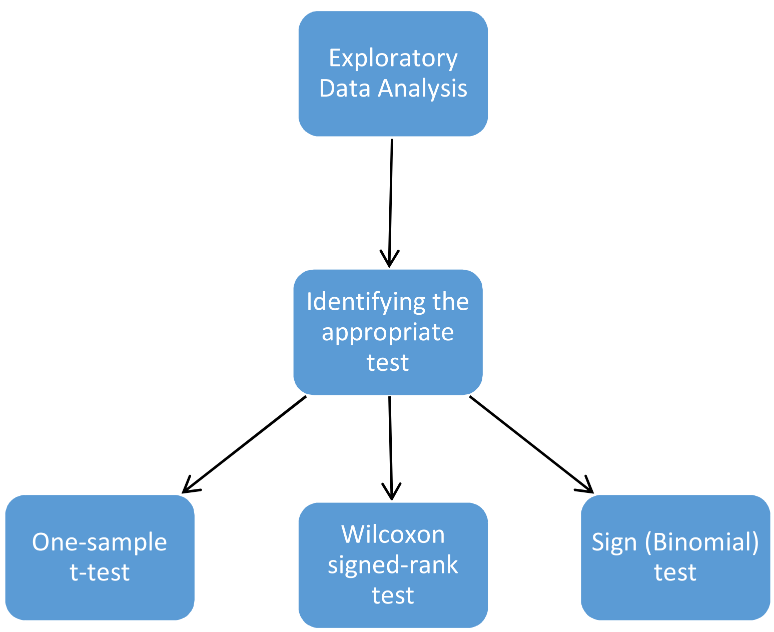
DEWIS Learning Modules

One-sample test of location



Interpreting the relevant SPSS output and decision rules

These resources have been made available under a Creative Commons licence by Iain Weir, Rhys Gwynllyw & Karen Henderson, University of the West of England, Bristol and reviewed by Nadarajah Ramesh, University of Greenwich.

Introduction

These notes concern the identification of an appropriate one-sample location test for a given data set. A set of recommended decision rules is presented for this purpose.

The accompanying set of five DEWIS Learning Modules allow the application of these decision rules to a variety of data sets that are generated from many practical contexts.

The notes emphasis is on using and interpreting output from the statistical package SPSS to perform the relevant analysis. In this document the SPSS output is merely presented; the creation is left to the accompanying set of videos for this set of DEWIS Learning Modules that utilise the example that is presented here.

One sample location tests

It is common to quantify the location of a sample by using either the mean or the median. The first choice is usually the mean, however in the presence of skewness and/or outliers it can be distorted to be an unrepresentative “typical” central value. In such cases the median is preferred as it is robust at giving a sensible mid-point estimate. Likewise variances can be overestimated by skewness and/or outliers in which case it may be prudent to consider inter quartile ranges as a robust alternative measure of spread.

The choice of test of location is also dependent on the properties of the sample data. The three tests we consider here are as following in the order of the desirability of application due to the power (probability of correctly rejecting a false hypothesis) they are capable of achieving.

- **One-sample t-test** is a parametric test for the mean that has the assumption the population is normal.
- **Wilcoxon signed-rank test** is a nonparametric test for the median which is the usual alternative test to the one sample t-test when we cannot assume normality. However, the test does assume that the population probability distribution is symmetrical.
- **Sign (Binomial) test** is a nonparametric test for the median that does not assume that the population probability distribution is symmetric. However, it is a less powerful alternative to the Wilcoxon signed-rank test.

All three tests also require that the sample consists of random and independent observations which is usually a consideration of the sampling design. We shall assume here that these requirements have been met.

The violation of test assumptions can result in an analysis that may be incorrect or misleading. The size of a sample is important to consider when considering which test to use.

- If the sample size is small it may be difficult to detect assumption violations; normality tests cannot be relied upon. Also outliers/skewness in the data will have a greater consequential effect in the distorting of means/variances and hence t-test statistics.

- If sample sizes are large then, as a consequence of the Central Limit Theorem, the one-sample t-test is somewhat robust to violations of the assumption of normality.

As such there exist in the literature various rules of thumb for the selection of an appropriate test that consider combinations of the following properties of a sample:

- Sample size;
- Outliers (mild and extreme);
- Skewness;
- Normality.

