TurboCalcStats

Nick Bergquist

COLLABORATORS			
	TITLE : TurboCalcStats		
ACTION	*NAME*	*DATE*	*SIGNATURE*
WRITTEN BY	Nick Bergquist	January 19, 2025	

REVISION HISTORY			
NUMBER	DATE	DESCRIPTION	NAME

Contents

Chapter 1

TurboCalcStats

1.1 Front page...

```
Welcome to the AmigaGuide documentation for

~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
     "TCalcStats v2.0c (pre-release)" - A series of statistical analysis tools ↩
        for
       TurboCalc v5.xx

~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
WARNING: This is an "almost ready for release" version. The Guide and HTML files ↩
   are not
completed. Some of the ARexx scripts may need correction/improvement.

Contents:
```

~~~~~~

by Rudy Kohut and Nick Bergquist, 1999

1.2 Introduction...

· Preface TCalcStats v2.0b.
~~~~~~~~~~~~~~~~~~~~~~~

What is it?! In brief... tools for number crunching!

The included ARexx scripts will allow you to perform a comprehensive range of statistical analysis tests on raw sample data. The statistical approach and the workings of all these scripts are extensively explained in this AmigaGuide and the associated HTML documentation. In effect, little prior knowledge of statistical analysis is assumed for application of many of the test procedures. This package is aimed at:

- Scientists researching in physical, biological and social sciences, etc.

- Students who are likely to need to conduct statistical analysis as part of their studies.

- Market researchers and anyone else who wants to analyze trends in sample data.

Although several statistical analysis programs exist for the Amiga platform this suite of ARexx scripts have been designed to provide an extensive range of summary statistics, parametric and additional non-parametric tests of hypotheses for a variety of data sources.

```
**********************************************************************
*                           *
*  This package is a series of analysis tools designed for use with  *
*  TurboCalc v5.x by Michael Friedrich.          *
*                           *
**********************************************************************
```

A similar set of tools are available for Microsoft Excel on the PC and Mac in the form of 'add-ins' but this package is intended to perform beyond these existing capabilities. It contains numerous extra features:

- Greater support for non-parametric tests such as:

  - Test for normality in data distributions.

- Mann-Whitney test.
- Wilcoxon paired sample or signed rank test.
- Kruskal-Wallis test.
- Friedman rank test... Etc.

- Support for unusual or specialised tests:
  - Shannon-Wiener biodiversity index test.
- Enhanced output. Results include probability and critical value figures for both one-tailed and two-tailed tests at both 0.05 and 0.01 levels of significance, where applicable.

1.3 An introduction to statistical analysis...

· Statistical analysis.
~~~~~~~~~~~~~~~~~~~~

Be warned that this is a bit of a 'bare bones' crash course in statistical analysis! If it is either incomprehensible or on the other hand you find it lacks sufficient detail then please obtain an introductory text or take a look at the references respectively.

Numerical data gathered as part of scientific investigations is often subject to some form of statistical analysis in order that objective evaluation of this data may be made in relation to experimental expectations and conclusions. Methods of statistical analysis can be categorised:

1. Descriptive statistics

This form of analysis is based on the organisation and summary of data and is aimed at describing particular characteristics of a sample set of data in order to obtain information about the overall population data. Under many circumstances the population data is usually unavailable and its characteristics may only be ascertained from a sample. For example, the true mean height of all lamp-posts in Germany is physically unobtainable as it would entail measuring them all. The subject of descriptive statistics may also be broken down into two broad areas of study:

(a.)Measures of central tendency

Analysis aimed at describing properties of statistical populations by determination of the most typical data element. In other words, measures of central tendency represent central or focal points in the distribution of data elements. Typical examples of this type of analysis include measurement of the mode (most frequently occurring element of data), the median (point constituting the middle measurement in a set of data) and the mean (the average or sum of all measurements divided by the number of measurements).
~~~~~~~~~~~~~~~~~~~~

(b.)Measures of dispersion and variability

Analysis designed to indicate the extent of scatter or clustering of data measurements around the centre of a data distribution. Typical examples of this type of analysis include calculation of the range (difference between the highest and lowest measurements in a group of data - a fairly crude measure of dispersion), the standard deviation (a measure of data dispersion in the form of a determination of the average extent of deviation of the data values from the mean) and the variance (a measure of data dispersion in the form of a determination of the average of the squared deviations from the mean). A further example is the standard error of the mean (the standard deviation of the means of an infinite number of data samples composing the population - in effect, an expression of the relationship between the standard deviation of the sample and the standard deviation of the population).

Related tools: Descriptive statistics

Measures of central tendency and variability are often used to describe the trends and characteristics of numerical data in conjunction with graphical means. Sample data may be arranged into classes or groups to illustrate the frequency of occurrence of data elements and the underlying distribution patterns of data.

In general, histograms and polygons are used to describe data which has been grouped into frequency, relative frequency and percentage distributions.

Related tools: Frequency distribution histograms

2. Inferential statistics

This form of analysis is based on making general conclusions and reasonable estimates of characteristics of a potentially large data set(s) which are derived from the characteristics of smaller (and usually more practical to quantify) sample data sets. It is this branch of statistics which is concerned with hypothesis testing and probability theory.

Numerical data is usually collected after consideration and formation of a research hypothesis. A hypothesis can be described as a suggested explanation of certain observed phenomena, tentatively adopted until it can be tested experimentally. For example, consider the following research hypothesis:

> "There are differences in the distribution and abundance of mole-hills at different grassland sites and this is likely to have been caused by differences in the mowing regimes applied".

In order to test the validity of this hypothesis, numerical data must be obtained from each grassland site using a sampling methodology. As the sites concerned may be very large it would be futile to attempt a count of all mole-hills. Numerical data is therefore collected from random sections of the sites. The quantity of all mole-hills at a given site forms what is known as a statistical population. Data derived from sections of the site forms what is known as a statistical sample of the population.

It is important to realise that when random samples are selected to represent the statistical population there will invariably be an element of error present. This should be minimized as much as possible but can be manifest as 'errors of measurement' or as 'sampling error'. In the latter case this simply means that the sample data may not be very representative of the true population data. In most scientific investigations 'sampling error' may be reduced by increasing the sample size.

Considerations in sampling

There are many types of sampling methodology which vary according to the requirements of the investigation. Detailed discussion is beyond the scope of this guide but all effective sampling methods should conform to the following essential requirements:

(a.) They must be repeatable (i.e., the technique can be standardised).

(b.) They should be as simple as possible so that errors introduced through application of the sampling technique are reduced to a minimum.

(c.) The sampling method and the frequency and size of the samples must be closely related to the method of statistical analysis. A well planned investigation should contain the simultaneous consideration of sampling procedure and subsequent handling of the data obtained.

(d.) It is important that relevant questions are applied to particular problems. The sample data collected to answer these problems should also be relevant. If the sampling is wrong, or the approach is wrong, then the results will also be wrong, leading to misinterpretation. On this note, there is a familiar saying: "garbage in; garbage out"!

Using sample data to test the hypothesis

Probability is an important component of any form of inferential statistical hypothesis test. In the various hypothesis tests that may be applied (dependent on the type of data and the question to be answered) the hypothesis is neither proved nor disproved to be correct. The test invariably produces a test statistic (a numerical value of t, x^2, F, etc.) which is used to determine the probability that the hypothesis is correct or not.

In this process, whereby a hypothesis is accepted or rejected, convention dictates that a probability level of 95% is usually used to determine whether acceptance or rejection is required. In other words analysis of sample data must indicate that there is at least 95% probability that the hypothesis is correct for it to be retained.

This method of testing the validity of hypotheses has important implications for science in general. In effect, scientists do not generally set out to ’prove’ that underlying mechanisms of particular observed phenomena are true. It is more accurate to state that that they make decisions about the probability of such mechanisms being operative based on available quantitative data.

Null hypotheses

Once a research hypothesis has been formulated it must be modified if it is to be tested for validity using inferential statistics. In other words, the research hypothesis must be re-stated in statistical terms as a null hypothesis (abbreviated as HO) and retained in its original form as the alternate hypothesis (abbreviated as HA). Consider the example research hypothesis provided earlier:

Null hypothesis:

> "Mowing frequency has no effect on mole-hill distribution and abundance and any observed differences of mole-hill distribution and abundance between the grassland sites is a result of chance sampling".

Alternate hypothesis:

> "There are differences in the distribution and abundance of mole-hills at different grassland sites and this is likely to have been caused by differences in the mowing regimes applied".

The title ’null hypothesis’ is so named because the hypothesis states that there is no difference between the sample data from each site, that they come from the same statistical population, and that any non-significant difference between them is due to sampling error. The reason for formulating a null hypothesis is that statistical hypothesis tests will provide a probability value that allows it to be rejected or accepted. If the null hypothesis is rejected the alternate hypothesis can then be tentatively accepted.

Null hypotheses and alternate hypotheses should be mutually exclusive and exhaustive. In other words, there must be no option for both to be true and there should be no possibility that some other unspecified alternate hypothesis is true.

Steps in hypothesis testing:

1. State the null hypothesis (H0).

2. State the alternative hypothesis (Ha).

3. Choose the level of significance (i.e., 0.05, 0.01, etc.).
4. Choose the sample size (n).
5. Determine the appropriate statistical technique and the corresponding statistical test to use.
6. Collect the data and set up the critical values that divide the rejection and non-rejection regions. This is performed by the included tools.
7. Calculate the sample value of the appropriate test statistic. This is also performed by the included tools.
8. Determine whether the test statistic has fallen into the rejection or the non-rejection region.
9. Make the statistical decision.
10. Express the statistical decision in terms of the problem.

Case example

In order to test the null hypothesis for the mole-hills at different grassland sites 20 x 4m^2 quadrats (a square frame) were randomly positioned at each of the sites and the quantity of mole-hills in each quadrat was recorded. The Student's t-test was chosen as a hypothesis test to determine the probability of a statistically significant difference in the means of samples of mole-hill populations at each grassland site. As the raw sample data was observed to exhibit a normal distribution and an F-test established that the variance of each sample was equivalent, it was decided to employ the ordinary t-test based on two independent sample means. The t-tests were carried out between pairs of samples (i.e., between samples taken from site A and site B, site A and C, and site B and C). The main characteristics of each site is shown below:

	Site A	Site B	Site C
Mowing regime:	Mown weekly	Mown annually	Never mown

For each t-test, calculations were made of the value of t (the test statistic), the t-critical value (at the 0.05 level of significance) and the probability figure that the samples were derived from the same statistical population.

	Site A & B	Site A & C	Site B & C
t-statistic:	1.74	4.706	3.17
t-critical:	2.02	2.059	2.04
P:	0.08	0.0000797	0.003

Test between samples from site A and B:

The value of the t-critical statistic represents a separation figure between rejection and non-rejection of the null hypothesis. In this test the t-statistic was found to be lower than the critical value at the 0.05 level of significance. This arbitrary level of significance should always be stated in results and represents the fact that the probability of making an error in deciding to retain or reject a null hypothesis is no higher than 5%. In brief, a null hypothesis is usually rejected if the t-statistic exceeds the critical value at a given level of significance (i.e., 0.05, 0.01, etc.)

In the case of the hypothesis test between the sample data from site A and site B the test statistics favoured a retention of the null hypothesis. Observation of the probability value shows that there was a greater than 5% probability that the sample means were derived from the same statistical population.

There is a good reason for the choice of very low levels of significance (i.e., 0.05, 0.01, etc.). A widely used analogy in explaining the concept of a level of significance is the justice system in courts of law. Usually, a defendant is presumed innocent until proven guilty beyond a reasonable doubt. The prosecution is required to prove the defendant guilty and it is deemed preferable to free a guilty person rather than imprisoning an innocent one. In a statistical sense this is similar to saying that it is preferable to accept a null hypothesis that is actually false (known as a Type II error) than to reject a null hypothesis that is actually true (known as a Type I error). The ’reasonable doubt’ in a hypothesis test is represented by the significance level and this allows little margin for error (i.e., no higher than 5% or 1%, etc.) in rejecting a null hypothesis.

