

**TEESSIDE UNIVERSITY
SCHOOL OF HEALTH & SOCIAL CARE**

SPSS Workbook 4 – T-tests

Research, Audit and data

RMH 2023-N

Module Leader: Sylvia Storey

Phone: 01642 0384969

s.storey@tees.ac.uk

SPSS – Workbook 4 – Differences between groups (T-tests)

A t-test is a statistical test that compares the mean score of the DV between 2 conditions of the IV eg *We are testing new drug B against the existing best treatment drug A and want to see which drug is most effective in reducing cholesterol levels.*

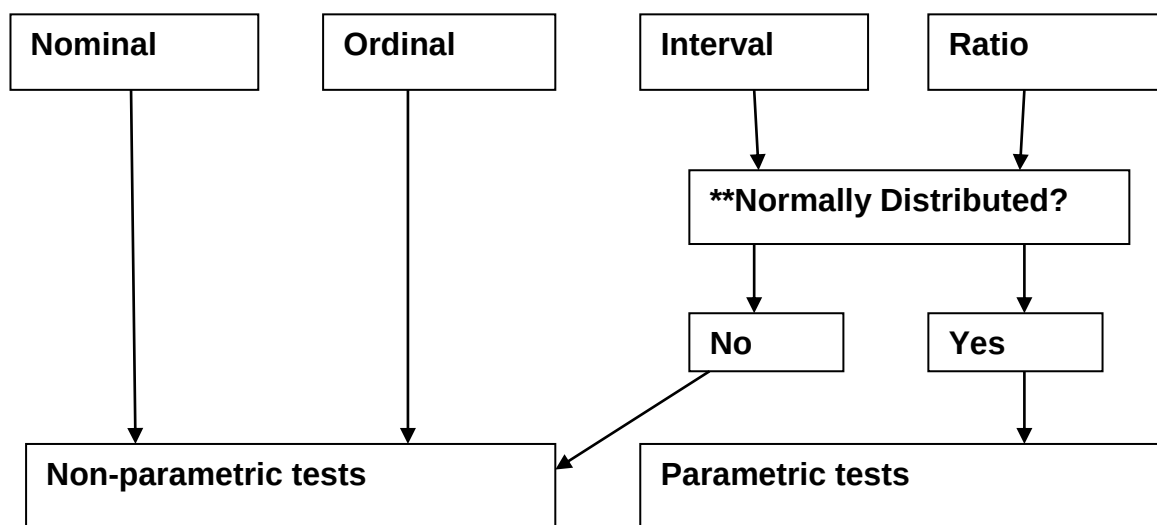
IV=treatment (2 levels : Drug A & Drug B)

DV=Cholesterol levels

The T-test test would take the mean score (cholesterol level) for each group(Drug A vs Drug B) and compare them to see if the difference is **significantly** different.

We have already mentioned that the choice of statistical test depends on various factors, the first being:

1.Level of measurement – nominal and ordinal levels of measurement are **discrete** or **categorical** variables and therefore the tests carried out on these levels of data are always **non-parametric**. (We have already looked at Chi-squared which is a type of non-parametric test). For data that is **at least interval level**, other parametric assumptions need to be met. The flow chart below shows that for data that is at least interval level, the assumption or **normal distribution** must be met. If the data is not normally distributed then a non-parametric test should be carried out.



2. Normal Distribution – you should already be familiar with this term or may have heard it referred to as a bell curve. The normal distribution of data is extremely important in statistics. Normal distribution has three important characteristics:

- it is symmetrical
- the mean, median and mode are all in the same place (ie centre of the bell curve)
- it is asymptotic (ie the tails of the distribution never touch the x-axis)