For this DEWIS based learning modules we shall use rules of thumb that consider the above where we have categorised sample size into the following;

- very small < 10
- small 10-15
- moderate 16-29
- large 30-39
- very large 40+

For each of these sample size categories there are decision trees in the Appendix that take into account outliers, skewness and normality and recommend the appropriate one-sample location test.

We shall now consider through an example the interpretation of the SPSS output for the Exploratory Data Analysis (EDA) of a one-sample data set to identify the appropriate test for location for it. After that the interpretation of the SPSS output for each of the three tests is considered.

Note that if we find evidence for non-normality then we can either transform the data or use **nonparametric tests** that don't require normally distributed data. In these notes and the DEWIS learning modules we consider the latter approach. We also restrict ourselves to considering just two sided alternative hypotheses but the approach here can be extended to one sided alternatives.

Example: water usage

In a study of water usage in a particular town the number of litres of water utilised per day in the following random sample of homes was recorded.

486 390 509 473 523 532 555 600 380 459

536 500 550 564 473 450 330 600 559 500

In the past the average water usage was 450 litres per day. Does this sample provide any evidence that the average water usage has changed?

Here we ideally want to test using the one-sample t-test whether;

$$H_0:\ \mu = 450$$

$$H_1: \mu \neq 450$$

However before performing this test we need to perform Exploratory Data Analysis to confirm that it is indeed the appropriate test.

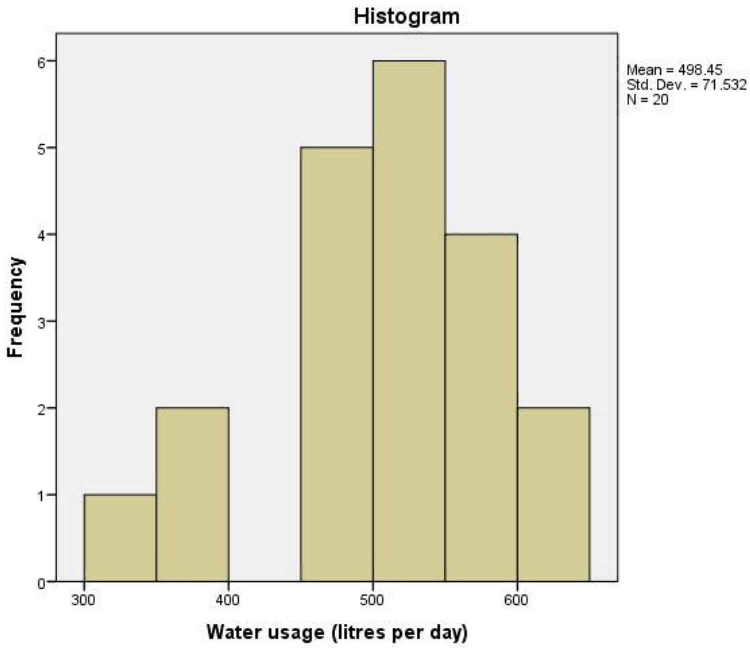
Exploratory Data Analysis

Explore

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Water usage (litres per day)	20	100.0%	0	.0%	20	100.0%

The table above shows us that we have (as expected) 20 observations and no missing data.



The above histogram displays the shape of the distribution and provides estimates of mean and SD. Superficially this data looks slightly negatively skewed, however you must recall that we have only 20 data points and we must leave the judgement of normality to the formal statistical test.

Descriptives

			Statistic	Std. Error
Water usage (litres per day)	Mean		498.45	15.995
	95% Confidence Interval for Mean	Lower Bound	464.97	
		Upper Bound	531.93	
	5% Trimmed Mean		502.17	
	Median		504.50	
	Variance		5116.787	
	Std. Deviation		71.532	
	Minimum		330	
	Maximum		600	
	Range		270	
	Interquartile Range		91	
	Skewness		-.784	.512
	Kurtosis		.364	.992

The table above presents a multitude of statistics that summarise the distribution of our data; we have information on location, dispersion and distribution shape.

