'type of class' into the '**Groups Based on'** list.
Click **OK**

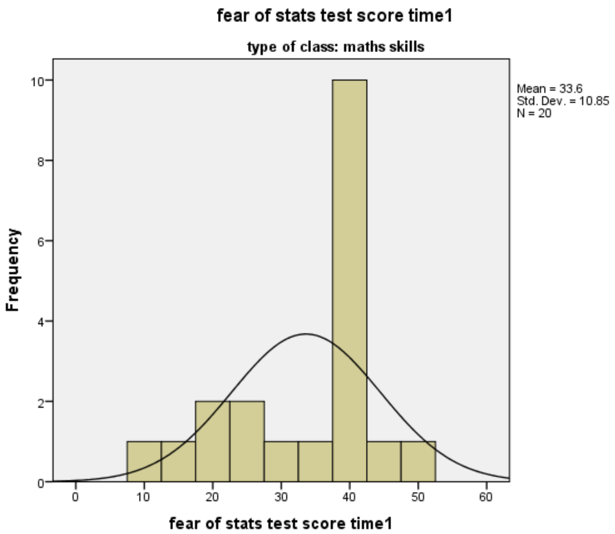
Click **Analyze → Descriptive statistics → Frequencies**
Select and move ⇨ 'fear of stats test score time1' into the **Variable(s) List**
Under **Statistics,** tick the **Skewness** box, Click **Continue**
Under **Charts** select **Histogram** and tick '**Show normal curve on histogram**' → Click **Continue**
Untick **Display frequency tables**
Click **OK,** then view output in the output window

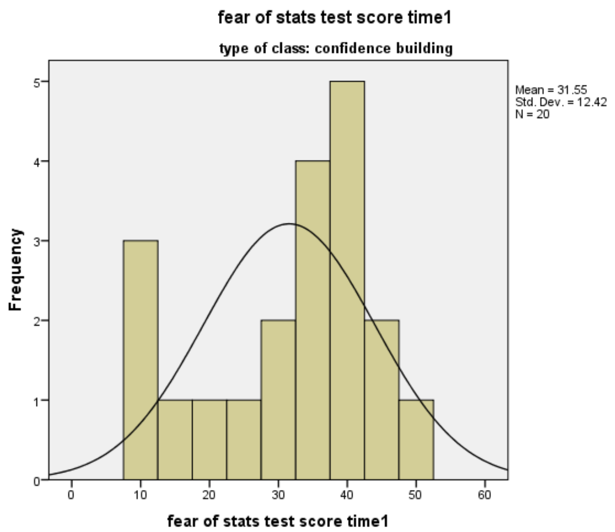
The below histograms are difficult to interpret, so it is useful to look at the skewness values shown in the statistics output. The skewness values indicate a negative skew for maths skills (-.722) and a negative skew for confidence building (-.587). In other words, if the skewness value is negative then you have a negative skew. If the value is positive then you have a positive skew. A perfect, normal distribution will have a skewness value of '0', however this is very unlikely and you will generally always have a very mild skew of some sort.

Statistics

fear of stats test score time1

maths skills	N	Valid	20
		Missing	0
	Skewness		-.722
	Std. Error of Skewness		.512
confidence building	N	Valid	20
		Missing	0
	Skewness		-.587
	Std. Error of Skewness		.512





Make sure 'Split File' is turned off (or RESET) prior to this next step!

To turn off split file:

Click **Data → Split File**
Click **Reset**
Click **OK**

To run statistical tests for normality:

Click **Analyze → Descriptive statistics → Explore**
Select and move ⇨ 'type of class' into the **Factor List**
Select and move 'fear of stats test score time1' into the **Dependent List**
Under **Plots** select **Normality test with plots**
Click **Continue** and **OK,** then view output in the output window
Our normality tests produce non-significant results for our Shapiro-Wilk test (>.05). This is our test of interest as there are less than 30 people in our group. This means the data can be considered as normally distributed.

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	type of class	Statistic	df	Sig.	Statistic	df	Sig.
fear of stats test score time1	maths skills	.257	20	.001	.909	20	.061
	confidence building	.214	20	.017	.911	20	.067

a. Lilliefors Significance Correction

What do these results mean?
As the criteria for normality are met in both categories, we conclude that the data has a normal distribution and choose parametric tests. As there are two groups of the categorical variable, we use the t-test. These are independent (same measure on different people), so we will use an independent t-test (unpaired).

STEP 3: State the H_0 and H_A.

The question asks us whether there is a significant difference in Fear of Statistics test score (time 1) between the two class types before interventions. This implies a two-tailed test.

H_0 : $\mu_{\text{maths skills}} = \mu_{\text{confidence building}}$
(i.e., the mean of the maths skills class is equal to the mean of the confidence building class training)

H_A : $\mu_{\text{mathss kills}} \neq \mu_{\text{confidence building}}$
(i.e., the mean of the maths skills class is not equal to the mean of the confidence building class)

Set the alpha level (α = 0.05) → *You do not actually do anything here – this is just a decision you need to make and write down.*

STEP 4: Calculate the test statistics.

- Click **Analyze → Compare means → Independent – samples T Test**
- Select 'fear of stats test score time1' as the **Test Variable**
- Select 'type of class' as the **Grouping Variable**
- Click **Define Groups**
- Type '1' for Group 1 (maths skills) and type '2' for Group 2 (confidence building)
- Click **Continue** and **OK**

STEP 5: Determine variability

An assumption for the unpaired t-test is that the variances of the numerical variable in the two categories of the categorical variable should not be statistically different. If they are different, corrections apply. Levene's test for equality of variances (F-test) is used, and if the F-test is significant, then the T test for unequal variances is the correct statistical procedure.