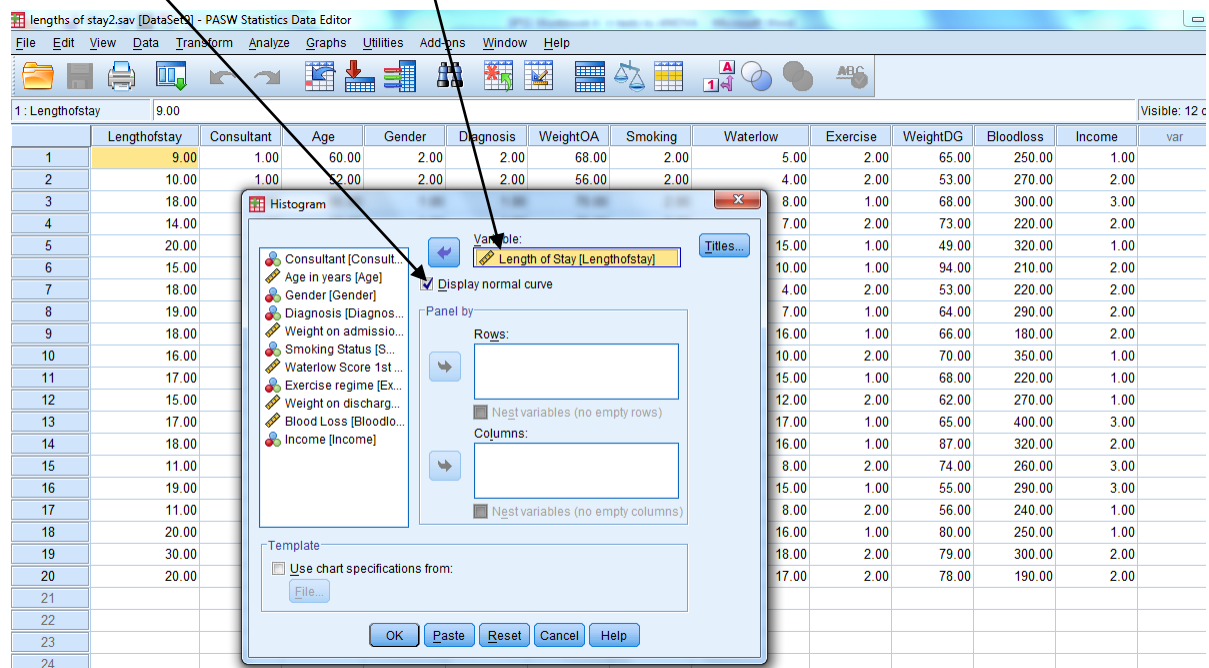
These characteristics are critical as they allow us to use **probability** statistics. A normal distribution is a theoretical concept in that your data is unlikely to ever form an exact normal distribution, but what we need to assess is that it approaches or is near to this distribution (refer back to lecture notes looking at skewed, platykurtic and leptokurtic distribution)s. We looked at sample size in a previous lecture (Measurement, Probability & Power), however in terms of normal distribution the central limit theorem states that as long as you have a reasonably large sample size (eg $n=30$), the sampling distribution of the mean will be normally distributed even if the distribution of scores in your sample is not.

There are several ways in SPSS to assess whether your data is normally distributed. The easiest way is to “eye-ball” the data. This is rather subjective and only looks at the scores of the sample and not the population.

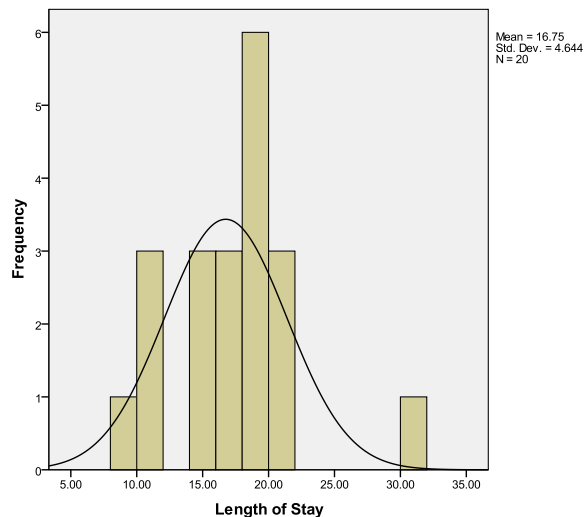
To do this open the data set from last week **lengthofstay.sav**.

1. Select **Graphs – Legacy Dialogs – Histogram**.

2. Move the variable **lengthofstay** into the **Variable** box and ensure that the **normal distribution** box is ticked.



Q1. The graph is shown below – are the data normally distributed?