The mean ($\bar{x}$) of the data is 498.45 litres and gives us an idea of where the data is centrally located; it estimates the unknown population mean (μ). However, quoting this statistic is of little use without some indication of the accuracy of this estimate. The corresponding standard error (SE) gives us an idea of the variability of this estimate. Mathematically these two are combined to obtain a 95% confidence interval (CI) for the unknown population mean; we are 95% sure that the interval from 464.97 to 531.93 litres covers the unknown population mean.

The CI is a plausible range of values that the unknown population mean could take. Thus thinking ahead to the future analysis of our data there already is cause for concern if the reservoir is set up to cope with an average usage of 450 litres. However, this estimate is obtained from a formula that assumes the data is normally distributed. As such we should also be assessing this assumption.

Alternative measures of central location are provided by the 5% Trimmed Mean (5% of lowest and highest scores are discarded before calculating the mean) and the median. Both of these estimates are robust against the presence of outliers. If we have outliers then $\bar{x}$ may be a severely biased estimate of the central location of the population. Do we have outliers in our data? Read on….

The variance, and standard deviation (SD) give us estimates of the spread of the data. The SD is more useful measure (and hence more often quoted) as it is on the same scale as the observed data. A layman's interpretation of SD is that it is an estimate of the average deviation of observations around the population mean.

The minimum and maximum values are also reported which give us estimates of extreme values that the data can take. The range is an alternative measure of spread which is the distance between these two values.

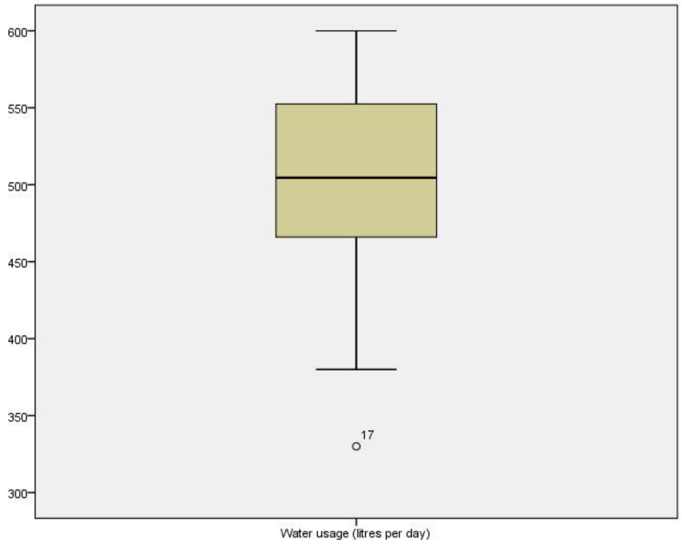
The above estimates of spread are potentially distorted in the presence of outliers, the interquartile range given in the Descriptives table (IQR) is 91 litres and is a robust measure of spread as it considers only the middle 50% of the data. The actual values of the quartiles (and other percentiles) can be obtained from the following table.

Percentiles

		Percentiles						
		5	10	25	50	75	90	95
Weighted Average (Definition 1)	Water usage (litres per day)	332.50	381.00	462.50	504.50	553.75	596.40	600.00
Tukey's Hinges	Water usage (litres per day)			466.00	504.50	552.50		

For instance we can see that the lower quartile (25^{th} Percentile) is 462.50 litres which gives us some concern because over 75% of the houses appear to have a usage above that value and hence above the hypothesised mean of 450 litres. Note; in the e-Assessment the lower and upper quartiles use the Weighted Average definition.

The minimum, lower quartile, median, upper quartile and maximum are 5 summary statistics that are used to draw a basic boxplot. SPSS produces one that is slightly more sophisticated in that it identifies outliers.



The boxplot is a useful way of summarising data;

- the median is the thick bar in the middle of the box;
- the box's top and bottom edges are respectively the upper and lower quartile;
- the range of representative values is defined by the whiskers;
- outliers are individually identified using circles for **mild outliers** and stars for **extreme outliers** along with their case numbers for identification.

We can see that the 17th observations is deemed to be a mild outliers. The actual value for this observation is 330 litres per day (see the Extreme Values table below for quick identification of this value).