SPSS provides output for Equal variance assumed and equal variance not assumed, but you need to read the F-test result to determine which is the relevant output.

The output provided by SPSS is as follows:

Group Statistics

	type of class	N	Mean	Std. Deviation	Std. Error Mean
fear of stats test score time1	maths skills	20	33.60	10.850	2.426
	confidence building	20	31.55	12.420	2.777

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference Lower	Upper
fear of stats test score time1	Equal variances assumed	.135	.716	.556	38	.582	2.050	3.688	-5.415	9.515
	Equal variances not assumed			.556	37.327	.582	2.050	3.688	-5.420	9.520

What do these results mean?

If the p-value of the Levene's test (see column F) is < 0.05, it indicates a significant difference in variances. Hence, we need to refer to the bottom row of the table (equal variances are NOT assumed). However, if the variances of the two groups are equal, we need to read the results from the top row.

The output from this analysis shows us that the variances are equal (the Levene's test is not significant i.e., > 0.05). Therefore we need to refer to the top row of the table (equal variances assumed).

<u>STEP 6:</u> Decide if we reject/retain H_0 and draw conclusions.

Reading the top row of the table, the output shows that the *p-value* is > .05 (in the Sig (2-tailed) column). Because the p-value is .582, we can retain the H_0. That is, the mean Fear of Statistics test score (time1) in the maths skills class is equal to the mean Fear of Statistics test score (time1) in the confidence building class. The mean Fear of Statistics test score (time1) in the maths skills class is not significantly different from the mean Fear of Statistics test score (time1) in the confidence building group.

<u>To answer the research question:</u>

An independent samples t-test was conducted to compare the first Fear of Statistics test scores between those in enrolled in a maths skills class and a confidence building class.

There was no significant difference in the scores for the maths skills class (*M*=33.60, *SD*=10.85) and the confidence building class (*M*=31.55, *SD*=12.42); *t*(38)=.556, *p*=.582). The results suggest that there was no difference between each class and students' fear of statistics. This also supports that the sample was randomly chosen for each intervention type.

<u>TEMPLATE used for writing your results:</u>

There was no/a significant difference in the scores for the (GROUP 1) (*M*=mean of group 1, *SD*=standard deviation group 1) and GROUP 2(*M*=mean of group 2, *SD*=standard deviation group 2); t(df)=t-statistic, *p*= sig.2-tailed value). The results suggest that…(explain here what the results tell us).

The research question (example 2):
Is there a significant difference in confidence scores before the intervention between those who have a pass grade and those who have a fail grade?

To find your answer to this research question, you will need to follow these steps:

STEP 1: We need to identify the variables and then classify the variables.
Variable 1: Type of grade = categorical, two categories (pass grade and fail grade)
Variable 2: Confidence scale score time1 (confid1) = numerical

STEP 2: Decide whether the numerical variable is parametric or non-parametric.
To do this we need to test the criteria for normality for each level of the categorical variable (grade) by following these steps:

To obtain skewness values and examine histograms:
Click **Data → Split File**
Select **Compare groups**
Select and move ⇨ 'grade' into the '**Groups Based on'** list**.**
Click **OK**

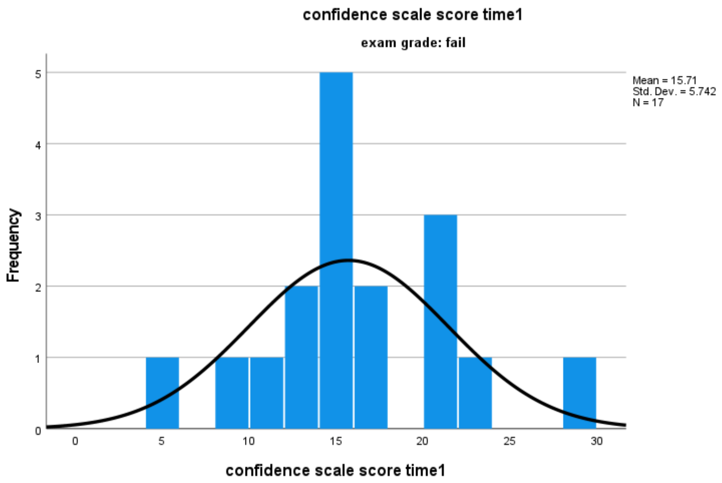
Click **Analyze → Descriptive statistics → Frequencies**
Select and move ⇨ 'confidence scale score time1 into the **Variable(s) List**
Under **Statistics,** tick the **Skewness** box, Click **Continue**
Under **Charts** select **Histogram** and tick '**Show normal curve on histogram**' → Click **Continue**
Untick **Display frequency tables**
Click **OK,** then view output in the output window

The below histograms are difficult to interpret, so it is useful to look at the skewness values shown in the statistics output. The skewness values indicate a negative skew for those with a pass grade (-.866) and a positive skew for those with a fail grade (.437). In other words, if the skewness value is negative then you have a negative skew. If the value is positive then you have a positive skew. A perfect, normal distribution will have a skewness value of '0', however this is very unlikely and you will generally always have a very mild skew of some sort.