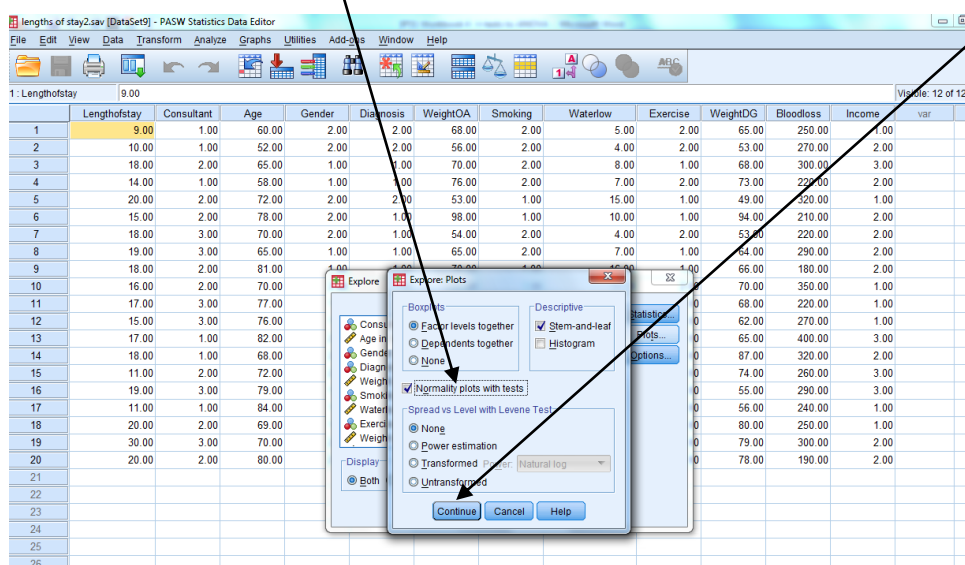
A more reliable method is to use an objective test of the distribution. The two main tests are: **Kolmogorov-Smirnoff (K-S)** and **Shapiro-Wilks (S-W)**

These tests compare the set of scores in a sample to a normally distributed set of scores with the same mean and standard deviation. If the test is non-significant (ie $p > 0.05$) then this shows that the data set is not significantly different from a normal distribution ie the data is normally distributed. If however the test statistic is significant (ie $p < 0.05$) then the data is not normally distributed. Like all statistical tests the power of these tests depends on the sample size, and in the test carried out below SPSS automatically quotes the S-W statistic when the sample size is less than 100.

1. Select – **Analyse – Descriptive Statistics – Explore**

2. Move the variables **lengthofstay**, **weightOA**, and **bloss** into the **Dependent list** box and click on **Plots**.

3. Ensure that the **Normality plots with tests** box is ticked and then click on **Continue**



The output will be large and much of it is not needed. There is a very good chapter in **SPSS: Analysis without anguish** (Coakes, 2008) that will take you through the output from this test. We will focus on the reported statistics:

Tests of Normality						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Length of Stay	.192	20	.052	.901	20	.043
Weight on admission (kg)	.109	20	.200 [*]	.964	20	.632
Blood Loss	.105	20	.200 [*]	.971	20	.774

a. Lilliefors Significance Correction

*. This is a lower bound of the true significance.

Q2. Referring to the table above which variables are normally distributed?

.....

.....

.....

Also look at the **normal Q-Q plots** for the variables. In these plots the central straight line relates to a normally distributed set of scores (ie the **expected values**) and the **observed values** (ie your data) are plotted individually.

3. Homogeneity of variance – we need to consider that the **variance** within the 2 groups is the same. For a t-test we use **Levenes test** which is produced as part of the statistical output of an independent samples t-test, so we will talk about this later.

Which t-test to use?

Today we will look at 4 types of t-test:

Independent samples t-test – a parametric test that compares the mean scores of 2 independent samples

Paired samples t-test – a parametric test that compares the mean scores of 2 paired (or repeated measures) samples

Mann Whitney (U) test – non-parametric equivalent of the independent samples t-test

Wilcoxon t-test – a non parametric equivalent of the paired samples t-test

The flow chart in the appendices shows the decision trail for deciding which t-test to carry out.

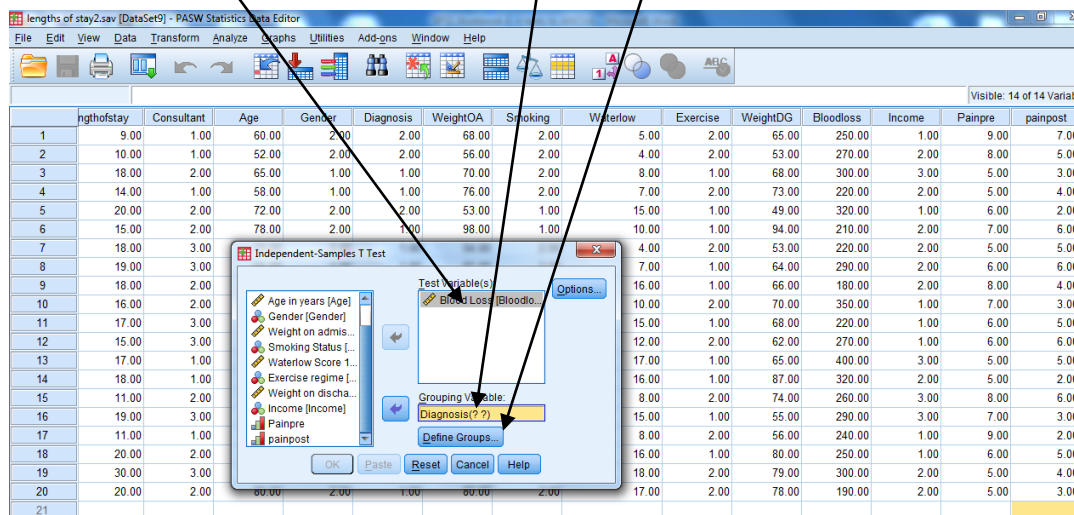
Using the **lengthofstay2** data (you will need to save this from BB) carry out the 4 t-tests described below