Extreme Values

			Case Number	Value
Water usage (litres per day)	Highest	1	8	600
		2	18	600
		3	14	564
		4	19	559
		5	7	555
	Lowest	1	17	330
		2	9	380
		3	2	390
		4	16	450
		5	10	459

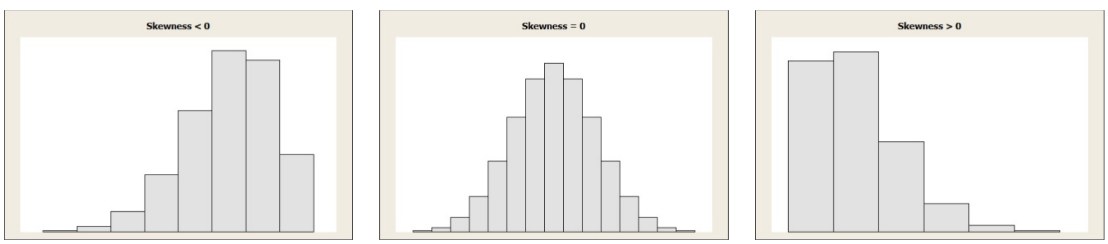
We can use the boxplot to visually assess normality:

- If a distribution is normal then the median should be positioned in the centre of the box and the whiskers should be of roughly equal length.
- If the median is closer to the top of the box and the corresponding whisker is shorter than the one at the bottom of the box then the distribution is negatively skewed.
- If the median is closer to the bottom of the box and the corresponding whisker is shorter than the one at the top of the box, then the distribution is positively skewed.

The spread of the data can be determined by either the width of the box (= IQR) or the length between the whiskers (= range after outliers removed).

Returning to the Descriptives table the remaining two statistics qualify the shape of the distribution and indicate how much a distribution varies from a normal distribution.

Skewness tells you the amount and direction of skew:



The skewness coefficient is thus a measure of the (lack of) symmetry.

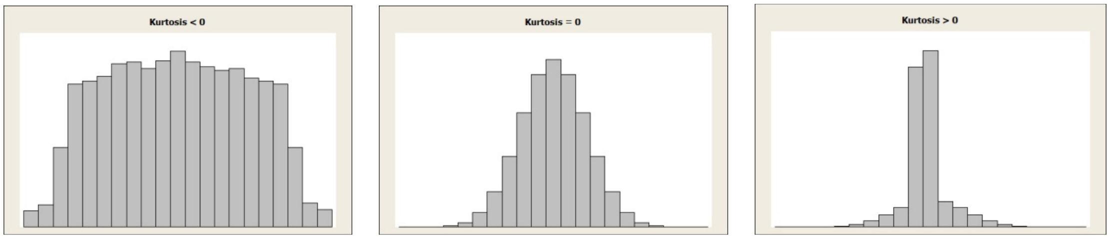
- The skewness coefficient of a symmetrical distribution is zero.
- A -ve skewness coefficient indicates a tendency for negative skewness.
- A +ve skewness coefficient indicates a tendency for positive skewness.

A normal distribution has skewness of 0, so if a skewness coefficient is close to zero then the distribution it is drawn from is consistent with being normal. The larger the absolute value of either statistic is then the greater the departure from normality.

We need to examine the skewness coefficient to see if it is significantly different to zero. The corresponding standard error (SE) helps us make this decision; if the absolute value of the symmetry statistic is less than $2 \times SE$ then the data is not deemed to have a sufficiently large enough skewness for us to believe it is different to zero. If this condition is ok, then our data is consistent with normality otherwise it is not.

Here twice the SE is 0.512 x 2 = 1.024 and the absolute value of the coefficient (0.784) is not greater than this. Thus we do not appear to have evidence of a skewed distribution.

Kurtosis tells you how tall and sharp the central peak is relative to a normal distribution:



Platykurtic Mesokurtic Leptokurtic

- The kurtosis of a normal distribution is zero ("mesokurtic").
- Data with -ve kurtosis ("platykurtic") tend to have a flat top near the mean rather than a sharp peak.
- Data with +ve kurtosis ("leptokurtic") tend to have a distinct peak near the mean, decline rather rapidly, and have heavy tails.

We need to examine the kurtosis coefficient to see if it is significantly different to zero. The corresponding standard error helps us make this decision; if the absolute value of the coefficient is greater than twice the standard error then this is indicative of evidence to suggest the distribution is not mesokurtic. Here twice the SE is 0.992 x 2 = 1.984 and the absolute value of the coefficient (0.364) is not greater than this. Thus we do not appear to have evidence of a non-mesokurtic distribution.