Statistics

confidence scale score time1

pass	N	Valid	23
		Missing	0
	Skewness		-.866
	Std. Error of Skewness		.481
fail	N	Valid	17
		Missing	0
	Skewness		.437
	Std. Error of Skewness		.550



Make sure 'Split File' is turned off (or RESET) prior to this next step!

To turn off split file:

Click **Data → Split File**
Click **Reset**
Click **OK**

To run statistical tests for normality:

Click **Analyze → Descriptive statistics → Explore**
Select and move ⇨ 'grade' into the **Factor List**
Select and move 'confidence scale score time1' into the **Dependent List**
Under **Plots** select **Normality test with plots**
Click **Continue** and **OK,** then view output in the output window

Our normality tests produce non-significant results for our Shapiro-Wilk test (>.05). This is our test of interest as there are less than 30 people in our group. This means the data can be considered as normally distributed.

Tests of Normality							
		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	exam grade	Statistic	df	Sig.	Statistic	df	Sig.
confidence scale score time1	pass	.185	23	.039	.916	23	.054
	fail	.185	17	.123	.956	17	.555
a. Lilliefors Significance Correction							

What do these results mean?
As the criteria for normality are met in both categories, we conclude that the data has a normal distribution and choose parametric tests. As there are two groups of the categorical variable, we use the t-test. These are independent (same measure on different people), so we will use an independent t-test (unpaired).

STEP 3: State the H_0 and H_A.

The question asks us whether there is a significant difference in confidence scale scores (time 1) between the pass and fail grade groups. This implies a two-tailed test.

H_0 : $\mu_{\text{pass group}} = \mu_{\text{fail group}}$
(i.e., the mean of the pass group is equal to the mean of the fail group)

H_A : $\mu_{\text{pass group}} \neq \mu_{\text{fail group}}$
(i.e., the mean of the pass group is not equal to the mean of the fail group)

Set the alpha level ($\alpha = 0.05$) → *You do not actually do anything here – this is just a decision you need to make and write down.*

STEP 4: Calculate the test statistics.

- Click **Analyze → Compare means → Independent – samples T Test**
- Select 'confidence scale score time1' as the **Test Variable**
- Select 'grade' as the **Grouping Variable**
- Click **Define Groups**
- Type '1' for Group 1 (pass) and type '2' for Group 2 (fail)
- Click **Continue** and **OK**

STEP 5: Determine variability

An assumption for the unpaired t-test is that the variances of the numerical variable in the two categories of the categorical variable should not be statistically different. If they are different, corrections apply. Levene's test for equality of variances (F-test) is used, and if the F-test is significant, then the T test for unequal variances is the correct statistical procedure.

SPSS provides output for Equal variance assumed and equal variance not assumed, but you need to read the F-test result to determine which is the relevant output.

The output provided by SPSS is as follows:

Group Statistics					
	exam grade	N	Mean	Std. Deviation	Std. Error Mean
confidence scale score time1	pass	23	20.74	5.479	1.142
	fail	17	15.71	5.742	1.393

Independent Samples Test										
		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference Lower	Upper
confidence scale score time1	Equal variances assumed	.147	.703	2.814	38	.008	5.033	1.788	1.413	8.654
	Equal variances not assumed			2.794	33.686	.009	5.033	1.801	1.371	8.695

What do these results mean?
If the p-value of the Levene's test (see column F) is < 0.05, it indicates a significant difference in variances. Hence, we need to refer to the bottom row of the table (equal variances are NOT assumed). However, if the variances of the two groups are equal, we need to read the results from the top row.

The output from this analysis shows us that the variances are equal (the Levene's test is not significant i.e., > 0.05). Therefore, we need to refer to the top row of the table (equal variances assumed).

<u>STEP 6:</u> Decide if we reject/retain H_0 and draw conclusions.

Reading the top row of the table, the output shows that the *p-value* is <.05 (in the Sig (2-tailed) column). Because the p-value is .008, we can reject the H_0. That is, the mean of the pass group confidence score (time1) is not equal to the mean of the fail group confidence score (time1). The mean confidence score (time1) for the pass group is significantly different from the mean confidence score (time1) in the fail group.

<u>To answer the research question:</u>

An independent samples t-test was conducted to compare the first confidence scale scores between those with a Pass grade and those with a Fail grade.

There was a statistically significant difference in the scores for those with a pass grade (*M*=20.74, *SD*=5.48) and those with a fail grade (*M*=15.71, *SD*=5.74); *t*(38)= 2.814, *p= .008*). The results suggest that those with a pass grade have more confidence before the intervention compared to those with a fail grade.

TEMPLATE used for writing your results:

There was no/a significant difference in the scores for the (GROUP 1) (*M*=mean of group 1, *SD*=standard deviation group 1) and GROUP 2(*M*=mean of group 2, *SD*=standard deviation group 2); t(df)=t-statistic, *p*= sig.2-tailed value). The results suggest that…(explain here what the results tell us).

2. Paired t-test

Purpose of Paired t-test:
To test for significantly significant difference between two related sample means.

When should this test be used?
This test is often used to test difference before and after an intervention. To use this test, the sample should be comprised of data using the same group of individuals who have provided data on two separate occasions. Alternatively, this test can be used to compare individuals in one sample with specifically linked individuals in a second sample.

Sample Research Questions to use this statistical test:

- Do people report increased energy and wellbeing levels after an 8-week targeted exercise program?
- Is there a difference between perceived coping skills of mothers and fathers after the birth of their first born child.

Using SPSS to run an Independent t-test

The research question: Do the *Confidence in coping with Statistics* scores improve after intervention?