Test between samples from sites A and C and sites B and C:

In both of these tests the t-statistic was found to be greater than the critical value at the 0.05 level of significance and the null hypothesis would be rejected in both cases. In other words, the results are statistically significant and a decision would then be made that the respective sample means in each test are highly likely to be from different statistical populations. This is reflected in the probability calculations: there was a less than 5% probability that this was the case.

The next logical step is the drawing of conclusions. As significant differences in sample means existed between sites A and C and sites B and C, but not between sites A and B, it may be considered that mowing caused an increase in mole-hill distribution and abundance but the frequency of mowing had little effect on this increase. However, it is important to remember that other unconsidered factors may have caused observed differences such as site exposure, geographical location, laziness of moles (!), etc.

This case example is meant to illustrate the main underlying mechanisms by which hypothesis tests function. Although in this case a relatively simple two-sample t-test is used, consideration of a one-way analysis of variance may be preferable if this was a real scenario.

Parametric and non-parametric tests

The t-test outlined above is a form of parametric statistical test and there are several others of this nature included within this package. The common basis of such tests is that they require a number of assumptions about one or more population parameters. The most common assumption when determining the nature of a statistical population from the characteristics of a sample is that it has a normal distribution. In addition, when comparisons of a sample parameter (i.e., the mean or variance, etc.) are made between two or more samples it is usually assumed that there is homogeneity of variances.

Non-parametric tests, or distribution-free tests, make less stringent assumptions about the form of the underlying population distribution. In general, a given parametric test will have an equivalent non-parametric version but it is preferable to use the former where possible as it is usually more powerful. There is a greater risk of committing a Type II error with non-parametric methods. There are several circumstances under which a non-parametric test should be generally employed:

(a.) If the sample data is on an ordinal scale it cannot be normally distributed and should be subjected to non-parametric methods. In other words, the data takes the form of being relative rather than quantitative and is ordered or ranked.

(b.) If the underlying population distribution is unknown or has not been specified or is known not to follow a normal distribution. For example, the distribution may be significantly skewed or be bimodal and even multimodal.

Related tools:

Descriptive statistics: Determination of modality and distribution curve shape, etc.

Goodness of fit (x^2) for normality: Detect normality using chi-square goodness of fit test.

(c.) If in a two-sample test there is a significant difference between sample variances.

Related tools:

F-Ratio: Variance ratio test.

One-tailed and two-tailed tests

In the t-test example above, no specifications were provided as to whether we were interested in detecting any significant difference in the means of the samples or whether we were interested in one sample mean being significantly greater or smaller than another.

In a two-tailed test (also known as a two-sided or nondirectional test) the object of the test is to determine whether there is any significant difference between two or more samples or one sample and a hypothesised parameter (i.e., the mean, variance, etc.). For example, in the t-test outlined above we could be trying to detect whether there was any significant difference in the sample means. As this method employs probability in the form of the t-distribution curve, and we are not investigating whether one particular sample mean is larger/smaller than the other in a particular direction, the rejection region is divided into two tails of the t-distribution. Any calculated t-critical value will also be accompanied by an equivalent negative t-critical value because an extreme, or very improbable value of 't' in any direction will cause the rejection of the null hypothesis.

In a one-tailed test (also known as a one-sided or directional test) the objective is determination of a significant difference between samples or parameters of samples in one direction only. As a result we are only interested in whether there is a small (i.e., 5%, 1%, etc.) probability of the test statistic occurring by chance alone in one tail of the given distribution (i.e., t, x^2, F, etc.). The one-tailed test is often preferable to the two-tailed test, if a direction may be specified in the alternate hypothesis, as it provides a higher chance (or power) of rejecting the null hypothesis.

3. Statistical relationships between variables

A further branch of statistical analysis is concerned with examining the relationships that may exist between two quantitative variables represented in sampling efforts.

Correlation analysis

Investigation of the degree of correlation between two variables effectively determines the strength of association between them. There are two commonly used correlation techniques. Spearman Rank correlation coefficient is calculated for data variables that are ranked or do not exhibit a normal distribution, and is therefore a non-parametric form of analysis. The Pearson product-moment correlation coefficient is more sensitive and therefore preferable but assumes that sample variables are normally distributed. Is there any statistically significant association between number of late hours spent writing AmigaGuides and number of accidental hard drive invalidating incidents per month?(!)

Regression analysis

Regression analysis is concerned with making predictions of the values of one variable (known as the dependent variable 'y') based upon the values of another variable (known as the independent variable 'x'). It is effectively concerned with investigation of the nature of the relationship that may exist between two or more variables. A typical application of such procedure may be to predict agricultural crop yields on the basis of quantities of fertilizer applied to a field.

Related tools:

Pearson Product Moment correlation: Bivariate correlation analysis.

Spearman Rank correlation: Non-parametric correlation.

Linear regression: Simple linear regression analysis.

1.4 Statistical analysis tools included...

· Details of the statistical analysis tools:
~~~~~~

This section contains details of specific use and workings of the statistical tools included in this package. All worked examples found in sections below and the scripts themselves will assume (where necessary) that statistical tests are conducted at the 0.05 or 0.01 level of probability or significance. Most of the information found in these sections will assume some prior knowledge of statistical analysis theory and application. If required, click here for further information. You may also want to pursue further references for information beyond the scope of this document.

Descriptive or summary statistics

Test: Descriptive statistics for sample data.
Tool: For description and operation select here.
Script: Descriptive_Stats.rexx

Test: Frequency distribution histogram.
Tool: for description and operation select here.
Script: Histogram.rexx

Test: Relative rank and percentile calculation.
Tool: For description and operation select here.
Script: Rank&Percentile.rexx

Parametric statistical tests

Test: T-test of two independent sample means.
Tool: For description and operation select here.
Script: T-test_IndSamples.rexx
~~~~~~

```
Test: T-test of two independent means-with unequal variances.
Tool: For description and operation select here.
Script: T-test_IndSamples_UneqVar.rexx

Test: T-test of two paired or correlated sample means.
Tool: For description and operation select here.
Script: T-test_Corr_Sample.rexx

Test: F-ratio test for homogeneity of sample variances.
Tool: For description and operation select here.
Script: F_Ratio.rexx

Test: One-way analysis of variance.
Tool: For description and operation select here.
Script: anova_oneway.rexx

Test: Two-way analysis of variance.
Tool: For description and operation select here.
Script: ANOVA_TwoWay.rexx

Test: Two-way analysis of variance without replication.
Tool: For description and operation select here.
Script: ANOVA_TWOWAY_NO_REP.rexx

Test: Pearson's product moment correlation between two variables.
Tool: For description and operation select here.
Script: Correlation.rexx
```

Non-parametric statistical tests

```
Test: Mann-Whitney U-test.
Tool: For description and operation select here.
Script: Mann-Whitney.rexx

Test: Smirnov Test on Two Independent Samples.
Tool: For description and operation select here.
Script: Smirnov_2W.rexx

Test: Cochrans Test for Related Observations.
Tool: For description and operation select here.
Script: Cochran.rexx

Test: Wilcoxon paired sample or sign rank test.
Tool: For description and operation select here.
Script: Wilcoxon_Sign_Rank.rexx

Test: Kruskal-Wallis H test.
Tool: For description and operation select here.
Script: Kruskal-Wallis.rexx

Test: Chi-square (x$^2$) sign test of two independent samples.
Tool: For description and operation select here.
Script: Chi-Sq_2_Ind_Sam.rexx

Test: Chi-square (x$^2$) test for independence of sample data.
Tool: For description and operation select here.
```

Script: Chi-Sq_Indep.rexx

Test: Chi-square (x^2) goodness of fit test.
Tool: For description and operation select here.
Script: Chi-Sq_GO_Fit.rexx

Test: "Normality" - Chi-square (x^2) GOF test for sample normal distribution.
Tool: For description and operation select here.
Script: Normality.rexx

Test: Shapiro-Wilk test for normality
Tool: For description and operation select here.
Script: Shapiro.rexx

Test: Fisher exact test
Tool: For description and operation select here.
Script: Fisher_Exact.rexx

Test: Friedman analysis of variance by rank test.
Tool: For description and operation select here.
Script: Friedman_Rank.rexx

Test: Spearman’s rank correlation between two variables.
Tool: For description and operation select here.
Script: Spearmans_Rho.rexx

Test: Kendall’s rank correlation between two variables.
Tool: For description and operation select here.
Script: Kendalls_Tau.rexx

Test: Durbin Test - Balanced Incomplete Block Design.
Tool: For description and operation select here.
Script: Durbin.rexx

Further statistical tools

Test: Linear regression analysis.
Tool: For description and operation select here.
Script: Regression.rexx

Test: Non-linear regression analysis.
Tool: For description and operation select here.
Script: Exponential_Smooth.rexx

Test: Moving average.
Tool: For description and operation select here.
Script: Moving_Average.rexx

Test: Covariance of 2 random variables.
Tool: For description and operation select here.
Script: Covariance.rexx

Test: Shannon-Wiener biodiversity estimation and t-test.
Tool: For description and operation select here.
Script: shannon-weiner.rexx

Test: Random Normal Deviates.
Tool: For description and operation select here.
Script: Deviates.rexx

1.5 Description and script operation...

- Descriptive and summary statistics.
  ~~~~~~

The output generated by this tool provides a variety of measures of central tendency and data dispersion for each column of sample data. This is advantageous before conducting further analysis in the form of hypothesis tests. The overall nature of the sample data gathered from this output may be influential in choice of subsequent parametric or non-parametric tests. For further details follow this linked section on statistical analysis.

- Script operation.
  ~~~~~~~~~~~~~~~~

This tool operates in much the same way as most of the others with no specific departures from the usual methods needed.

Click here for information about general script usage.

Raw sample data may be entered as single or multiple samples, the data being arranged in columns. Typical output is shown below. Note that the statistics are computed for the three example columns and these have been labelled in the output under the headings ’Group A’, ’Group B’ and ’Group C’. Notice how the original samples were labelled in the same manner and that these titles have been incorporated in the output by inclusion in the input data range.

Raw data:

```
Group A   Group B   Group C
     22       6.2        18
     15       5.0        16
      9       8.9        31
      7       7.8         8
      4       6.4         2
     45      11.4        36
     19       5.8        12
     26       6.2        16
     35       7.1        47
     49      10.4        22
        9.7        37
        8.7        52
        6.4
        7.7
        7.8
        7.4
```

```
Spreadsheet output:

Statistics       Group A     Group B          Group C
Count:          10          16               12
Sum:           231       122.9             297
Mean(Arith.):       23.1      7.6813           24.75
Mean(Geo.):            17.7873      7.4992          18.8119
Mean(Harm.):           12.8471      7.3259          11.3004
Mean(Quad.):           27.4645      7.8688          28.9698
Mode:          None  Multi_(See_below)       16
First Quartile:        9          6.4              16
Median:        20.5        7.55              20
Third Quartile:       35          8.9              37
Range:          45          6.4              50
Maximum:              49        11.4             52
Minimum:               4     5         2
Std. Error:             4.9519      0.4409          4.5396
Std. Deviation:        15.6592      1.7638         15.7256
Mean Deviation:    12.52      1.3688          13.2083
Variance:             245.2111       3.111        247.2955
Skewness:               0.4588      0.5742           0.355
Kurtosis:              -1.0607     -0.4642          -1.016
Confidence Level (95%)-low:  13.3943       6.817         15.8524
Confidence Level (95%)-high: 32.8057      8.5455         33.6476
Confidence Level (99%)-low:  10.3242      6.5436         13.0378
Confidence Level (99%)-high: 35.8758      8.8189         36.4622

              Mode  Count
               6.2      2
               6.4      2
               7.8      2
```

- Interpretation.
  ~~~~~~~~~~~~~~

There is much useful information contained within the output and the following points are particularly worth noting:

- Several mean values are calculated for each sample. Of these, the arithmetic mean is the usual, most widely used value required. The others have specialised applications (i.e., the geometric mean is used in averaging ratios and % change and is inappropriate if any data element is a negative value).

- In calculation of the median (the middle measurement in a set of data) the sample is divided into two groups. The sample may be further fractionally split into four equal parts, represented by the upper (3rd) and lower (1st) quartiles.

- In this particular example output there appears to be a significant difference in the variances of particular samples and there are also differences in modality. At first glance it would appear that increased sample sizes are necessary. If in this particular case this was not possible or if the samples still exhibited the same characteristics after further sampling, then the sample data could be assumed to violate the assumptions required by parametric testing.
  ~~~~~~~~~~~~~~

In this case it is preferable to employ a non-parametric testing approach.