Note we do not have an immediate way of obtaining a p-value to test for zero skewness or kurtosis.

The skewness and kurtosis coefficients are therefore both supporting the belief that our data may have come from a normal distribution. We will see later in the decision rules that when we have data that fails a normality test it is sometimes of interest to know the cause and in particular to know whether it is due to skewness or not.

We can formally test for normality using the following output.

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Water usage (litres per day)	.111	20	.200*	.945	20	.292

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

a. Lilliefors Significance Correction

The above allows us to test between the following hypotheses:

H_0: The data are normally distributed

H_1: The data are not normally distributed

There are two tests presented but only one should be used. It is recommended to use the Shapiro-Wilk test.

The Significance value (or p-value) is an indication of the likelihood of our observed data if the Null hypothesis is true (i.e. the data was from a normal population). If we have a low significance value (generally less than 0.05) then we reject the Null Hypothesis in favour of the Alternative Hypothesis which would indicate that the distribution of the data differs significantly from a normal distribution. However here we have a non-significant p-value of 0.292 which indicates that we have no evidence to suggest non-normality, that is we can assume normality. We can report this as follows:

The data was tested for normality using the Shapiro-Wilk (S-W) test and resulted in there being no evidence of any departure from normality (S-W(20) = .945, p = .292).

We have now explored the data and produced some graphics and summary statistics that summarise our data. We have established that this data set:

has a sample size of 20;

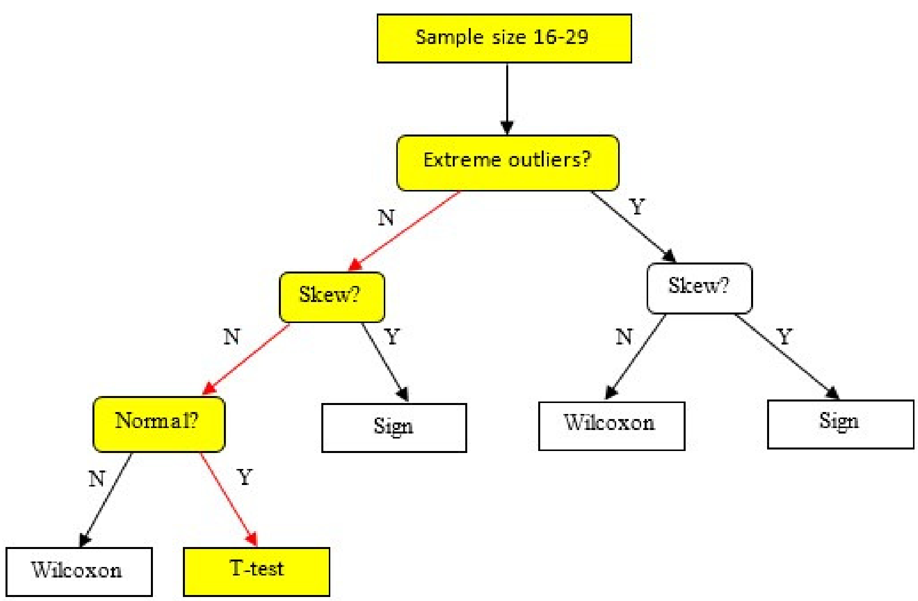
has 1 mild outlier;

is not skewed;

is normal.

In order to identify the appropriate one-sample location test we shall now refer to the decision tree in the appendix which covers data with sample sizes of 16-29.

With data sets of this size if the data has extreme outliers and/or is skewed it is unwise to use the one-sample t-test; we have no concerns here. Our data appears to be normal and thus from the following we can see that for ***this*** data set the recommended test is the one-sample t-test.



We shall now consider the interpretation of the SPSS output for all three tests.

One-sample t-test

T-Test

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Water usage (litres per day)	20	498.45	71.532	15.995

The first table reproduces some statistics that we have already seen via the *Explore* command.

One-Sample Test

	Test Value = 450					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Water usage (litres per day)	3.029	19	.007	48.450	14.97	81.93

From the above we see that the p-value is less than 0.05 thus we reject H_0 in favour of H_1 which indicates that the mean water usage is not 450 litres per day.