To find your answer to this research question, you will need to follow these steps:

STEP 1: Classify the variables.
Variable 1: Categorical, grouped by time - before intervention [time 1] and after intervention [time 2]
Variable 2: Numerical – Confidence in coping with Statistics Scale

STEP 2: To decide whether the distribution of the numerical variable is parametric or non-parametric, we need to test the assumptions for normality.

This is a paired (or repeated measures) study, so we need to consider whether the distribution of DIFFERENCES between the two measures is normal. To do this, we need to compute a new variable, called 'condiff', which means the difference between time 1 and time 2 confidence scores.

- Click **Transform→ Compute Variable**
- Type 'condiff' in **Target Variable** box
- Highlight and move variables to form the **Numeric Expression** 'confid2 – confid1' (confidence scale score time2 *minus* confidence scale score time1)
- Click **OK** and you will see the new variable called 'condiff' in the Variable View Window

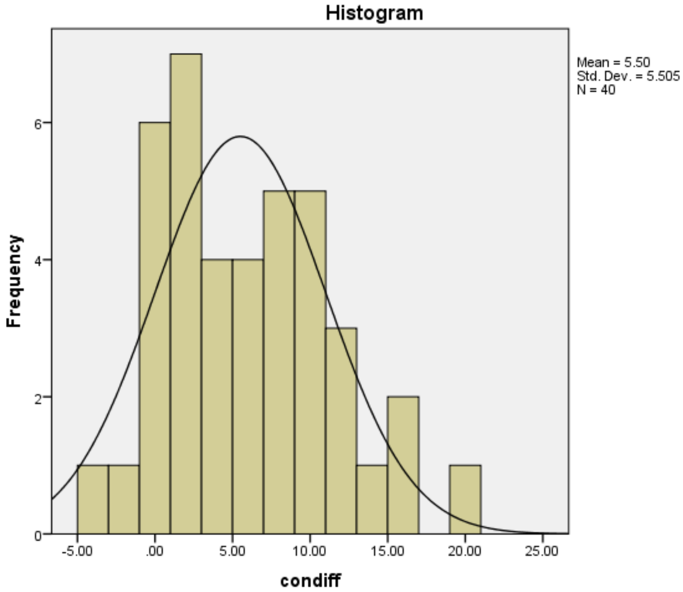
To obtain skewness values and examine histograms:

Click **Analyze → Descriptive statistics → Frequencies**
Select and move ⇨ 'condiff' into the **Variable(s) List**
Untick **Display frequency tables**
Under **Statistics,** tick the **Skewness** box, Click **Continue**
Under **Charts** select **Histogram** and tick '**Show normal curve on histogram**'
Click **Continue** and **OK,** then view output in the output window

Statistics

condiff

N	Valid	40
	Missing	0
Skewness		.575
Std. Error of Skewness		.374



We can now consider whether the distribution of the variable 'difference' is normal
Look at the histogram and check skew. The histogram appears to be approximately normally distributed. The skewness value indicates a positive skew.

To run statistical tests for normality:

Click **Analyze → Descriptive statistics → Explore**
Select and move 'condiff' into the **Dependent List**
Under **Plots** select **Normality test with plots**
Click **Continue** and **OK,** then view output in the output window

As we have 30 people in our sample, we can use the Shapiro-Wilk test for normality. We can see by looking at the output that this test is statistically non-significant because the Shapiro-Wilk *p-value* is .173 (not <0.05), so we assume our difference variable is normally distributed.

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
condiff	.118	40	.168	.960	40	.173

a. Lilliefors Significance Correction

STEP 3: Before we conduct the test, we need to state the H_0 and H_A.

The question asks us whether there is an improvement in the Confidence in coping with Statistics Scale scores after intervention. This implies a one-tailed test.

H_0: $\mu_{time1} = \mu_{time2}$ (i.e., mean of score before intervention is equal to the mean of score after intervention)
(Alternatively, $\mu_{difference} = 0$)

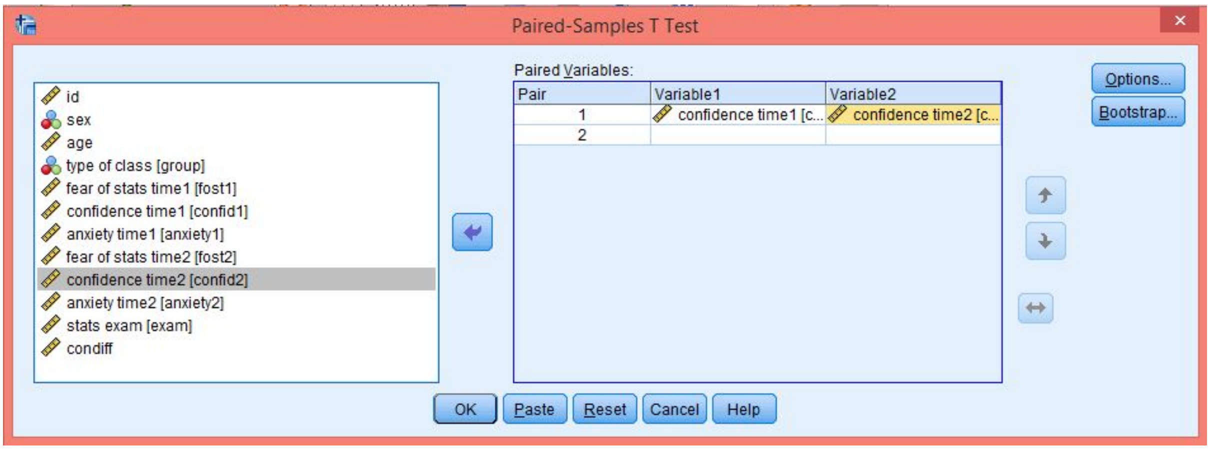
H_a: $\mu_{time1} < \mu_{time2}$
(Alternatively $\mu_{difference} > 0$)
Set the alpha level ($\alpha = 0.05$).