Related tools:

Goodness of fit (x^2) for normality: Detect normality using chi-square goodness of fit test.

F-Ratio test: Variance ratio test.

- Skewness (the measure of degree of asymmetry in data distributions) and kurtosis (the shape of the data distribution) are also calculated. Neither calculations have units of measurement but their relationship to zero is important:

  In general, if the mean and median of sample data are equal the data distribution may be considered to be symmetrical, or zero-skewed. If the mean exceeds the median then positive, or a right-skewed distribution is evident (i.e., the distribution 'tails off' towards the right-hand side). If the median exceeds the mean then a negative or left-skewed distribution is evident. In the example above it is seen that samples 'Group A' and 'Group B' have a slightly positive skewed distribution and sample 'Group C' has a slightly negatively skewed distribution. A score of zero would indicate a symmetrical distribution. It is possible to test for a statistically significant departure from zero skewness using a parametric t-test but this is at present not undertaken.

  Although sample data distributions may exhibit symmetry their shape, or kurtosis may be measured in a similar way to skewness. A kurtosis score of zero is indicative of a mesokurtic or normal distribution. A leptokurtic distribution is represented by a positive score above zero and is characterised by many data elements clustered around the mean and in the 'tails' of the distribution. A platykurtic distribution is represented by a negative score less than zero (as in all samples 'Groups A, B, and C' above) and is characterised by many data elements in between the mean and the 'tails' of the distribution.

  Related tools:

  Goodness of fit (x^2) for normality: Detect normality using chi-square goodness of fit test.

  Frequency histograms: Analyse data frequency distributions.

1.6 Description and script operation...

· Descriptive and summary statistics - histograms.
~~~~~~

In order to gain an indication of sample data distribution patterns the data may be reorganised.
~~~~~~

Sample data is often reorganised to create a frequency distribution table. In other words the raw data is split into groups, or classes, based on size.

Resulting size classes enable observation of the frequency with which data elements fall into specific numerical ranges within the overall range of the data. It may then be possible to see whether or not the data is clustered in any sub-range of the overall data range or dispersed throughout the overall range.

· Script operation.
~~~~~~~~~~~~~~~~

This tool operates in a slightly different way to most others and this subsequently requires that the instructions outlined below are followed.

Click here for information about general script usage.

The Histogram.rexx script analyses a single column of data to generate frequencies, cumulative frequencies and cumulative percentages. This allows you to see the distribution of the data set before embarking on more elaborate statistical procedures.

NOTE: Only one column of data is allowed.

The program will request you to nominate the start cell of a range of "bin" values (ie. class interval values). If no range is nominated, the program will create its own bin values. The following example illustrates the use of bin values supplied by the user and also those calculated by the program automatically (the second set of outputs).

```
Raw data:        Spreadsheet output 1:

Exam scores   Bin values     Exam scores

   42           40      Bin   Frequency   Cum.Freq.     Cum.%
   45           50
   46           60       40           6     6     31.58
   46           70       50           3     9     47.37
   42           80       60           2    11     57.89
   55           90       70           2    13     68.42
   60          100       80           3    16     84.21
   94        90            2    18     94.74
   86             100           1    19    100.00
   72
   64
   59
   52            Spreadsheet output 2:
   44
   86
   99            Exam scores
  100
   77             Bin   Frequency   Cum.Freq.     Cum.%
   84
       42           7     7     36.84
```
~~~~~~~~~~~~~~~~

```
53.6        4    11     57.89
65.2        1    12     63.16
76.8        4    16     84.21
88.4        3    19    100.00
```

· Interpretation.
~~~~~~~~~~~~~~

It is important to realise that frequency distributions and histograms are simply a visual aid to gain an insight into trends in the sample data frequency distribution. If it is necessary to test data for the presence of a normal distribution, before choosing subsequent parametric hypothesis testing methods the use of this histogram output data may be useful to an extent.
A more powerful way of testing for normality can be gained by applying a x^2 goodness of fit hypothesis test which compares the observed data distribution with that of the expected normal distribution.

Related tools:

Goodness of fit (x^2) for normality: Detect normality using chi-square goodness of fit test.

Several useful assumptions may be made from the data generated from the analysis tool:

The easiest way to gain familiarity with any trends in the sample data distribution is to plot the distributions generated using TurboCalc’s graphing capabilities. A simple line graph gives the clearest insight.

Click here to view a plot of the frequency data displayed above. Note that in this case the data seems very unlikely to follow a normal distribution and appears to be distinctly positively skewed.

Click here to view a plot of the cumulative percentage frequency data for the example above. If the plot had resembled a sigmoid (S) curve it would be expected that the sample data followed a normal distribution. In this case the shape of the curve produced by the plot resembles the upper half of a sigmoid curve, reinforcing the likelihood that the sample data is from a positively skewed distribution. This in turn seems to imply that the exam was a particularly difficult one or the students were mostly a bit daft or that the tutor was useless. Take your pick...

1.7 Description and script operation...

· Rank and percentile conversion of sample data.
~~~~~~

Sample data may often be converted so that it is represented in an alternative format, without actually altering the relationship of one data element to the others. This modified data can then be used for specific further analysis.

Data is usually ranked, or arranged in order so that rank ’1’ is assigned to either the lowest or highest data element and subsequent rankings are assigned to the other data elements on the basis of magnitude. The ranking of data often forms the basis of many non-parametric hypothesis tests. Data converted to ranks using this tool may also be used for non-parametric hypothesis tests of a specialised nature which are beyond the scope of this package contents.

When summarizing or describing the properties of sample data the data itself may be represented by measures of non-central location known as quantiles. Typical measures of central location include the median and mean. Typical quantile measures include quartiles and percentiles. For example, the median splits ordered sample data in half, quartiles split it into four equal portions and percentiles split it into hundredths. Measures of non-central location have limited applications but may be used to locate the position of single data elements and their relationship to others (eg. finding the position of a student’s score among a large data set of examination scores, etc.).

· Script operation.
~~~~~~~~~~~~~~~~

This tool operates in much the same way as most of the others with no specific departures from the usual methods needed.

Click here for information about general script usage.

Typical input data and script output are shown in the example below. Note that any column titles or labels may be included in the output data by being included in the input range. This tool will accept single or multiple columns of sample data.

Raw data:

Sample1	Sample2
60	75
65	89
90	93
45	66
66	78
66	84
89	97
92	100
75	75

Spreadsheet output:

Sample1	Rank	Percentile Rank	Percentile Point	Sample2	Rank	Percentile Rank	Percentile Point
45	9	6	0.0	66	9	6	0.0
60	8	17	12.5	75	7	22	12.5
65	7	28	25.0	75	7	22	12.5
66	5	44	37.5	78	6	39	37.5
66	5	44	37.5	84	5	50	50.0
75	4	61	62.5	89	4	61	62.5
89	3	72	75.0	93	3	72	75.0
~~~~~~~~~~~~~~~~

```
    90     2    83        87.5       97     2           83        87.5
    92     1    94       100.0      100     1           94       100.0
```

1.8 Description and script operation...

```
· The t-test of two independent sample means.
  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
 In common with other t-tests, this tool is based on detection of
 statistically significant differences in the mean values obtained from two
 data samples.

 A very common form of t-test that should not be employed if there are
 significant differences in the variance of both samples. In common with
 many other t-tests it also assumes that the samples are drawn from a
 normally distributed population.

 The alternative non-parametric equivalent of this test is the Mann-Whitney
 test.


· Script operation.
  ~~~~~~~~~~~~~~~~
 This tool operates in much the same way as most of the others with no
 specific departures from the usual methods needed.

 Click here for information about general script usage.

 Click here for the formula used in this t-test.

 Raw sample data is entered in two columns as shown. Note that in this
 example summary statistics (i.e., the count (or 'n'), the sum, and the
 mean, etc.) are computed for the two columns and these have been labelled
 in the output under 'Group A' and 'Group B'. If sample data columns
 contain titles this can be reproduced in the output by including these in
 the input data range.

 Raw data:             Spreadsheet output:

 Group A   Group B     T-Test: Two Means for Independent Samples
             Group A       Group B
      16  20      Count:              8       6
       9   5      Sum:               88      48
       4   1      Mean:              11       8
      23  16
      19   2      Pooled variance:    60.1667
      10   4      Std. Deviation:      7.7567
       5          Std. Error:          4.1891
       2          t:              0.7161
          d.f.:              12
          P(T<=t) one-tail:  0.243803
          T-Critical (95%):    1.7823
          T-Critical (99%):     2.681
          P(T<=t) two-tail:  0.487606
          T-Critical (95%):    2.1788
```

```
          T-Critical (99%):    3.0545
```

· Interpretation.
~~~~~~~~~~~~~~
Using the example above, the results are provided for both one-tailed and two-tailed tests at both the 95% and 99% levels of probability or significance.

The null hypothesis for a two-tailed test may take the form of:

"There is no statistically significant difference in the means of the two samples and they are derived from the same data population. Any observed difference is the product of chance sampling".

This null hypothesis should be rejected when the value of ’t’ determined is seen to exceed the critical value for a given one-tailed or two-tailed test at a given probability level.
i.e.,

As the t-distribution curve is symmetrical about the mean the rejection region for a two-tailed test at P=0.05 (95% level) falls beyond +2.1788 and -2.1788. This is not exceeded by the test statistic value of ’t’ (0.7161) and therefore the null hypothesis should be retained at this level.

For both the one-tailed test and two-tailed tests the exact probability of obtaining the computed t-statistic is presented in the output.

1.9 Description and script operation...

· The t-test of two independent sample means-assuming unequal variances.
~~~~~~
This is similar to the usual t-test but is designed for testing samples which are normally distributed but with statistically significant differences in variances. This is effectively a compromise, indicating that the t-test is fairly ’robust’ when its requirements (i.e., normal distribution and equal variances) are not entirely met. There is a greater danger of falsely rejecting or retaining a null hypothesis with this method, although the chance of this can be reduced by using large sample sizes with equal amounts of data in each.

The comparison of means of two normally distributed samples with unequal variances is known as the ’Behrens-Fisher problem’. Although several solutions exist, the script included is one of the most reliable ways of undertaking such procedures and produces what is known as ’Welch’s approximate t’. An associated approximated degrees of freedom is also determined.

· Script operation.
~~~~~~~~~~~~~~~~
This tool operates in much the same way as most of the others with no specific departures from the usual methods needed.