The Mean Difference informs us that it appears that the true average water usage is 48.450 litres per day higher than the previous value of 450. A 95% CI for this difference is from 14.97 to 81.93 litres per day. We can report this as follows:

Application of the one sample t-test provides strong evidence that the mean is not 450 litres per day (t(19) = 3.029, *p* = .007). It appears to be 48.450 (14.97, 81.93) litres per day higher.

Note when reporting the degree of evidence against for Ho:

- p > .05 “no evidence”;
- p ≤.05 “some evidence”;
- p ≤.01 “strong evidence”;
- p ≤.001 “very strong evidence”.

NOTE: The first choice for ***this*** data set is the t-test as this is the most powerful test as the data is normal (i.e. most likely to detect a difference).

One-sample Wilcoxon signed-rank test

The **one-sample Wilcoxon signed-rank test** is a test about medians rather than means:

$$H_0: \text{median} = 450$$

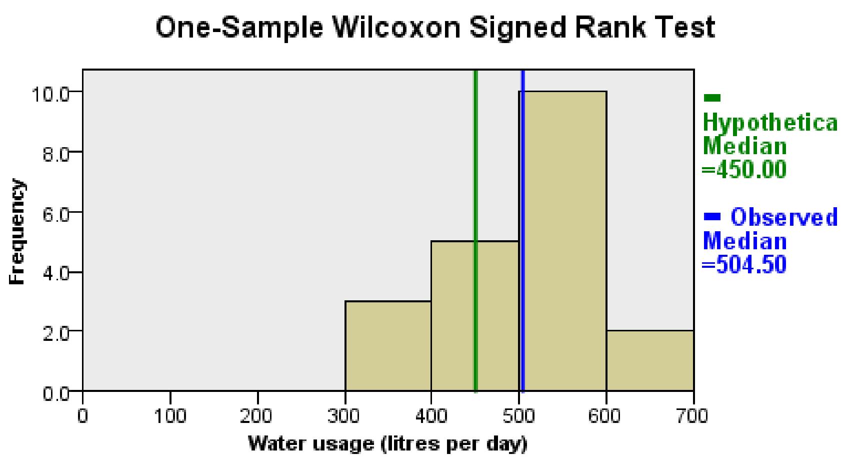
$$H_1: \text{median} \neq 450$$

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The median of Water usage (litres per day) equals 450.00.	One-Sample Wilcoxon Signed Rank Test	.014	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

From the above we see that the p-value is less than 0.05 thus we reject H_0 in favour of H_1 which indicates that the median water usage is not 450 litres per day.



Total N	20
Test Statistic	156.000
Standard Error	24.842
Standardized Test Statistic	2.456
Asymptotic Sig. (2-sided test)	.014

As $p < 0.05$ we reject the Null Hypothesis and conclude that the median is not 450 litres per day; the median appears to be higher at around 504.50 litres per day. We can report this as follows:

Application of the one sample Wilcoxon signed-rank test provides evidence that the median is not 450 litres per day ($z = 2.456$, $N=20$, $p = .014$). It appears to be higher at around 504.50 litres per day.

NOTE: The first choice for ***this*** data set is the t-test as this is the most powerful test as the data is normal (i.e. most likely to detect a difference).

Sign (Binomial) test

The Sign (Binomial) test is a nonparametric test of medians rather than means:

$$H_0: \text{median} = 450$$

$$H_1: \text{median} \neq 450$$

The Sign (Binomial) test requires as part of its calculation the number of observations that are both smaller and bigger than the median under the null hypothesis; ties are excluded. In this data set we have one value equal to the null hypothesis which is excluded before requesting the output.