Because this is a one-tailed test, the p-value that is produced in the output (sig 2-tailed) will need to be divided by 2, to get the p-value required for the 1-tailed test.

STEP 4: Calculate the test statistics.

Click **Analyze → Compare means → Paired samples t-test**
Select 'confid1' and ⇨ into box **Variable 1**
Select 'confid2' and ⇨ into box **Variable 2**
Click **OK**

Note: Make sure both variables are entered into the same row (Pair 1) as in the screenshot.



The output provided by SPSS is as follows:

T-Test

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	confidence scale score time1	18.60	40	6.067	.959
	confidence scale score time2	24.10	40	5.674	.897

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	confidence scale score time1 & confidence scale score time2	40	.562	.000

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	confidence scale score time1 - confidence scale score time2	-5.500	5.505	.870	-7.261	-3.739	-6.319	39	.000

<u>STEP 5:</u> Decide if we reject/retain the null hypothesis and draw a conclusion.

The output shows that the *p-value* is .000 (in column headed 'Sig (2-tailed). But this is the *p-value* for a 2-tailed test and we have stated a 1-tailed test. So the true *p-value* is .000/2 = 0.000.

This is less than 0.05, so we can accept the H_A and reject the H_0. The conclusion is that the Confidence in coping with statistics scores improve significantly after intervention.

REMEMBER: here we looked at the data set as a whole (i.e. both math skills and confidence building interventions). If we wanted to look at each one separately we would have to split the file by class type.

To answer the question:

A paired samples t-test was conducted to compare confidence in coping with statistics scores of students pre- and post-intervention.

There was a significant difference in the scores between the pre-intervention test (*M*= 18.60, *SD*= 6.07) and post intervention test (*M*= 24.10, *SD*= 5.67); t(39)=-6.32, *p*<.01. This result may demonstrate that the intervention appears to increase students' confidence with statistics. *(Remember we did time2-time1 to get the difference score. Here, because it is measuring confidence, we would expect time2's scores to be higher than time1's. Our t-statistic is negative therefore showing it increased confidence like expected)*

TEMPLATE used for writing your results:

There was no/a significant difference in the scores for the (GROUP 1) (*M*=mean of group 1, *SD*=standard deviation group 1) and GROUP 2(*M*=mean of group 2, *SD*=standard deviation group 2); t(df)=t-statistic, *p*= sig.2-tailed value). The results suggest that…*(explain what the results tell us).*

Section 2: Two categorical variables

2. Chi-square (χ^2) test of independence
(Also called Chi-square (χ^2) test of Contingencies)

Purpose of Chi-square test of independence:
To test whether two categorical (nominal) variables are related.

When should this test be used?
This test is often to test whether group or category membership on one variable is influenced by group membership on a second variable. To use this statistical test, each participant should participate only once in the research. As a general rule, expected frequencies should be at least five (5).

Sample Research Questions to use this statistical test:
- Do women under the care of a breast care nurse have lower unmet needs to women without a breast care nurse?
- Are people classified as obese more likely than people who are not classified obese, to fall ill at least once during winter?

Using SPSS to run a Chi-square test of independence

The area of interest: The subject coordinator is interested in whether gender influences grade of the final exam.

The research question: Is there a significant association between gender and pass rate of the final exam?

The frequency data is as follows:

Grade	Gender	
	Male	Female
Pass	10	13
Fail	9	8

To find your answer to this research question, you will need to follow these steps:

STEP 1: Classifying the variables.
Variable 1: sex - categorical (2 levels – Male and Female)
Variable 2: grade – categorical (2 levels – pass and fail)

STEP 2: Before we can use the chi-square test for independence, we need to check the assumptions, which are:

- Sample data are randomly selected, and are represented as frequency counts in the contingency table
- Maximum of 25% of "expected frequencies" below 5 (or expected cell count <5 in 0 cells).
- No "expected frequency" below 1

We assume data are randomly selected, so we need to check the second assumption.
The easiest way to do this is to go ahead and run the chi-square test.

How to perform Chi-square tests using SPSS (if the data is provided in tabular form):

- Select **Analyze** → **Descriptive Statistics** → **Crosstabs**
- Select the row variable 'sex' and ⇨ to move into the **Row(s)** box
- Select the column variable 'exam grade' and ⇨ to move into the **Column(s)** box
    - Note: It does not matter which variable you chose to be in the **Row** or **Column** box – this will not influence your results.
- Click on **Statistics** and tick **Chi-square**. Click **Continue.**
- Click on **Cells** to open the **Cell Display**
- In the **Counts** box, tick **Observed** and **Expected**
- In the **Percentages** box, click **Row**, **Column** and **Total**.
- Click on **Continue**. Click **OK**

sex * exam grade Crosstabulation

			exam grade		Total
			pass	fail	
sex	male	Count	10	9	19
		Expected Count	10.9	8.1	19.0
		% within sex	52.6%	47.4%	100.0%
		% within exam grade	43.5%	52.9%	47.5%
		% of Total	25.0%	22.5%	47.5%
	female	Count	13	8	21
		Expected Count	12.1	8.9	21.0
		% within sex	61.9%	38.1%	100.0%
		% within exam grade	56.5%	47.1%	52.5%
		% of Total	32.5%	20.0%	52.5%
Total		Count	23	17	40
		Expected Count	23.0	17.0	40.0
		% within sex	57.5%	42.5%	100.0%
		% within exam grade	100.0%	100.0%	100.0%
		% of Total	57.5%	42.5%	100.0%