(State why these variables are suitable for the test allocated to them – this is in terms of level of measurement and study design as we have not checked these data for the assumption of normal distribution)

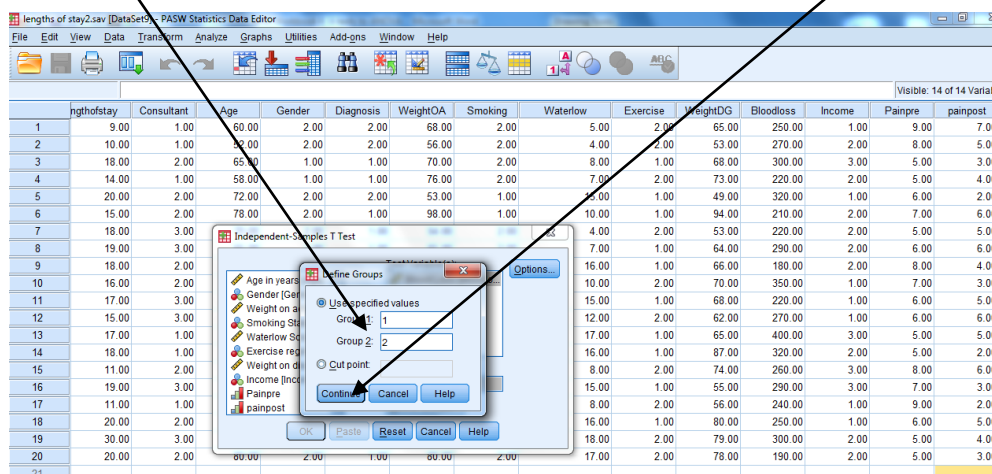
Independent Measures t-test (bloss/type)

1. Select – **Analyse – Compare Means – Independent Measures**

2. Move the **IV (Diagnosis)** into the **Grouping variable** box and the **DV (Bloss)** into the **Test variable** box and click on **Define groups**.



Enter **1 & 2** as below (why do you do this?) and click on **Continue**



The output appears as:

Group Statistics					
Diagnosis		N	Mean	Std. Deviation	Std. Error Mean
Blood Loss	Chronic Illness	14	260.7143	51.06019	13.64641
	Trauma	6	283.3333	66.53320	27.16207