Click here for information about general script usage.
~~~~~~~~~~~~~~~~

Clich here for the formula used by this t-test.

Raw sample data is entered in two columns as shown. Note that in this example summary statistics (i.e., the count (or 'n'), the sum, and the mean) are computed for the two columns and these have been labelled in the output under the headings 'Column 1' and 'Column 2'. Notice how the original samples were labelled as 'Sample A' and 'Sample B' but that these titles were not incorporated in the output by inclusion in the input data range.

```
Raw data:         Spreadsheet output:

SampleA  SampleB  T-Test:Two Means for Independent Samples:Unequal Variances
     74       66           Column 1   Column 2
     69       62  Count:          8          8
     70       63  Sum:          557        499
     71       64  Mean:      69.625     62.375
     73       65
     68       61  Standard Error:        1.8504
     71       63  Welchs approx. t:       3.918
     61       55  First d.f.:      7
      Second d.f.:     7
      Approx. d.f.:      13.62
      P(T<=t) one-tail:     0.000883
      T-Critical (95%):       1.7709
      T-Critical (99%):       2.6503
      P(T<=t) two-tail:     0.001765
      T-Critical (95%):       2.1604
      T-Critical (99%):       3.0123
```

· Interpretation.
~~~~~~~~~~~~~~

Using the example above, the results are provided for both one-tailed and two-tailed tests at both the 95% and 99% levels of probability or significance.

The null hypothesis for a two-tailed test may take the form of:

"There is no statistically significant difference in the means of the two samples and they are derived from the same data population. Any observed difference is the product of chance sampling".

This null hypothesis should be rejected when the value of 't' determined is seen to exceed the critical value for a given one-tailed or two-tailed test at a given probability level.
i.e.,

As the t-distribution curve is symmetrical about the mean the rejection region for a two-tailed test at P=0.05 (95% level) falls beyond +2.1604 and -2.1604. The test statistic value of 't' (3.918) is seen to exceed the critical value and the null hypothesis should therefore be rejected at this level.

For both the one-tailed test and two-tailed tests the exact probability of obtaining the computed t-statistic is presented in the output.
~~~~~~~~~~~~~~

Note that if use of statistical distribution tables is being employed, the approximation that assigns a number of degrees of freedom to 't' does so in order that the ordinary t-distribution table may be used. Although the degrees of freedom may be non-integer this must be rounded to the nearest integer when entering the t-distribution table.

1.10 Description and script operation...

· The t-test of two paired or correlated sample means.
~~~~~~

A further commonly used form of t-test but should be used with caution. It is used when two samples are directly associated or occur in pairs.

For example, a biologist may want to determine whether there is a significant difference between foreleg and hindleg length in cats. Although these leg lengths may well differ the analysis is complicated by the fact that cats are of different sizes. Whilst we have different different cat sizes, the length of the foreleg in any one cat is proportional to that of the hindleg in the same cat and the data for each can therefore be associated in pairs for each animal.

Although this type of t-test is very powerful it should not be used in preference to the others unless you are sure that your two samples have a direct association. For example, you would not normally use this test if you were attempting to detect any statistical significant difference between the mean concentration of a particular chemical in streams at two sites. In this case, one of the other t-tests would be used, after testing the samples for normal distributions and homogeneity of variances.

This test does not deal with the original data measurements but rather the differences within each pair of measurements taken from the samples. For this reason, this test does not require samples with normal distributions or equality of variances. However, normality must be observed in the population of differences.

The non-parametric equivalent is the Wilcoxon paired sample test.

· Script operation.
~~~~~~~~~~~~~~~~

This tool operates in much the same way as most of the others with no specific departures from the usual methods needed.

Click here for information about general script usage.

Click here for the formula used in this t-test.

Raw sample data is entered in two columns as shown. Note that this tool will not accept sample data with unequal sample sizes. Follow the worked example below to observe how summary statistics are computed for the two samples. Although the difference values for each data pair are calculated they are not printed to the spreadsheet except in the form of the mean value. Note that any raw data labels included within the input range are not reproduced in the output because this deals with statistics based on difference values only.

```
Raw data:        Spreadsheet output:

Stress  Non-stress     T-Test: Two Means for Correlated Samples
     7     5
     9    15     Mean of Diff.:        -1.3
     4     7     Variance:          12.2333
    15    11     St.Dev.:      3.4976
     6     4     St.Err.:       1.106
     3     7     t:           -1.1754
     9     8     Count:            10
     5    10     d.f.:              9
     6     6     P(T<=t) one-tail:       0.135004
    12    16     T-Critical (95%):         1.8331
     T-Critical (99%):         2.8214
     P(T<=t) two-tail:       0.270007
     T-Critical (95%):         2.2622
     T-Critical (99%):         3.2498
```

· Interpretation.
~~~~~~~~~~~~~~~

In the matched-pairs or correlated sample t-test the hypotheses in the standard two-tailed test may take the form of:

HO: mean difference = 0; HA: mean difference is not = 0

Therefore retain HO. At the 0.05 level of significance (95% probability) the t-critical value of +2.2622 or -2.2622 is not exceeded by the value of the t-statistic. There is no significant difference between the two samples.
i.e.,

```
     t:     -1.1754
     d.f.:        9
     P:       >0.05   (i.e., 0.270007)
```

For both the one-tailed test and two-tailed tests the exact probability of obtaining the computed t-statistic is presented in the output.

1.11 Description and script operation...

· F-test: Testing for difference between two sample variances.
~~~~~~

One of the requirements of adopting a parametric testing procedure for determining differences in a parameter of two or more samples is that the samples have equivalent variances (s^2). Although any two sample variances are extremely unlikely to be identical, the usual approach is to determine whether there is any statistically significant difference at a given level of confidence or probability.

If it is found that two sample variances are significantly different then it is preferable to employ a non-parametric hypothesis test. In some cases it may be possible to transform the original sample data by calculating log values and this may reduce the inequality in variance of two samples.

The other frequent requirement of parametric testing methods is that the sample data follows a normal distribution.

Related tool:

Goodness of fit (x^2) for normality: Detect normality using chi-square goodness of fit test.

• Script operation.
~~~~~~~~~~~~~~~~

This tool operates in much the same way as most of the others with no specific departures from the usual methods needed.

Click here for information about general script usage.

Click here for the formula used by this F-test.

Raw sample data is entered in two columns as shown. Note that this tool will also accept sample data with unequal sample sizes. Follow the worked example below to observe how summary statistics are computed for the two samples. Titles used for the raw data samples may be transferred and printed within the output by inclusion in the input range.

```
Raw data:            Spreadsheet output:

Sample I   Sample II       F-test:Two-sample for variances
      64          26         Sample_I   Sample_II
      52          35       Count:              9     9
      48          34       Mean:          50.2222    31
      52          32       Variance:      49.6944       10.75
      43          34       d.f.(v):            8     8
      44          28
      46          29       F Ratio:        4.6227
      58          28       P(F<=f) one-tail:   0.022173
      45          33       F-Critical (95%):     3.4381
         F-Critical (99%):     6.0289
         P(F<=f) two-tail:   0.044346
         F-Critical (95%):     4.4333
         F-Critical (99%):     7.4959
```

• Interpretation.
~~~~~~~~~~~~~~

If a two-sample F-test is being conducted simply as a precursor to making a decision as to whether to embark on parametric testing procedures, then the null hypothesis is likely to take the following form:

HO: The variance of the 1st sample [s^2(1)] is not significantly different from the variance of the 2nd sample [s^2(2)].

As we are not interested in whether one variance is larger or smaller than the other but just whether significant difference exists, this can be defined as a two-tailed hypothesis test. Under all other circumstances of the F-test (i.e., ANOVA) the test is a one-tail variety.

In the example output above, it can be seen that at the 0.05 level of significance the F-critical value of 4.4333 is exceeded by the computed F-statistic of 4.6227 and that this therefore implies that the null hypothesis should be rejected. Note that in this particular case the null hypothesis would be retained at the much more stringent 0.01 level of significance.

1.12 Description and script operation...

· Single factor or one-way analysis of variance (ANOVA).
~~~~~~

The analysis of variance (often termed ANOVA or AOV) is a technique used to test multi-sample hypotheses whereby a variable (the mean) is measured from three or more samples.

In cases where multi-samples are to be tested, series of two-sample tests (i.e., t-tests, etc.) should not be employed. Aside from the fact that such an approach would be very time consuming, it may also be statistically invalid. For example, testing at the 0.05 level of significance, if three sample means are compared two at a time using the two-sample t-test, then there is a 13% chance of committing a Type I error (i.e., one in which the null hypothesis is rejected when it is actually true). As the number of samples increase the chance becomes greater (i.e., 63% when comparing ten sample means and 92% when comparing twenty sample means).

The single factor analysis of variance test should be applied when a test for the effect of only one factor on the variable in question is required. For example, an investigation may need to determine whether four different fertilizers result in different heights of a plant. In this case the single factor is fertilizer type and the variable is plant height. Each type of fertilizer is termed a level of the factor.

Two approaches may be employed which involve identical calculations but differ in construction of the null and alternate hypotheses. In the first case, known as a Model I or fixed effects ANOVA, the levels of the factor are specifically chosen. In the example mentioned above the fertilizer treatments have been selected to form levels of the factor and the null hypothesis would take the form of:

HO: $\mathrm{\mu}$(1) = $\mathrm{\mu}$(2) = $\mathrm{\mu}$(3) = $\mathrm{\mu}$ ↩
(4) ...where $\mathrm{\mu}$ represents the population mean.

In the second approach, known as a Model II or random effects ANOVA, the levels of the factor may be randomly selected to represent a random sample of a population. For example, in the case mentioned above we may be interested in whether there is great variability in the heights of the specific type of plant irrespective of fertilizer treatment. In this case the locations or sites of plant growth may be sampled. If specific sites with particular characteristics are chosen then a Model I ANOVA may be employed - the sites, or factorial levels may be influential in plant height attained. If the intention is to generalize so that all sites or locations are considered to comprise a random sample from all possible
~~~~~~

sites then a Model II ANOVA may be employed. In this case, the null hypothesis would simply be:

HO: There is no variability in plant height at different locations.

The parametric analysis of variance technique makes certain assumptions about the sample data used. Sample data used should follow a normal distribution and exhibit homogeneity of variances. However, some test 'robustness'is present. Tests will not be too adversely affected by small departures from normality particularly if the sample sizes are large. Similarly, differences in the variance between each group may be accomodated as long as sample sizes are the same.

The altenative non-parametric equivalent is the Kruskal-Wallis single factor analysis of variance by ranks.

- Script operation.
~~~~~~~~~~~~~~~~

This tool operates in much the same way as most of the others with no specific departures from the usual methods needed.

Click here for information about general script usage.

Raw sample data must be entered as multiple samples, the data being arranged in columns. Follow the example output below and note that the statistics are computed for the four example columns and these have been labelled in the output under the headings 'Method I', 'Method II', etc. Notice how the original samples were labelled in the same manner and that these titles have been incorporated in the output by inclusion in the input data range.

Note that in this example summary statistics (i.e., the count (or 'n'), the sum, the mean and the variance) are computed for the columns.

Raw data:

```
Method I   Method II   Method III    Method IV
 5           9      8            1
 7          11      6            3
 6           8      9            4
 3           7      5            5
 9           7      7            1
 7            4             4
 4            4
 2
```

Spreadsheet output:

ANOVA - One way

```
Group          Count        Total      Mean       Variance
Method_1           8           43    5.375         5.4107
Method_II          5           42      8.4     2.8
```
~~~~~~~~~~~~~~~~

```
Method_III         7           43   6.1429        3.8095
Method_IV          6           18        3     2.8

ANOVA
Source of Variation     SS      d.f.(v)  Variance Est.       F-Ratio

Between Groups:    82.2217           3  27.4072        7.0167
Within Groups:     85.9321          22    3.906
Total:      168.1538           25

P(F<=f) one-tail:  0.00175
F-Critical (95%):   3.0488
F-Critical (99%):   4.8167
```

· Interpretation.
~~~~~~~~~~~~~~

The example test outlined above may be considered to have been made after formulation of a null hypothesis such as:

HO: Weights of cakes do not differ when baked using four different methods.