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	.351[a]	1	.554		
Continuity Correction[b]	.074	1	.785		
Likelihood Ratio	.351	1	.553		
Fisher's Exact Test				.750	.393
Linear-by-Linear Association	.342	1	.559		
N of Valid Cases	40				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 8.08.

b. Computed only for a 2x2 table

When we look at the data, we see the note under the bottom table which says "0 cells have expected count less than 5". In addition, we can see from the 2nd table that the expected cell count for all cells is >5. This tells us that our assumptions are not violated, and we can proceed with the chi-square test for independence.

STEP 3: Before we conduct the test, we need to state the H_0 and H_A.

H_0: Proportion of individuals with a pass grade in the exam is independent from gender
H_A: Proportion of individuals with a pass grade in the exam is not independent from gender

Set alpha level, $\alpha = 0.05$

STEP 4: Conduct the statistical test – which we have already done in Step 2!

STEP 5: Decide if we reject / retain H_0 and draw conclusions

.
The output shows that the p-value for the chi-square test for independence is .554 (we look at the row for Pearson's chi-square). This is >0.05, so we accept the H_0. That is, there is no difference between the number of individuals with passing grades on the exam, based on gender.

To answer this question:

A chi-square test of independence was calculated which compared the frequency of pass and fail grades in men and women.

No significant interaction was found (χ^2 (1, *N*=40) = .351, *p*=.554).

TEMPLATE used for writing your results:

No/A significant interaction was found (χ^2 (df, *N*= number of participants in the total sample) = Pearson chi-square value, *p*=sig.value). **IF interaction was found, report percentages belonging to each group → this is found in the crosstabs table**

4. Chi-square test for trend

Purpose of Chi-square test for trend:
To assess whether the association between variables follows a trend.

When should this test be used?
The chi-square test for trend is used where 1 variable is binary and the other is ordered categorical. *Reminder:* binary variables are variables which only take two values. For example, male or female or in the example below, maths skills and confidence building.

Sample research questions to use this statistical test:
- Is the frequency of using oral mouth rinse associated with the presence or absence of gingivitis in orthodontic patients?

Using SPSS to run a Chi-square test for trend

The area of interest: The subject coordinator wanted to know whether different interventions results in the effort of students put into their final exam.

The research question: Is there a significant difference between two interventions to lessen the anxiety of statistics classes and the considered 'effort' students put into their exam?

		Effort		
		Poor	**Average**	**Excellent**
Intervention	**Maths skills**	6	7	7
	Confidence building	6	5	9

To find your answer to this research question, you will need to follow these steps:

STEP 1: Classifying the variables.
Variable 1: type of class - categorical (2 levels – maths skills and confidence building)
Variable 2: effort – categorical (3 levels – poor, average, excellent)

STEP 2: Before we can use the chi-square test for trend, we need to check the assumptions, which are:

- Sample data are randomly selected, and are represented as frequency counts in the contingency table
- Maximum of 25% of "expected frequencies" below 5 (or expected cell count <5 in 0 cells).
- No "expected frequency" below 1

We assume data are randomly selected, so we need to check the second assumption. The easiest way to do this is to go ahead and run the chi-square test.

How to perform Chi-square tests using SPSS (if the data is provided in tabular form)

- Select **Analyze → Descriptive Statistics → Crosstabs**
- Select the row variable i.e. 'type of class' and ⇨ to move into the **Row(s)** box
- Select the column variable i.e. ' effort in class' and ⇨ to move into the **Column(s)** box
- Click on **Statistics** and click **Chi-square** in the box. Click **Continue.**
- Click on **Cells** to open the **Cell Display** box
- In the **Counts** box, tick **Observed** and **Expected**
- In the **Percentages** box, click **Row**, **Column** and **Total**.
- Click on **Continue**. Click **OK**

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
type of class * effort in class	40	100.0%	0	0.0%	40	100.0%

type of class * effort in class Crosstabulation

			effort in class			Total
			poor	average	excellent	
type of class	maths skills	Count	6	7	7	20
		Expected Count	6.0	6.0	8.0	20.0
		% within type of class	30.0%	35.0%	35.0%	100.0%
		% within effort in class	50.0%	58.3%	43.8%	50.0%
		% of Total	15.0%	17.5%	17.5%	50.0%
	confidence building	Count	6	5	9	20
		Expected Count	6.0	6.0	8.0	20.0
		% within type of class	30.0%	25.0%	45.0%	100.0%
		% within effort in class	50.0%	41.7%	56.3%	50.0%
		% of Total	15.0%	12.5%	22.5%	50.0%
Total		Count	12	12	16	40
		Expected Count	12.0	12.0	16.0	40.0
		% within type of class	30.0%	30.0%	40.0%	100.0%
		% within effort in class	100.0%	100.0%	100.0%	100.0%
		% of Total	30.0%	30.0%	40.0%	100.0%

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	.583[a]	2	.747
Likelihood Ratio	.586	2	.746
Linear-by-Linear Association	.141	1	.707
N of Valid Cases	40		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 6.00.