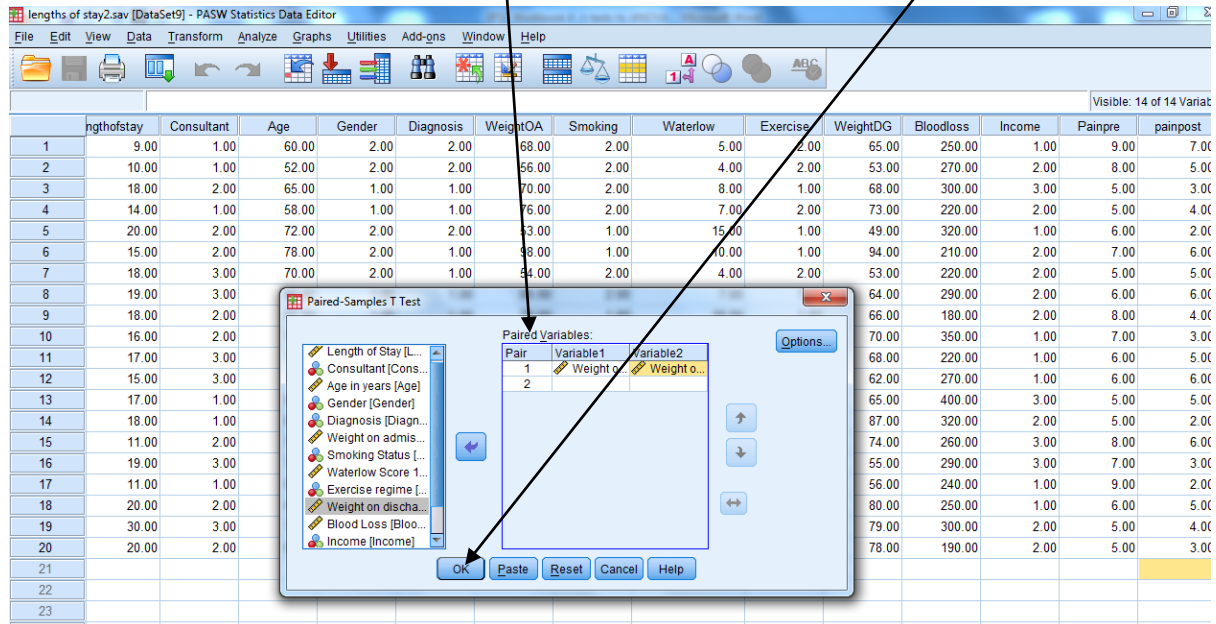
Independent Samples Test										
		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Blood Loss	Equal variances assumed	.393	.539	-.831	18	.417	-22.61905	27.22292	-79.81227	34.57418
	Equal variances not assumed			-.744	7.655	.479	-22.61905	30.39741	-93.26914	48.03104

Discuss the findings and produce a graph appropriate to the data.

Paired samples t-test (weight OA/WeightDG)

1. Select – Analyse – Compare means – Paired samples

2. Move the variables into the **Paired variables** box and click on **OK**



The output will appear as below:

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Weight on admission (kg)	70.7500	20	12.22971	2.73465
	Weight on discharge (kg)	67.9500	20	11.84316	2.64821

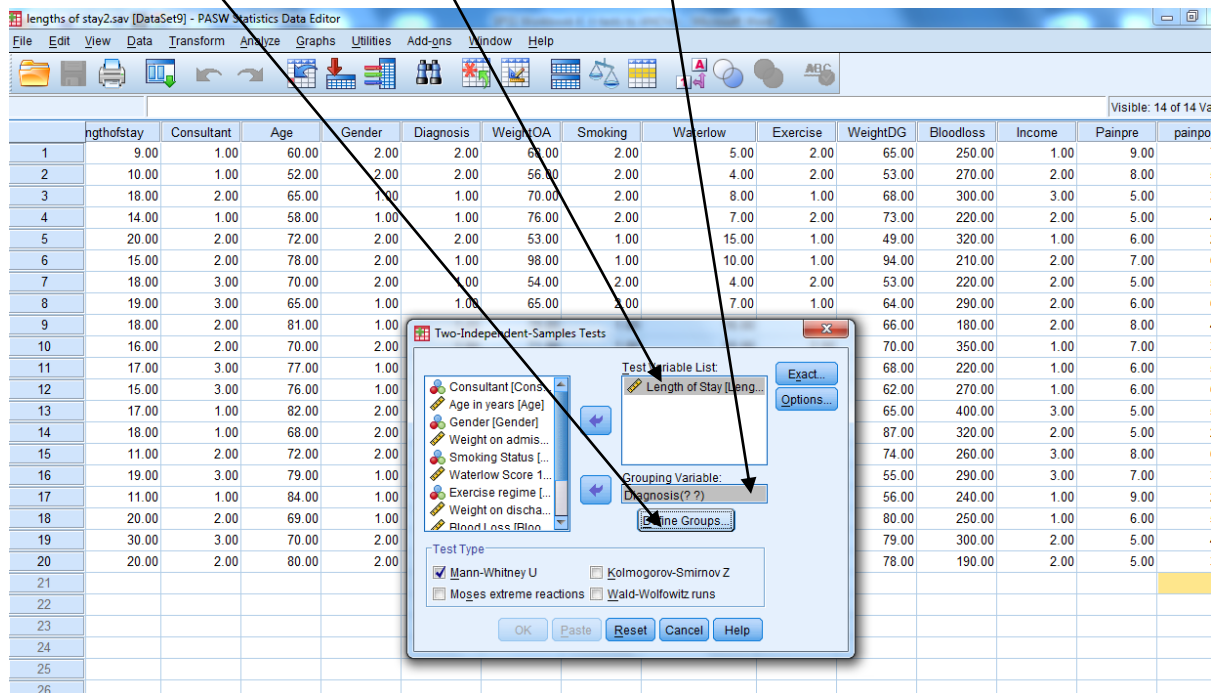
Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	Weight on admission (kg) - Weight on discharge (kg)	2.80000	1.23969	.27720	2.21981	3.38019	10.101	19	.000

Again produce a graph and discuss the findings.