In effect this ANOVA test would then be a random effects model in this case. The statistics produced in the output are known as sources of variation:

The between groups SS (sum of squares) and within groups SS are used in conjunction with their respective degrees of freedom to produce two variance estimates known as MS (an abbreviation of mean squared deviations from the mean). These are the between groups MS (often known as just groups MS) and the within groups MS (often termed the error MS) respectively.

If the null hypothesis is correct then both MS (or variance estimates) will be an estimate of the variance common to all four statistical populations. If the four population means are not equal then the between groups MS will be greater than the within groups MS. This can be tested by a simple one-tailed variance ratio test:

F = Between groups MS / Within groups MS

In this case:

F = 27.4072 / 3.9060 = 7.0167

In this example case, the critical value at the 0.05 level of significance can be seen to be 3.0488 and this is exceeded by the F value of 7.0167. In other words, the null hypothesis must be rejected and it can be stated that population means of the weights of cakes are significantly different dependent on baking method used.

1.13 Description and script operation...
~~~~~~~~~~~~~~

· Two way analysis of variance with equal replication (ANOVA).
~~~~~~

The analysis of variance (often termed ANOVA or AOV) is a technique used to test multi-sample hypotheses whereby a variable (the mean) is measured from three or more samples.

In cases where multi-samples are to be tested, series of two-sample tests (i.e., t-tests, etc.) should not be employed. Aside from the fact that such an approach would be very time consuming, it may also be statistically invalid. For example, testing at the 0.05 level of significance, if three sample means are compared two at a time using the two-sample t-test, then there is a 13% chance of committing a Type I error (i.e., one in which the null hypothesis is rejected when it is actually true). As the number of samples increase the chance becomes greater (i.e., 63% when comparing ten sample means and 92% when comparing twenty sample means).

The two-factor (or two-way) analysis of variance is a means of conducting simultaneous analysis of the effect of two factors on a population variable. For example, an investigation may need to determine whether fertilizer and temperature differences result in different heights of a plant. In this case the first factor is fertilizer application, the second factor is ambient temperature and the variable is plant height. As an example, in an experimental situation the application or non-application of fertilizer and the temperature under which different samples of plants are grown are known as levels of the factor.

The advantage of factorial analysis of variance (i.e., analysis of the effect of more than one factor on population means) is that it is more economical. Use of this form of analysis makes it unnecessary to conduct a one-way analysis of variance for each factor in an investigation. It also enables testing for interactions between factors.

In an ANOVA test with equal replication there are equal numbers of measurements of the variable in question in each factorial level.
This parametric analysis of variance technique makes certain assumptions about the sample data used. Sample data used should follow a normal distribution and exhibit homogeneity of variances.

· Script operation.
~~~~~~~~~~~~~~~~

This tool operates in a slightly different way to others. Enter sample data in the way shown below. After entering the data into the input requestor, a further requestor will demand the number of rows in each sample. In other words the script demands the amount of replicate results. In the example data below this would be 5.

** This means you must select a 3 column range! **

Click here for information about general script usage.

```
Raw data:            Spreadsheet output:


 Hormone treatmnt:   ANOVA - Two Way: With Replication
```

```
    FEMALE  MALE
NO    16.5  14.5    Summary Statistics
      18.4    11
      12.7  10.8           FEMALE     MALE     Total
  14  14.3
      12.8    10    NO
YES   39.1    32    Count:            5   5        10
      26.2  23.8    Sum:           74.4      60.6       135
      21.3  28.8    Mean:         14.88     12.12      13.5
      35.8    25    Variance:     6.217     4.477    6.8689
      40.2  29.3

        YES
        Count:            5   5        10
        Sum:           162.6     138.9     301.5
        Mean:          32.52     27.78     30.15
        Variance:     69.717    11.182   42.1961


        Col. Total

        Count:           10  10
        Sum:             237     199.5
        Mean:           23.7     19.95
        Variance:   120.1844   75.0806

        ANOVA

        Source of        Sum of  Degrees of   Variance  F Ratio
        Variation       Squares     Freedom   Estimate
        Sample:       1386.1125           1  1386.1125  60.5336
        Column:         70.3125           1    70.3125   3.0706
        Interaction:     4.9005           1     4.9005    0.214
        Within Cells:   366.372          16    22.8983
        Total:        1827.6976          19

        P(F Sample <=f) one-tail:       0.000001
        F-Critical (95%):            4.494
        F-Critical (99%):           8.5309
        P(F Column <=f) one-tail:       0.098852
        F-Critical (95%):            4.494
        F-Critical (99%):           8.5309
        P(F Interaction <=f) one-tail:         0.64987
        F-Critical (95%):            4.494
        F-Critical (99%):           8.5309
```

· Interpretation.
~~~~~~~~~~~~~~

In the example detailed above the variable being examined may be considered to be blood calcium concentration and the the two factors thought to be influencing this are sex and an unspecified hormone treatment. Note that both factors have two levels (i.e., male/female and treatment/no treatment) and that there are 5 replicate results (i.e., blood calcium concentration in each factorial level of 5 animals).

In this example case the null hypotheses may take the following form:
~~~~~~~~~~~~~~

1. Hormone treatment has no effect on the mean blood calcium concentration.

2. There is no difference in the mean blood calcium concentration between males and females.

3. There is no influence of both hormone treatment and sex on the mean blood calcium concentration.

The output produced by the script has two components. In the first block above the title 'ANOVA' summary statistics are calculated for each factorial level and these are then summed across the rows and down the columns. The second block contains the ANOVA sources of variation.

The 'Sample'(representing the hormone factor) sum of squares (or SS), the 'Column' (representing the sex factor) SS, and the 'Interaction' (representing the combined factors) SS are used in conjunction with their respective degrees of freedom to produce three variance estimates known as MS (an abbreviation of mean squared deviations from the mean). There is also a further MS value calculated known as the 'Within Cells' MS, or error MS.

If the means of the factorial levels are not equal, and the null hypotheses must be rejected, then each MS value will be greater than the within cells MS. This can be tested by a simple one-tailed variance ratio test:

Hypothesis 1:

F = Sample MS / Within cells MS

In this case:

F = 1386.1125 / 22.8983 = 60.5336

In order to test the significance of the calculated values of F, the probability of obtaining these statistics and the critical values at the 0.05 and 0.01 levels of significance are provided in the output.

In the case of hypothesis No.1, the critical value at the 0.05 level of significance may be seen to be 4.494 and this is exceeded by the F value of 60.5336. In other words, the null hypothesis must be rejected and it can be stated that there is a statistically significant difference between blood calcium concentration dependent on hormone treatment.

If this process is continued for the further remaining two hypotheses it can be seen that both are retained at the 0.05 level of significance and the overall conclusion would be that hormone treatment has a highly significant effect on blood calcium concentration but that there is no significantly different blood calcium concentration between males and females. Furthermore, in relation to hypothesis No.3, there is no significant interaction between hormone treatment and sex - the effect of the hormone treatment in terms of blood calcium concentration is no different in males and females.

1.14 Description and script operation...

· Two-way analysis of variance with no replication (ANOVA).
~~~~~~

The non-parametric equivalent to this test is the Friedman ANOVA by ranks test.

· Script operation.
~~~~~~~~~~~~~~~~

This tool operates in a slightly different way to most of the others. Sample data should be entered as multiple samples but the column titled 'Subject' in the example below must also be included within the input range. The title of this column may be anything you wish. In the example below, the column headers 'A', 'B', etc. should also be included in the input range. The reason for this is that this information represents the two factors within the analysis.

Click here for information about general script usage.

Raw data:

Subject	A	B	C	D
1	31	42	14	80
2	42	26	25	106
3	84	21	19	83
4	26	60	36	69
5	14	35	44	48
6	16	80	28	76
7	29	49	80	39
8	32	38	76	84
9	45	65	15	91
10	30	71	82	39

Spreadsheet output:

ANOVA-Two Way:No Replication

Subject	Count	Total	Mean	Variance
1	4	167	41.75	782.9167
2	4	199	49.75	1466.9167
3	4	207	51.75	1344.9167
4	4	191	47.75	404.25
5	4	141	35.25	230.25
6	4	200	50	1072
7	4	197	49.25	486.9167
8	4	230	57.5	691.6667
9	4	216	54	1030.6667
10	4	222	55.5	621.6667
A	10	349	34.9	390.9889
B	10	487	48.7	388.9
C	10	419	41.9	751.8778
D	10	715	71.5	513.6111

ANOVA

```
Source of    Sum of    Degrees of    Variance        F Ratio
Variation    Squares   Freedom       Estimate

Rows      1565      9    173.8889         0.2787
Columns      7553.1      3      2517.7         4.0359
Remainder    16843.4     27    623.8296
Total        25961.5     39

P(F Column <=f) one-tail:            0.017098
F-Critical (95%):          2.9606
F-Critical (99%):          4.6011
```

· Interpretation.
~~~~~~~~~~~~~~

1.15 Description and script operation...

· Pearson's product-moment correlation coefficient.
~~~~~~
Investigation of the degree of correlation between two variables in effect determines the strength of association between them. There are two commonly used correlation techniques. Spearman Rank correlation coefficient is calculated for data variables that are ranked and do not exhibit a normal distribution, and is therefore a non-parametric form of analysis.

Related tool:

The non-parametric equivalent to this test: Spearman Rank or rho test.

The Pearson product-moment correlation coefficient is more sensitive and therefore preferable but assumes that sample variables are normally distributed. The statistic calculated (usually defined as 'r') is independent of the raw data units and can only fall within the range of -1 and +1.

· Script operation.
~~~~~~~~~~~~~~~~
This tool operates in much the same way as most of the others with only one specific departure from the usual methods needed. After the data input and script output range has been defined a further requestor will appear demanding the hypothesized value of the true population correlation coefficient. By default this is set as '0' and for the majority of cases can be left as this. This parameter can be changed (within the range of -1 to +1) but the considerations in doing so will be discussed later.

Click here for information about general script usage.

This test requires the input of equal sample sizes. Note that the sample labels 'x' and 'y' have been incorporated into the output by inclusion within the data range. Note also that the hypothesized population correlation coefficient was left as the default '0'.
~~~~~~~~~~~~~~~~

```
Raw data:          Spreadsheet output:

       Correlation: UnGrouped Data
   x       y
   5       1            x       y
  10       6      x          1         0.575
   5       2      y       0.575       1
  11       8
  12       5      Pearson r:    0.575
   4       1      r sq.:            0.3307
   3       4      Std. Error of r:   0.2893
   2       6      n:         10
   7       5      d.f.:            8
   1       2      t:           1.9880
       P(T<=t) one-tail:  0.0410
       T-Critical (95%):  1.8595
       T-Critical (99%):  2.8965
       P(T<=t) two-tail:  0.0820
       T-Critical (95%):   2.306
       T-Critical (99%):  3.3554
       Confidence Intervals:
       95% (+1.96):       1.3958
       95% (-1.96):      -0.0858
       99% (+2.58):       1.6301
       99% (-2.58):      -0.3201
```

· Interpretation.
~~~~~~~~~~~~~~

In the example above the correlation coefficient calculated (0.575) indicates that there is an intermediate degree of positive correlation between the two data variables.

If the correlation coefficient is -1, correlation is known as perfect and negative; a value of +1 means perfect positive correlation; a value of 0 means that there is no correlation.

Although we can see that the positive correlation between the two example variables is not particularly strong it is orthodox practice to express this result in the usual statistical terms of probability. There are two ways in which this may be accomplished.