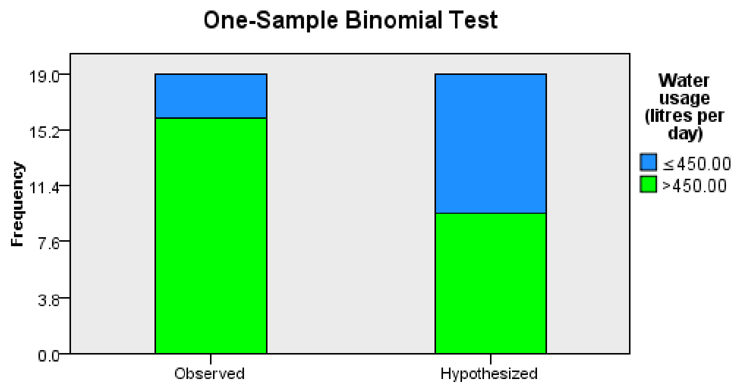
Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The categories defined by Water usage (litres per day) <=450.00 and >450.00 occur with probabilities 0.5 and 0.5.	One-Sample Binomial Test	.004[1]	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

[1]Exact significance is displayed for this test.

From the above we see that the p-value is less than 0.05 thus we reject H_0 in favour of H_1 which indicates that the median water usage is not 450 litres per day.



Total N	19
Test Statistic	3.000
Standard Error	2.179
Standardized Test Statistic	-2.753
Asymptotic Sig. (2-sided test)	.006
Exact Sig. (2-sided test)	.004

If the sample size is 25 or less, SPSS reports an Exact p-value as well as the usual Asymptotic one. If the Exact p-value is available we should report this rather than the Asymptotic one.

As Exact $p < 0.05$ we reject the Null Hypothesis and conclude that the median is not 450 litres per day; the median appears to be higher at around 504.50 litres per day. Note that the observed median is not reported here; we have to look back to our EDA output.

We can report this as follows:

Application of the one-sample Sign (Binomial) test provides strong evidence that the median is not 450 litres per day (N=20, Exact p = .004). It appears to be higher at around 504.50 litres per day.

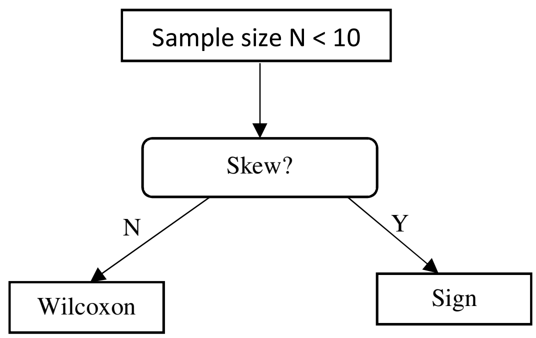
Note N here refers to our sample size pre-exclusions.

NOTE: The first choice for ***this*** data set is the t-test as this is the most powerful test as the data is normal (i.e. most likely to detect a difference).

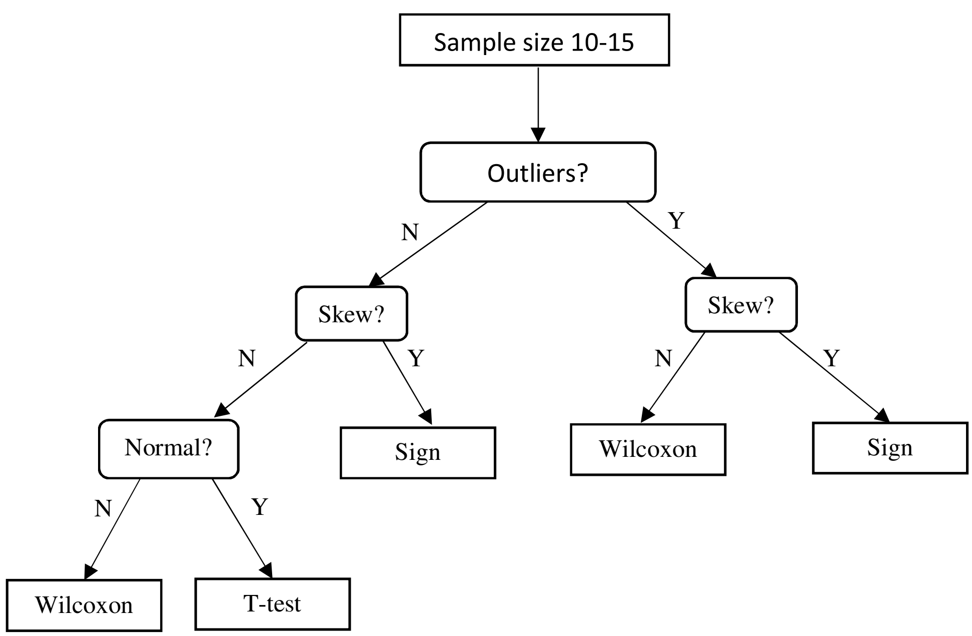
Appendix 1: decision trees identifying the appropriate one-sample location test

The following decision trees are used in the DEWIS learning modules to recommend the appropriate one-sample location test.