Non-parametric Mann – Whitney (lengthofstay vs type)

1. Select – **Analyze – Non-parametric – Legacy dialogs - 2 Independent samples**
2. Move the variable **lengthofstay** and **Diagnosis** into the boxes as shown below and select **define groups**.



Define groups **1 & 2** as before and select **Continue** and **OK**.

The output is reported below – what does this mean

Ranks				
Diagnosis		N	Mean Rank	Sum of Ranks
Length of Stay	Chronic Illness	14	11.89	166.50
	Trauma	6	7.25	43.50
	Total	20		

Test Statistics ^b	
	Length of Stay
Mann-Whitney U	22.500
Wilcoxon W	43.500
Z	-1.619
Asymp. Sig. (2-tailed)	.105
Exact Sig. [2*(1-tailed Sig.)]	.109 ^a

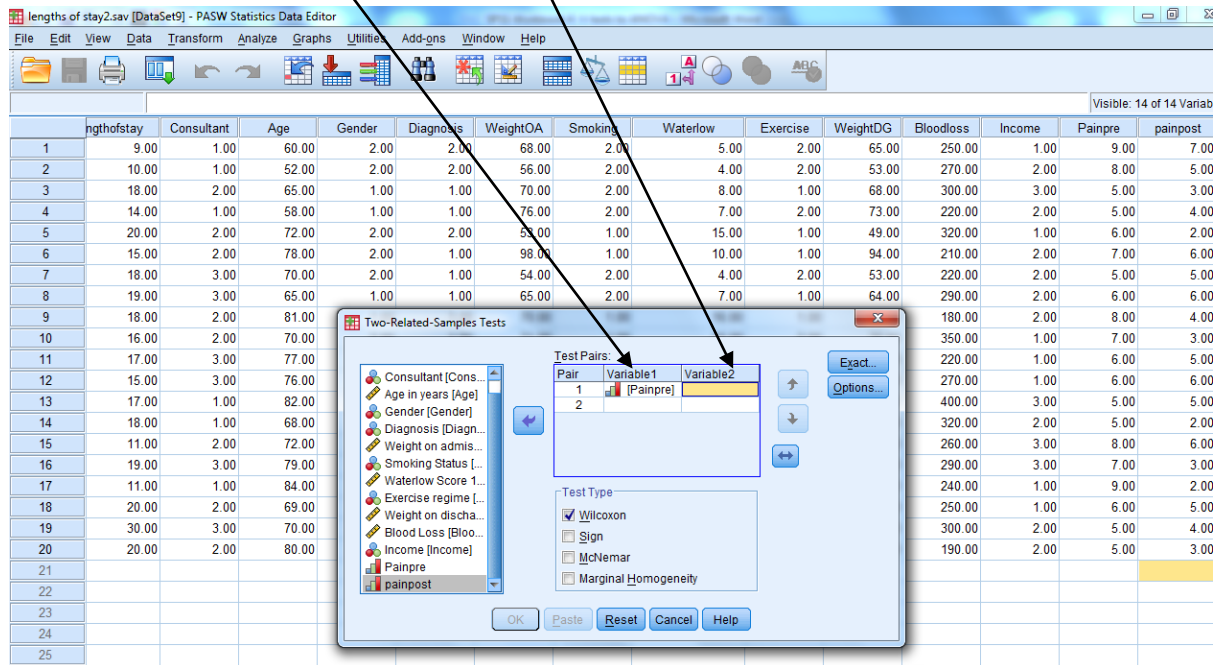
a. Not corrected for ties.

Produce a graph.

Non-parametric Wilcoxon t-test (PainPre/PainPost)

1. Select – Analyze – Non-parametric – Legacy dialogs - 2 related samples

Move the variables **Painpre** and **Painpost** into the **Test Pairs** box and select **OK**.



The data is shown below:

Ranks		N	Mean Rank	Sum of Ranks
painpost - Painpre	Negative Ranks	16 ^a	8.50	136.00
	Positive Ranks	0 ^b	.00	.00
	Ties	4 ^c		
	Total	20		

a. painpost < Painpre

b. painpost > Painpre

c. painpost = Painpre

Test Statistics ^b	
	painpost - Painpre
Z	-3.541 ^a
Asymp. Sig. (2-tailed)	.000

a. Based on positive ranks.

b. Wilcoxon Signed Ranks Test

Discuss the findings and produce a graph

ANSWERS

Appendix.