The first method, based on determination of the significance of the 'r' statistic obtained, will require access to a table of significance levels of the value of 'r' for this particular sample size. This table is widely available in the appendices of most text books.

In the example above, r=0.575 based upon 10 pairs of observations. This means that there are n-2 degrees of freedom (i.e., 10-2=8 d.f.) and this is used to enter the table. Reference to the correlation coefficient table indicates that with d.f.=8 the value of 0.575 is less than the tabulated critical value of 0.632 found at the 0.05 level of significance.
In other words as P>0.05, we are less than 95% certain that there is statistically significant correlation between the two variables and the null hypothesis that there is no significant correlation must be retained.

In the second method the Student's t-test may be used. The correlation
~~~~~~~~~~~~~~

coefficient (r) is calculated from a pair of samples and is taken to be an estimate of the correlation present in the population data that was sampled. Where the population correlation is concerned the coefficient is termed 'p'. In this case the null hypothesis may take the form of:

HO: p = 0 and therefore HA: p is not equal to 0.

As with the first method this is again a two-tailed test as we are not particularly interested in directional change, only whether there is significant correlation in the form of a coefficient significantly different from 0 (which is effectively no correlation).

In the two-tail output above it can be seen that the calculated value of 't' (1.9880) does not exceed the t-critical value at the 0.05 level of significance (2.306). As P>0.05 there is less than 95% certainty of statistically significant correlation between the two variables and again the null hypothesis must be retained.

Note that in the second method if it is hypothesized that 'p' is anything other than '0' then the t-test is unsuitable. In this case 'r' must be converted to a 'z' value using Fisher's z-transformation. If you alter the default hypothesized population correlation coefficient of '0' when the appropriate requestor appears then a Fisher z-transformation will be carried out and the results will be output to the spreadsheet.

1.16 Description and script operation...

- Mann-Whitney U-test
  ~~~~~~~~~~~~~~~~~~~

This non-parametric test, as with many others, examines sample data only after it has been transformed into ranked data. It also makes no estimation or hypothesis about population parameters (i.e., the mean, etc.) and does not require the raw data to be derived from a normal distribution. This test is based on detection of statistically significant differences in two independent samples.

The alternative parametric equivalent of this test is the ordinary t-test of two independent sample means.

- Script operation.
  ~~~~~~~~~~~~~~~~

This tool operates in much the same way as most of the others with no specific departures from the usual methods needed.

Click here for information about general script usage.

Raw sample data is entered in two columns as shown. Note that in this example summary statistics (i.e., the count (or 'n'), the sorted data and the ranked sorted data) are computed for the two columns and these have been labelled in the output under 'Morning' and 'Afternoon'. If sample data columns contain titles this can be reproduced in the output by including these in the input data range.

Note in the example that all sample data is output to the spreadsheet after being both sorted on the basis of size and corresponding rank.

```
Raw data:          Spreadsheet output:

 Morning   Afternoon     Mann-Whitney U Test
      66    53     Sorted Data
      70    81          Morning    Afternoon
      73    83         66           53
      75    84         70           81
      75    84         73           83
      79    84         75           84
      82    85         75           84
      87    86         79           84
      95    88         82           85
      90          87            86
      91          95            88
      92                90
                 91
                 92

       Table of Ranks for Sorted Data
             2            1
             3            8
             4           10
           5.5           12
           5.5           12
             7           12
             9           14
            16           15
            21           17
                 18
                 19
                 20


       U(Sample 1):            80
       U(Sample 2):            28
       No. of Cases:           21
       n(Sample 1):             9
       n(Sample 2):            12
       Normal Approximation:
       Mean:             54
       Variance:              198
       St.Dev.:          14.0712
       z:          -1.8477
       P(z to left tail):       0.03232
       P(z to mean):     0.46768
       P(z to right tail):      0.96768
       P Density Function:       0.0724
       Z-Critical One-tail(95%):  1.65
       Z-Critical One-tail(99%):  2.33
       Z-Critical Two-tail(95%):  1.96
       Z-Critical Two-tail(99%):  2.58
```

· Interpretation.

~~~~~~~~~~~~~~

There are two ways in which the output from the script may be interpretated for the purpose of drawing conclusions. These are based on the two sample sizes used.

If the size of the smaller sample is <20 and the larger sample size is <40 then the distribution of U can be used in conjunction with the U statistics calculated, in order to determine the statistical significance of the result. If the sample sizes do not fall within these categories and are subsequently larger, then the normal approximation method may be used. When larger sample sizes are available the distribution of U approaches the normal distribution. In effect this then means that a value of z may be calculated and compared with critical values of the t-distribution, with degrees of freedom of infinity, which is identical to the normal distribution.

In the example above, the following two-tailed null hypothesis may be proposed:

HO: Morning and afternoon figures are the same, or have no significant difference.

The first method: using the U distribution.

Access is required to a table of critical values of the Mann-Whitney U distribution (widely available in most text books) in order to ascertain whether a statistically significant result has been obtained. Using such a table, the larger of the two U statistics determined by the analysis tool is compared with the critical value at a given level of significance (i.e., 0.05 or 0.01 etc.). In order to do this, the table must be entered using the n(Sample 1) and n(Sample 2) values provided in the output. There are two ways in which you may proceed:

If n(Sample 1) < n(Sample 2) then the table should be entered using the value of n(Sample 1) first and then followed by the value of n(Sample 2).

If n(Sample 1) > n(Sample 2) then the table should be entered using the value of n(Sample 2) first, to be followed by the value of n(Sample 1).

In the example output above it can be seen that the former method should be employed as 9 < 12 and that the critical value derived from a table for U(0.05)(2-tail)(9,12) is 82.

As the U-critical value (82) is not exceeded by the largest calculated value of U (80) then the null hypothesis should be retained at the 0.05 level of significance.

The second method: using the normal approximation.

As the t-distribution (with v=infinity) is identical to the normal distribution, the z-critical value is equal to the t-critical value.
As the value of z is computed (i.e., -1.8477) it can then be compared with the two-tailed t-critical value at a given level of significance in order to determine the validity of the null hypothesis.
~~~~~~~~~~~~~~

At the 0.05 level of significance the t-critical value with v=infinity is 1.96. As this is not exceeded by the computed value of z this also indicates that the null hypothesis should be retained.

1.17 Description and script operation...

- Smirnov Test on Two Independent Samples.
  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

This test is useful where two samples are drawn, one from each of two populations (which could be different), and we wish to know whether the the two distribution functions associated with the populations are identical or not.

Both the t-test (parametric) and the Mann-Whitney test are sensitive to differences between the two means or medians but may not be sensitive to other differences, eg. variances. The Smirnov test is sensitive to all types of differences between the two distribution functions.

The data consists of two independent continuous random samples that we assume are mutually independent, and that have a measurement scale which is at least ↩ ordinal.
(If the random variables are discrete instead of continuous, the test is still valid but becomes conservative).

There are three test statistics based on the two distribution functions (S1(x) and S2(y)):

```
Two-sided T = sup|S1(x)-S2(y)|
One-sided T1= sup[S1(x)-S2(y)]
One-sided T2= sup[S2(y)-S1(x)]
```

Hypotheses:

```
(a) Two-sided:   Ho: F(x)=G(x) for all x
   H1: F(x)~=G(x) for at least one value of x

(b) One-sided: Ho: F(x)<=G(x) for all x
   H1: F(x)>G(x) for at least one value of x

(c) One-sided: Ho: F(x)>=G(x) for all x
   H1: F(x)<G(x) for at least one value of x
```

- Script operation.
  ~~~~~~~~~~~~~~~~

This tool operates in much the same way as most of the others with no specific departures from the usual methods needed.

Click here for information about general script usage.

In this example, two random samples of size 9 and 15 are tested.

```
X   Y
7.6  5.2
```

```
8.4  5.7
8.6  5.9
8.7  6.5
9.3  6.8
9.9  8.2
10.1 9.1
10.6 9.8
11.2 10.8
 11.3
 11.5
 12.3
 12.5
 13.4
 14.6
```

```
Spreadsheet output:

 Smirnov Two Sample Test

 S1(x)  S2(x) S1(x)-S2(x) S2(x)-S1(x) Smirnov T
 0  0.0667  -0.0667   0.0667    (Two-Sided)
 0  0.1333  -0.1333   0.1333    0.4
 0  0.2 -0.2    0.2   (One-Sided:S1>S2)
 0  0.2667  -0.2667   0.2667    0.4
 0  0.3333  -0.3333   0.3333    (One-Sided:S2>S1)
 0.1111 0.3333  -0.2222   0.2222    0.333333
 0.1111 0.4 -0.2889   0.2889
 0.2222 0.4 -0.1778   0.1778
 0.3333 0.4 -0.0667   0.0667
 0.4444 0.4 0.0444    -0.0444
 0.4444 0.4667  -0.0222   0.0222
 0.5556 0.4667  0.0889    -0.0889
 0.5556 0.5333  0.0222    -0.0222
 0.6667 0.5333  0.1333    -0.1333
 0.7778 0.5333  0.2444    -0.2444
 0.8889 0.5333  0.3556    -0.3556
 0.8889 0.6 0.2889    -0.2889
 1  0.6 0.4   -0.4
 1  0.6667  0.3333    -0.3333
 1  0.7333  0.2667    -0.2667
 1  0.8 0.2   -0.2
 1  0.8667  0.1333    -0.1333
 1  0.9333  0.0667    -0.0667
 1  1 0   0

· Interpretation.
  ~~~~~~~~~~~~~~
 Specific tables need to be consulted that give quantiles of the Smirnov test
 statistic for two samples to assess the output. For n=9 and m=15 with a
 two-sided test, the table indicates that W0.95=0.5333, therefore Ho is accepted
 at the 95% level.
```

1.18 Description and script operation...

- Cochran's test for related observations.
  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

We often have treatments or tests that result in one of two possible outcomes eg. yes/no; success/failure.

The Cochran test is used where every test or treatment is applied independently to every block (or subject) and the results are compared with each other. Thus each of "c" treatments is applied independetly to each of "r" blocks and the result is recorded as either a 1 or a zero. The results are presented in the form of a table with "r" rows and "c" columns.

This is called a "randomised complete block design".

The important assumptions are that each subject or block is selected randomly from the population, and that treatment outcomes can by presented as either 1 or zero.

The test assumes a large sample of blocks, so that the null distribution is the chi-squared distribution with c-1 degrees of freedom.

Hypotheses:

Ho: The treatments are equally effective.
H1: There is a difference in effectiveness between treatments.

The test statistic "T" is computed and the Ho hypothesis is rejected if T is greater than the (1 - level of significance) quantile of the chi-squared distribution with c-1 degrees of freedom. "P" is the probability of a chi-squared random variable with c-1 degrees of freedom exceeding the observed value of "T".

- Script operation.
  ~~~~~~~~~~~~~~~~

This tool operates in much the same way as most of the others with no specific departures from the usual methods needed. The data set can include the column titles for the calculation although they are discarded by the script.

Click here for information about general script usage.

In the following table we have 3 columns representing the outcome of the choices of three persons concerning the outcome of a randomly selected series of questions. If they answered correctly they received a "1"; if not, they received a "0".

Outcome of Answers

```
Person A B C
 1 1 1
 1 1 1
```

```
  0 1 0
  1 1 0
  0 0 0
  1 1 1
  1 1 1
  1 1 0
  0 0 1
  0 1 0
  1 1 1
  1 1 1

Spreadsheet output:

 Cochrans Test for Related Observations

 Cochrans T:  2.8
 df(v):   2
 P(Chi>T):  0.24659701
 Chi-Critical (95%):  5.9915
 Chi-Critical (99%):  9.2103

· Interpretation.
  ~~~~~~~~~~~~~~~
 In this example, Ho is accepted as the "T" value of 2.8 is less than the .05
 (95%) critical region of rejection (with 2 degrees of freedom).
```

1.19 Description and script operation...