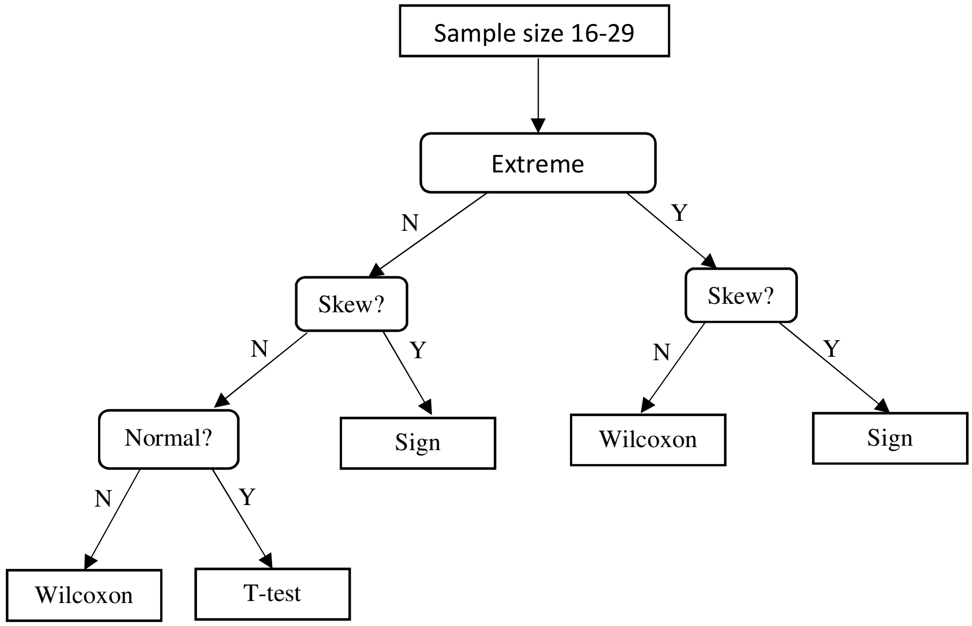
Very small < 10: normality tests cannot be relied upon and thus it is unwise to use the one-sample t-test. Considering skewness will allow you to decide between the two nonparametric tests.



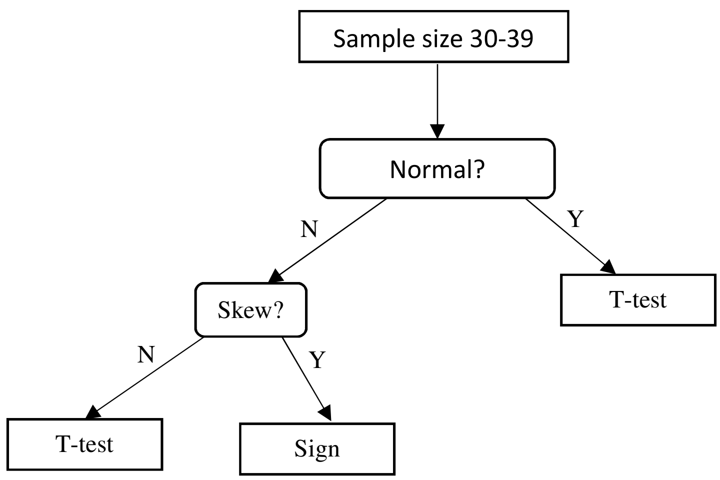
Small 10-15: the data set is big enough to consider the one-sample t-test. However if the data has outliers and/or is skewed it is unwise to use the one-sample t-test.



Moderate 16-29: if the data has extreme outliers and/or is skewed it is unwise to use the one-sample t-test.



Large 30-30: if the data is not normal but is symmetrical it is ok to use the one-sample t-test.



Very large 40+: Due to the Central Limit Theorem it is safe to use the one-sample t-test regardless of outliers, skewness or failing a normality test.