Q1. From the graph it would be feasible to say that the data is normally distributed, however, this is a subjective method and not very reliable. Look at the answer to Q2 to see if the data is in fact normally distributed.

Q2.

Tests of Normality						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Length of Stay	.192	20	.052	.901	20	.043
Weight on admission (kg)	.109	20	.200 [*]	.964	20	.632
Blood Loss	.105	20	.200 [*]	.971	20	.774

a. Lilliefors Significance Correction

*. This is a lower bound of the true significance.

As $p > 0.05$ the variables **Blood loss & Weight on admission** are normally distributed.

However, the p-value for **length of stay** is 0.043 and as this is less than 0.05, the data is not normally distributed.

Independent samples t-test.

The DV "Bloss" is of at least interval level data and is normally distributed (see Q2 above). This indicates that a parametric t-test can be carried out. As the study design is independent measures ie patients are admitted either through trauma or chronic illness. Using the flow chart you will see that the "independent measures t-test" is the appropriate choice of test.

The findings show:

Group Statistics					
Diagnosis		N	Mean	Std. Deviation	Std. Error Mean
Blood Loss	Chronic Illness	14	260.7143	51.06019	13.64641
	Trauma	6	283.3333	66.53320	27.16207

The first box in the output shows that :

1. there were 14 patients admitted due to chronic illness and 6 due to trauma.
2. The mean blood loss for patients admitted due to trauma was more than that for patients admitted due to chronic illness (283.333 compared to 260.7143)

The assumption of equal variances is reported in the box below and like assessing for normal distribution the p-value should be >0.05 ie no significant difference between the groups (they are equal). As $p=0.539$ (>0.05) then we use the values from the top line of the table "Equal variances assumed").

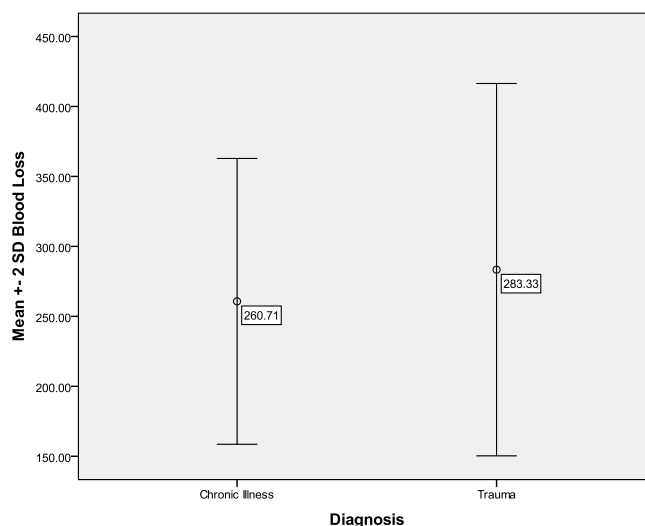
Although the statistics suggest that there is a difference in the amount of blood lost – this is in fact not significant as shown by the p value below which is >0.05 .

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
Blood Loss	Equal variances assumed	.393	.539	-0.831	18	.417	-22.61905	27.22292	-79.81227	34.57418
	Equal variances not assumed			-.744	7.655	.479	-22.61905	30.39741	-93.26914	48.03104

In a report you would express the results of this test as:

There was no significant difference in the amount of blood lost between patients admitted for trauma and those admitted through chronic illness ($t=-0.831$, $df=18$, $p=0.417$)

An Error – bar should be produced as below:



Paired samples t-test (weight on admission/weight on discharge)

The data is at least interval level and from Q2 we can see that the variables weight on admission is normally distributed. (You would also need to check that weight on discharge is also normally distributed).

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Weight on admission (kg)	70.7500	20	12.22971	2.73465
	Weight on discharge (kg)	67.9500	20	11.84316	2.64821

The descriptive statistics shown above suggest that patients lose weight during their stay in hospital ie the mean weight on admission is 70.75kg but on discharge this is 67.95kg: a mean reduction of 2.7kg.

The table below shows the results of the t-test test. You will notice that there are no boxes referring to the Levene's test – this is because we do not have to assess paired samples t-test for homogeneity of variances.