```
· Wilcoxon paired sample or signed-ranks test.
  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
 This tool is based on detection of statistically significant differences in
 two paired data samples. If sample data is paired then this test should be
 used in preference to the Mann-Whitney U test when choosing a
 non-parametric test as the latter may result in a higher probability of
 commiting a Type II error (i.e., failure to reject a false null hypothesis).

 This test does not deal with the original data measurements but rather the
 differences within each pair of measurements taken from the samples. It has
 an underlying assumption that the data population represented by the
 samples is symmetrical around the median value.

 The parametric equivalent is the Matched-pair t-test test.


· Script operation.
  ~~~~~~~~~~~~~~~~
 This tool operates in much the same way as most of the others with no
 specific departures from the usual methods needed.

 Click here for information about general script usage.

 Raw sample data is entered in two columns as shown. If sample data columns
 contain titles this can be reproduced in the output by including these in
 the input data range.
```

Note in the example that all sample data is output to the spreadsheet after being both sorted on the basis of difference, absolute values and corresponding rank.

```
Raw data:           Spreadsheet output:

    X      Y        Wilcoxon Matched-Pairs Signed Ranks Test
   15     19
   19     30         (X-Y)      Sorted ABS Values     Rank
   31     26
   36      8           -4              0       --
   10     10          -11              2      1.5
   11      6      5               2      1.5
   19     17            28              4       -3
   15     13      0               5      4.5
   10     22      5               5      4.5
   16      8      2               8        6
      2              11       -7
          -12              12       -8
      8              28        9

        Sum of Neg. Ranks:          18
        Sum of Pos. Ranks:          27
        Wilcoxon T:           18
        Ranked Obs.(n):              9
        z:          -0.5331
        Prob. (Z<=z):         0.296977
```

· Interpretation.
~~~~~~~~~~~~~~

There are two ways in which the output from the script may be interpretated for the purpose of drawing conclusions. These are based on the two sample sizes used. Either method may be used for the same purpose but the second, normal approximation method, is often used where raw data consists of more than 100 pairs (i.e., beyond the limits of a normal Wilcoxon T-distribution critical values table).

When larger sample sizes are available the Wilcoxon T-distribution approaches the normal distribution. In effect this then means that a standardized value of z may be calculated and compared with critical values of the t-distribution, with degrees of freedom of infinity, which is identical to the normal distribution. The reason for this, although beyond the scope of this AmigaGuide, can be found in many text books under the title 'central limit theorem'.

In the example above, the following two-tailed null hypothesis may be proposed:

HO: 'X' and 'Y' sample data is drawn from the same statistical population,
    or have no significant difference.

The first method: using the Wilcoxon T-distribution.

Access is required to a table of critical values of the Wilcoxon T-distribution found in most text books. In the example case printed above
~~~~~~~~~~~~~~

the table would be entered with n=9 (labelled as ’Ranked Obs.(n)’ in the output). At the 0.05 level of significance the T-critical value may be found to be 5 for a two-tailed test. The calculated value of T can be seen to be 18 in the output. In most other test circumstances the null hypothesis is rejected if the calculated test statistic is found to exceed the critical value at a given level of confidence. In this case the null hypothesis is retained unless the T-statistic is less than or equal to the critical value.

Note that if one-tail tests are being analysed using this method, then the following mechanisms apply:

1. If HO: Sample 1 measurements < or = Sample 2 measurements, then
   Reject HO if ’Sum of Neg. Ranks’ < or = T(0.05)(1-tail)(d.f.=9).

2. If HO: Sample 1 measurements > or = Sample 2 measurements, then
   reject HO if ’Sum of Pos. Ranks’ < or = T(0.05)(1-tail)(d.f.=9).

In the output the ’Sum of Neg. Ranks’ and the ’Sum of Pos. Ranks’ are often also referred to as ’T-’ and ’T+’ respectively.

The second method: using the normal approximation.

As the t-distribution (with v=infinity) is identical to the normal distribution, the z-critical value is equal to the t-critical value.
As the value of z is computed (i.e., -0.5331) it can then be compared with the two-tailed t-critical value at a given level of significance in order to determine the validity of the null hypothesis.

At the 0.05 level of significance the t-critical value with v=infinity is 1.9600. As this is not exceeded by the computed value of z this also indicates that the null hypothesis should be retained.

1.20 Description and script operation...

· Kruskal-Wallis single factor analysis of variance by ranks.
~~~~~~

The analysis of variance (often termed ANOVA or AOV) is a technique used to test multi-sample hypotheses whereby a variable (the mean) is measured from three or more samples. In this case, inter-group differences in data are subject to a non-parametric test whereby no population parameters or sample statistics are stated or used in the test calculations.

The parametric equivalent to this test is the one-way ANOVA test.

As may be expected from a non-parametric test, the sample data is not required to be from mormal distributions and the sample variances may be heterogeneous. As with the parametric ANOVA test equivalent, rejection of a null hypothesis in such a test will only provide an indication that there is at least one difference between the groups (i.e., samples, or ’k’) but no information about which groups differ from which other groups.

· Script operation.
~~~~~~

~~~~~~~~~~~~~~~~

This tool operates in much the same way as most of the others with no specific departures from the usual methods needed.

Click here for information about general script usage.

Raw sample data must be entered as multiple samples, the data being arranged in columns. Follow the example shown below and note that the statistics are computed for the five example columns and these have been labelled in the output under the headings 'A', 'B', 'C', etc.
Notice how the original samples were labelled in the same manner and that these titles have been incorporated in the output by inclusion in the input data range.

Note in the example that all sample data is output to the spreadsheet after being both sorted on the basis of size and corresponding rank.

```
Raw data:       Spreadsheet output:

  A   B   C   D   E    Kruskal-Wallis H Test
 68  72  60  48  64    Sorted Data
 72  53  82  61  65         A       B       C       D       E
 77  63  64  57  70        42      48      60      48      53
 42  53  75  64  68        53      53      64      50      64
 53  48  72  50  53        68      53      72      57      65
         72      63      75      61      68
         77      72      82      64      70

           Table of Ranks for Sorted Data
          1     2.5      10     2.5     6.5
        6.5     6.5      14       4      14
             17.5     6.5      21       9      16
         21      12      23      11    17.5
         24      21      25      14      19


           H Statistic:            6.44
           No. of Cases:       25
           d.f.(v):       4
           Sample Size allows Chi-Square Test:
           P(x$^2$<=H):     0.83139
           Chi-Critical(95%):    9.4877
           Chi-Critical(99%):   13.2767
```

· Interpretation.

~~~~~~~~~~~~~~

There are two common ways in which the results of this test may be interpretated. Both function by comparison of the H statistic generated with critical values derived from distribution tables. If the number of samples is less than five then the the H statistic is compared, at a given level of significance, with a critical value derived from a table of such values of the H distribution. If there are more than five samples being compared for significant differences, or the sample sizes are fairly large then the H statistic may be considered to be approximated by the the chi-square (x^2) distribution with k - 1 degrees of freedom (where k is the number of samples).

The first method will require access to a table of H distributions published in most statistical analysis texts.

In order to illustrate this process, the comparison of the H statistic with critical values of the x^2 distribution will be employed as this is likely to be the most useful and more widely employed method. In the example provided, the H statistic is determined as 6.44 and the approximated degrees of freedom is 4 (i.e., k - 1 = 5 - 1 = 4).
In the output (above) it can be seen that the critical value at the 0.05 level of significance is 9.4877. As the critical value is not exceeded by the test statistic value the null hypothesis that there is no significant difference between the samples may be retained.

1.21 Description and script operation...

· Chi-squared (x^2) sign test of two independent samples.
~~~~~~

This test is also known as the "median test" as it compares the medians of two independent samples.

The null hypothesis is that there is no difference between the medians of the populations from which the samples were drawn.

The data consist of two independent samples with N1 and N2 observations. The median of the combined observations is calculated, and in each sample observations above and below the combined median are assigned either a "+" or a "-". The number of "positive" and "negative" signs is calculated and the Chi-square test is used to determine whether the observed frequencies of + and - signs depart significantly from the null hypothesis.

· Script operation.
~~~~~~~~~~~~~~~~

This tool operates in much the same way as most of the others.

Click here for information about general script usage.

Raw sample data must be entered as multiple samples, the data being arranged in columns. Note that sample titles may be included within the output by inclusion within the input range.

```
Raw data:           Spreadsheet output:

 Sample I   Sample II      Chi-Square Non-Parametric Sign Test
       10           6      of Two Independent Samples
       10           7             Positive  Negative
       10           8      Sample_I            7         5
       12           8      Sample_II           3         6
       15          12
       17          16      Median:            16
       17          19      Count:             21
       19          19      Chi-Square:    0.4813
       20          22      d.f.:         1
       22            P(CHI<=chi):       0.512143
       25            Chi-Critical(95%)    3.8431
```

```
      26           Chi-Critical(99%)     6.637
```

· Interpretation.
~~~~~~~~~~~~~~
The calculated Chi-square probability of 0.512 includes the Yates correction for continuity. This value is below the value of 3.84 required for significance at the 5% level indicating that we have no grounds for rejecting the null hypothesis.

1.22 Description and operation...

· Chi-square (x^2) test for independence.
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
The chi-square test for independence, or contingency table test, is a non-parametric test which uses the x^2 distribution to determine whether there is statistical independence between two variables or whether a relationship between them exists. The x^2 statistic is calculated by comparison of observed data values of two categories of two variables with the corresponding expected values which should occur if a null hypothesis is true. By reference to the input data contingency table (below) it can be seen that a typical null hypothesis would take the form of:

HO: The survival of a plant is independent of whether a drug treatment has been administered.

and the alternate hypothesis would take the form of:

HA: The survival of the plants is associated with the administration of the drug treatment.

The raw observed data is arranged in a contingency table such as the one below. Note that this accomodates the testing of the independence of the row data with that of the columns, and vice versa.

· Script operation.
~~~~~~~~~~~~~~~~
The tool operates in a similar way to others with the exception that it will incorporate both row and column headings if they are included in the data input range. Once the input range has been entered, a requestor will ask you if the range includes row headings to the left-hand side of the data.

Click here for information about general script usage.

Click here for the formulae used by this test.

Note that this tool currently deals only with 2x2 contingency tables.
~~~~~~~~~~~~~~~~

```
Raw cont. table data:   Spreadsheet output:

       Dead  Alive      Chi-square test for independence
 Treated   9     15
 Not treated    15     10      Expected Frequencies

                Dead     Alive
             Treated           11.7551   12.2449
             Not treated             12.2449   12.7551

             Chi-square:        2.4806
             d.f.:             1
             P(CHI<=chi):       0.8847
             Chi-Critical(95%):       3.8431
             Chi-Critical(99%):        6.637
             With Yates correction
             Chi-square:         1.662
             P(CHI<=chi):       0.8027
             Measures of Association:
             Phi.Coeff.:         0.225
             Yules Q:          -0.4286
             Odds Ratio:           0.9
             Pearson Cont.Coeff.:     0.2195
             Cramers Phi.Coeff.:       0.225
```

· Interpretation.
~~~~~~~~~~~~~~

In the first section of the output the expected frequencies are computed and printed for each of the original observed data elements. These are the expected frequencies if the null hypothesis is true.

Two chi-square values are calculated: one is a straightforward calculation and the other is modified after the application of Yate's continuity correction. In the case of a 2x2 contingency table where the degrees of freedom is always '1', the corrected chi-square value should always be used as it provides a more conservative estimate of x^2. In chi-square analysis the result, in the form of the x^2 statistic, is only an approximation to the theoretical distribution. At any other value of d.f. it is a reasonably accurate approximation and the Yate's correction factor can be ignored (it has little effect with larger d.f. and large sample sizes anyway).