When we look at the data, we see the note under the bottom table which says "0 cells have expected count less than 5". In addition, we can see from the 2nd table that the expected cell count for all cells is >5. This tells us that our assumption is not violated, and we can proceed with the chi-square test for trend.

STEP 4: Before we conduct the test, we need to state the H_0 and H_A.

The question asks us whether there is a significant difference in the class type and perceived effort. This implies a two-tailed test.

H_0: Proportion of individual's perceived effort is independent from intervention
H_A: Proportion of individual's perceived effort is not independent from intervention

Set alpha level, α = 0.05

STEP 5: Conduct the statistical test.
We already performed the chi-square test when we were checking the assumptions. We can now refer to the output from that step (see above).

STEP 6: Decide if we reject / retain H_0 and draw conclusions.
The output shows that the p-value for the chi-square test for trend is 0.707 (we look at the row for Linear-by-Linear Association, we use this row because variables are ordinal). This is > 0.05, so we accept the H_0, and reject H_A. That is, there is no difference between class types and individuals effort scores.

To answer this question:
A chi-square test of independence was calculated which compared the frequency of pass and fail grades in men and women.

No significant interaction was found (χ^2 (1, *N*=40) = .141, *p*=.707).

TEMPLATE:
No/A significant interaction was found (χ^2 (df, *N*= number of participants in the total sample) = Linear-by-linear Association value, *p*=sig.value).

IF an interaction was found, report percentages belonging to each group → this is found in the crosstabs table

For example:

For the maths skills class, 30% of the sample demonstrated poor effort, 35% average effort and 35% excellent effort. For the confidence building class, 30% of the sample demonstrated poor effort (the same as the maths class), 25% average effort and 45% excellent effort, which was higher than the maths class.

Section 3: Two numerical variables

4. Bivariate Correlation

Purpose of Pearson's correlation coefficient:
Pearson's correlation coefficient (or Pearson's *r*) is used to determine the strength and direction of the linear association/relationship between two continuous variables.

When should this test be used?
In this workbook, we will focus on using a bivariate correlation which is used to measure the linear association between two continuous variables.

Note: A correlation coefficient is between -1.00 and +1.00. When reporting a correlation, we usually describe the relationship between variables as "weak", "moderate" or "strong". This is determined by the correlation coefficient.

Description	Correlation coefficient (+ or -)
Weak	0 – 0.3
Moderate	0.4 – 0.7
Strong	0.8 – 1.0

**these values are just "guides" for the description of the relationship between variables. These values can be either positive or negative.*

Sample Research Questions to use this statistical test:
1. Is there a relationship between socio-economic status and level of education?
2. Is there a relationship between alcohol consumption and weight?
3. Is there a relationship between depression and suicide rates?

Using SPSS to run a Correlation

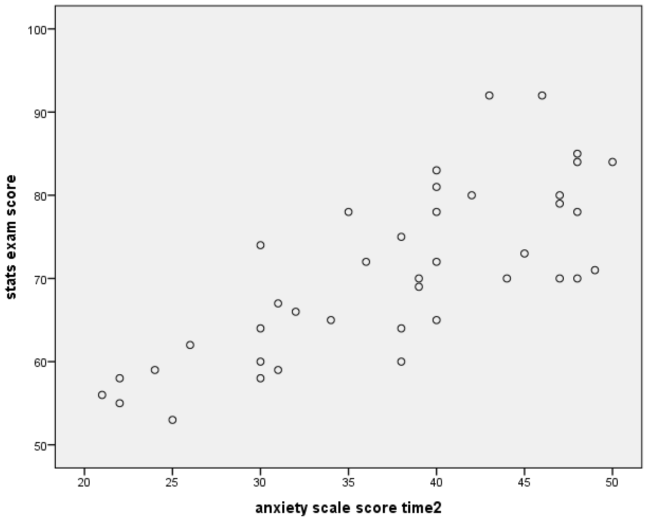
The research question: The subject coordinator wanted to determine if there was a relationship between the anxiety of students at the end of the intervention and their exam score.

1. Construct a scatterplot examining the relationship between Anxiety Scale score (time 2) and the exam score.
2. Obtain the correlation coefficient using SPSS.

To find your answer to this research question, you will need to follow these steps:

STEP 1: Classify the variables:
Variable 1: numerical (anxiety2)
Variable 2: numerical (exam)

STEP 2: Construct a scatter plot to determine the direction of linear association:
Click **Graphs** → **Legacy Dialogs** → **Scatter/Dot** → **Simple Scatter** → **Define**
Select the 'anxiety2' variable and ⇨ into the X-Axis box
Select the 'stats exam score' variable and ⇨ into the Y Axis box
Click **OK** and view the output in the output window



The output suggests that there is a positive linear association between Anxiety Scale Score (time 2) and Stats exam score.

STEP3: We now want to find how strong the linear correlation is, so conduct the statistical test.