In this test the result is significant – which is demonstrated by the p-value being <0.05

Paired Samples Test

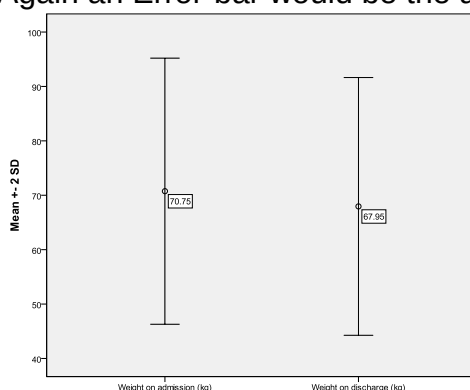
		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	Weight on admission (kg) - Weight on discharge (kg)	2.80000	1.23969	.27720	2.21981	3.38019	10.101	19	.000

In a report you would express the results of this test as:

NB – you need to round this up to $p < 0.001$

There was a significant difference between weight on admission and weight on discharge ($t=10.101$, $df=19$, $p<0.001$), with patients weighing less on discharge than they did on admission. (Discharge Mean = 67.95: Admission Mean = 70.75kg).

Again an Error-bar would be the appropriate graph:



Non-parametric Mann-Whitney

Normal distribution of the DV “length of stay” was assessed and shown to be not normally distributed (Q2). We therefore have to carry out a non-parametric t-test. Using the flow chart we can see that the appropriate choice when the study is an independent measures design (IV – diagnosis 2 conditions: trauma, chronic illness), is the Mann Whitney (U) test ie this test is the **non-parametric equivalent** of the independent samples t-test carried out earlier.

		Ranks		
	Diagnosis	N	Mean Rank	Sum of Ranks
Length of Stay	Chronic Illness	14	11.89	166.50
	Trauma	6	7.25	43.50
	Total	20		

The **Mean rank** shows the mean rank of scores within each group, whilst the **Sum of ranks** shows the total sum of all ranks within each group.

If there were no differences between the 2 groups we would expect these to be roughly equal for each group.

Test Statistics ^b	
	Length of Stay
Mann-Whitney U	22.500
Wilcoxon W	43.500
Z	-1.619
Asymp. Sig. (2-tailed)	.105
Exact Sig. [2*(1-tailed Sig.)]	.109 ^a

The test statistic is **Mann Whitney U** (22.5)

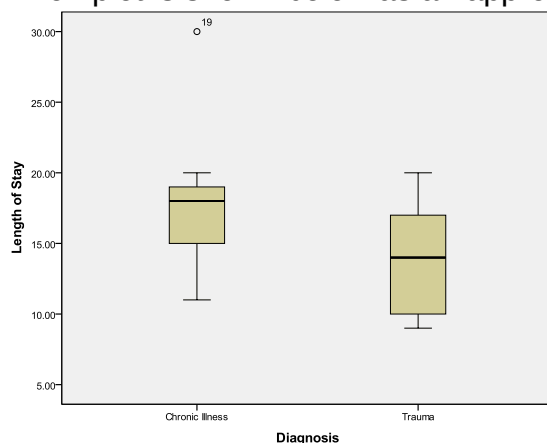
The **p value** at 0.105 shows that there is no significant difference between the 2 groups.

a. Not corrected for ties.

In a report you would write:

There was no significant difference in the length of stay between patients admitted due to trauma and those admitted due to chronic illness ($U=22.5$, $p=0.105$).

A Box-plot is shown below as an appropriate graph:



Non-parametric Wilcoxon t-test

The variables **painpre** and **painpost** are classed as ordinal level data (although you may find references that would justify them being interval/ratio level data). For ordinal level data **non-parametric** tests are carried out. As this is repeated measures design (ie same set of patients in both condition) then the test to be carried out is the **Wilcoxon t-test**.

Ranks		N	Mean Rank	Sum of Ranks
painpost - Painpre	Negative Ranks	16 ^a	8.50	136.00
	Positive Ranks	0 ^b	.00	.00
	Ties	4 ^c		
	Total	20		

a. painpost < Painpre

b. painpost > Painpre

c. painpost = Painpre

The values above are sorted into **positive** and **negative** ranks and **ties**.

These relate to the equations shown below ie the negative ranks (a) relate to patients where the painpost score is less than the painpre score.

Test Statistics ^b	
	painpost - Painpre
Z	-3.541 ^a
Asymp. Sig. (2-tailed)	.000

The test statistic is shown as **Z** (we will look at Z scores later).

And the **p-value** is <0.001 (remember you need to round this up) which shows that the result of the test is significant.

In a report you would write:

There was a significant difference in pain scores pre and post admission to hospital ($Z=-3.541$, $p<0.001$)

Produce a graph (box-plot):