Click **Analyze** → **Correlate** → **Bivariate**
Select both variables ('anxiety2' and 'exam') of interest then click ⇨
Select **Pearson** under the correlation coefficient
Click **Options**, select **Means and standard deviations,** click **Continue**
Click **OK**

Descriptive Statistics

	Mean	Std. Deviation	N
anxiety scale score time2	37.58	8.536	40
stats exam score	70.78	10.172	40

Correlations

		anxiety scale score time2	stats exam score
anxiety scale score time2	Pearson Correlation	1	.763**
	Sig. (2-tailed)		.000
	N	40	40
stats exam score	Pearson Correlation	.763**	1
	Sig. (2-tailed)	.000	
	N	40	40

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

The table above shows the Pearson correlation coefficient of anxiety scale score time2 vs stats exam score. The Pearson correlation coefficient is .763**, which is significantly different from zero (p <.01). Therefore, the linear relationship between two variables is strong. (Remember that the Pearson's Correlation Coefficient r ranges between –1and +1, and the closer it is to +1 or -1, the stronger is the linear relationship.

**NOTE: in the correlation table, you will notice the results are mirrored. You only need to report one set of the values because they are the same.*

To answer this question:

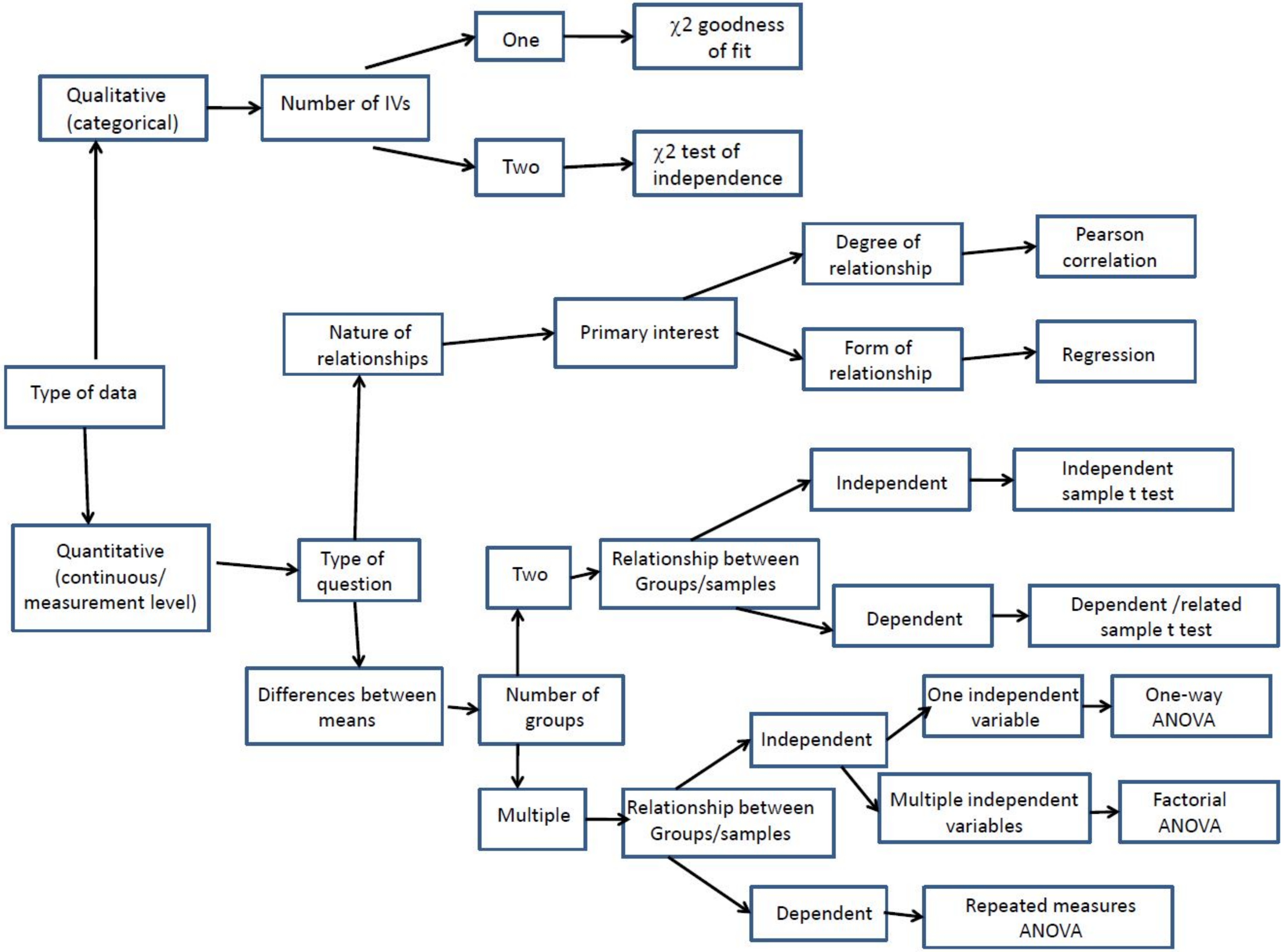
A correlation was performed to examine the relationship between scores on the anxiety scale and the scores on the stats exam.

A strong positive correlation was found between students' scores on the anxiety scale score (time 2) (*M*=37.58, *SD*= 8.54) and their stats exam score (*M*= 70.78, *SD*= 10.17); *r*(38)=.763, *p* <.01. This result indicates that as scores on the anxiety scale increase, so too does the stats exam scores and vice versa.

TEMPLATE used for writing your results:

A strong/moderate/weak positive/negative correlation was found between VARIABLE 1 (*M*= variable 1 mean, *SD*= variable 1 standard deviation) and VARIABLE 2 (*M*=variable 2 mean, *SD*= variable 2 standard deviation); *r*(N-2 (to give df))=correlation coefficient, *p* = sig. value. This result indicates…(explain what the results tell us).

Decision Trees



This is the end of this workbook