Before analyzing the output it is worthwhile noting another proviso of the chi-square analysis method. It should not be used if any of the expected frequencies are seen to be less than '5'. This can be remedied by increasing the sample size or by using an alternative test that determines direct probability values:

Related tool:

Fisher Exact test

From the output the x^2 statistic (corrected) can be seen to be 1.662 and the critical value at the 0.05 level of significance can be seen to be 3.8431. If testing at this level the null hypothesis would therefore be retained as the critical value is not exceeded by the test statistic. It
~~~~~~~~~~~~~~

```
would then be concluded that there is no statistically significant
association between drug administration and survival.
```

1.23 Description and operation...

```
· Chi-square (x$^2$) goodness of fit analysis.
  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
 Goodness of fit (GOF) analysis is concerned with testing a hypothesis that
 sampled data may be randomly drawn from a theoretical distribution or that
 a random variable has a specific theoretical probability distribution.
 The data for this test consist of N independent observations of a
 random variable which are grouped into c classes. The measurement
 scale must be at least nominal. If the number of observations in any one cell
 is very low, the observations may be combined into adjacent cells.

· Script operation.
  ~~~~~~~~~~~~~~~~
 This tool operates in much the same way as most of the others but with
 specific departures from the usual methods. The script requires that the
 Observed frequencies be in column 1, while the second column contains either ↩
    Expected
 frequencies or the Probabilities. The script will ask the user to indicate which
 type of data is in column 2.

 Click here to examine the Chi-square script that allows the
 user to work with the computer to group raw data for analysis.

 Note that if Probabilities are used, the total of all probabilities must add
 up to either 1 or 100, or the script will detect an error condition.
 The script also allows the user to vary the degrees of freedom used in the ↩
    calculations.
 A requester will open allowing the user to indicate a number to be subtracted ↩
    from the
 calculated df (which is the number of rows -1). The script will not allow less ↩
    than 1
 degree of freedom to be used!

 Click here for information about general script usage.

 The script operates on raw data which is arranged in columns. Note that sample ↩
    titles
 may be included within the input range but will not be used on output.
 The following shows the same observed data in two examples, one using
 Expected frequencies and the other using Probabilities. The output is the same
 for both examples.

 Example 1
 Cell Observed F  Expected F
 1  20    20
 2  14    20
 3  18    20
 4  17    20
 5  22    20
 6  29    20
```

```
Example 2
Cell Observed F  Probability
1  20    0.1666666666667
2  14    0.1666666666667
3  18    0.1666666666667
4  17    0.1666666666667
5  22    0.1666666666667
6  29    0.1666666666667
```

Spreadsheet output:

```
Chi-Sq Goodness of Fit

Residuals
0
-1.3416
-0.4472
-0.6708
0.4472
2.0125

Chi-Square:  6.7
d.f.:  5
P(CHI<=chi): 0.756075
Chi-Critical (95%):  11.0705
Chi-Critical (99%):  15.0863
```

· Interpretation.
~~~~~~~~~~~~~~

1.24 Description and script operation...

· Chi-square (x^2) goodness of fit analysis for normal distribution.
~~~~~~
Goodness of fit (GOF) analysis is concerned with testing a hypothesis that sampled data may be randomly drawn from a theoretical distribution or that a random variable has a specific theoretical probability distribution. The data for this test consist of N independent observations of a random variable which are grouped into c classes. The measurement scale must be at least nominal. If the number of observations in any one cell is very low, the observations may be combined into adjacent cells.

· Script operation.
~~~~~~~~~~~~~~~~
This script allows the user to interact with the computer to group raw data for the analysis. It requires the user to respond to a series of requesters:
* to provide the desired class interval (if 0 is entered/confirmed, the script will calculate an interval;
* to provide the first lower class boundary (l.c.b.) (if 0 is entered/confirmed, the script
will calculate the l.c.b);
* to allow the "degress of freedom" to be altered based on the parameters being estimated (if 0 is entered/confirmed, the default of c-1 is used).
~~~~~~~~~~~~~~~~

```
Otherwise, the script operates as per normal. Note that the first cell in the
column can be a label which is used by the script on output.

Click here for information about general script usage.

Here is an example (note only the "DO_value" column is used):

Dissolved oxygen values for 50 sampled lakes.
Lake DO_value
 1 5.1
 2 5.6
 3 5.3
 4 5.7
 5 5.8
 6 6.4
 7 4.3
 8 5.9
 9 5.4
10 4.7
11 5.6
12 6.8
13 6.9
14 4.8
15 5.6
16 6.4
17 5.9
18 6
19 5.5
20 5.4
21 4.4
22 5.1
23 5.6
24 5.8
25 5.7
26 4.9
27 6.6
28 5.7
29 5.4
30 5.9
31 5.6
32 6.7
33 5.4
34 4.8
35 6.4
36 5.8
37 5.3
38 5.7
39 6.3
40 4.5
41 5.6
42 6.2
43 4.2
44 5.2
45 5.8
46 6.1
47 5.1
```

```
48 5.9
49 5.5
50 4.7
```

Here is the output based on requesting a starting l.c.b of 4 and accepting the other defaults:

```
Chi-Sq Goodness of Fit of Normal Frequency Distribution

DO_value
   Mean    Standard  Proportion  Expected  Observed
l.c.b  u.c.b Deviation Score   Within    Frequency Frequency
6.57           0.0605    3.02    4
6.14 6.57  0.99    1.5509    0.1289    6.44    5
5.71 6.14  0.56    0.8804    0.2275    11.38   10
5.29 5.71  0.13    0.2099    0.2606    13.03   18
4.86 5.29  -0.29   -0.4606   0.1935    9.68    5
4.43 4.86  -0.72   -1.131    0.0932    4.66    5
4  4.43  -1.15   -1.8015   0.0358    1.79    3


Chi-Square:  5.8042
d.f.:  6
P(CHI<=chi): 0.554527
Chi-Critical (95%):  12.5916
Chi-Critical (99%):  16.812
```

- Interpretation.
  ~~~~~~~~~~~~~~~

The calculated Chi-Sq is very sensitive to the classes being generated. The objective should be to minimise the Chi-Sq statistic as much as possible. This may require additional adjustments to the class boundaries etc.

Look at the operation of the other Chi-Sq Goodness of Fit script here.

1.25 Description and script operation...

- Shapiro-Wilk Test for Normality.
  ~~~~~~~~~~~~~~~~~~~~~~~~~~~

The test is used to see if the (unknown) distribution function of a random sample is normal or not.

The sample is of size N where N<=50. If the sample is >50, then the script will estimate the W statistic using the Shapiro-Francis test. This will be indicated to the user through on-screen messages.

- Script operation.
  ~~~~~~~~~~~~~~~~

This tool operates in much the same way as most of the others. Only one column of data is allowed. The column heading may be included in the selection range but is discarded by the script.
Click here for information about general script usage.

The test statistic,W , should be close to 1 if the sample behaves like a normal sample, or close to 0 if non-normal.
  ~~~~~~~~~~~~~~~~

The script will calculate W, G (the approximate standard normal of W), and the p-value. For sample sizes less than 7, the G and p-value may be less accurate and should be checked against appropriate tables of values.

If the Shapiro-Francis test is used, only the W statistic is calculated.

Hypotheses:

H0: F(x) is a normal distribution function
H1: F(x) is non-normal

In this example, a sample of 50 numbers is generated by computer:

```
23
23
24
27
29
31
32
33
33
35
36
37
40
42
43
43
44
45
48
48
54
54
56
57
57
58
58
58
58
59
61
61
62
63
64
65
66
68
68
70
73
73
74
```

```
 75
 77
 81
 87
 89
 93
 97

Spreadsheet output:

 Shapiro-Wilk_W_Test_of_Normality_for_Small_Samples
 W: 0.96413
 G: -0.75303
 p-value:0.2257
```

· Interpretation.
  ~~~~~~~~~~~~~~~
 Specific tables need to be consulted that indicate quantiles of the W statistic in order for the level of significance to be determined. In this example, the W value lies between the .1 and .5 quantiles indicating a p-value of approximately .29. However, the calculated p-value is more precise at .2257, indicating we can accept the hypothesis H0.

1.26 Description and script operation...

· Fisher Exact Test for 2x2 contingency table data.
  ~~~~~~
 In experimental situations where there is simultaneous collection of data for two variables with two categories, a hypothesis may be formulated to detect independence of the frequencies of occurrence in the categories of one variable with the other. When the raw data is set out in a contingency table the table itself is denoted by rows(r) x columns(c) to form ’cells’ that compose the table. Note that a contingency table may have a configuration other than 2x2, although this is not currently supported by this test.

 The usual method of analyzing contingency table data is to employ a chi-square (x^2) test to detect any significant difference between observed and expected data frequencies. In order to do this, a null hypothesis must be constructed of the form ’frequencies of observed row values are independent of the frequencies of observed column values’, or vice versa. The use of chi-square analysis to test such a hypothesis is found elsewhere:

 Related tool: Chi-square test for independence of sample data.

 The Fisher Exact Test is sometimes referred to as the Fisher-Irwin test or the Fisher-Yates test. In this test the output takes the form of direct binomial probabilities of the observed frequencies being derived at random from the sum of row and column data. In effect, the calculations involve determination of the probability of the whole contingency table occurring by random chance. This probability is calculated once for the observed data table and then for corresponding tables of the next extreme data data values. Any one-tail or two-tail hypothesis is tested for significance by analysis of the total probabilities of all the tables being considered.

Click here for the probability formula used by this test.

· Script operation.
  ~~~~~~~~~~~~~~~~

This tool operates in much the same way as most of the others with only one departure from the usual methods needed (i.e., row and column titles for the 2x2 contingency table should not be included in the data input range).

Click here for information about general script usage.

Raw sample data is entered as a 2x2 contingency table so that the observed data values, or frequencies, for each variable are present. See the example raw data input range below for further details:

```
Raw data (2x2 cont. table):        Spreadsheet output:

       Beetles  Snails       Fisher Exact Test for Independence
 Upper leaf       12       7
 Lower leaf        2       8       Prob. for Obs. Freq.:     0.0292
           Prob. Extreme Freq. (RH Tail):  0.0015
           Total Prob. (RH Tail):          0.0308
           Prob. Extreme Freq. (LH Tail):  0.0036
           Total Prob. (LH Tail):          0.0329
           Total of Observations:              29
```

· Interpretation.
  ~~~~~~~~~~~~~~

In the example data input the number of beetles and snails were obtained from the upper and lower surfaces of leaves. For one reason or another we may suspect that the proportions of animals may differ on either side of the leaf and we may set up appropriate null hypotheses.
For example the following one-tail (i.e., directional) null hypothesis may be proposed (amongst others):

HO: The proportion of beetles is less than, or equal to, the proportion of snails on the upper surface of the leaf.

Note that it is possible to test a two-tail null hypothesis using the Fisher Exact Test but that this will not be further mentioned here and the spreadsheet output itself provides no direct test results for this. Although a two-tail test may be useful when small sample sizes or expected values are encountered, this form of hypothesis test may be tested using the usual chi-square test for independence where the independence of row and column data is determined.

In the case of HO above and with reference to the spreadsheet output it can be seen that the analysis tool first determines the probability of obtaining the observed values by random chance (P=0.0292). The nature of the null hypothesis means that we are then interested in the probabilities associated with the left-hand distribution tail. The total probability for contingency table data more extreme than the observed data is computed to be P=0.0036. This is then added to the probability value previously calculated to give a total probability (P=0.0329) of obtaining all the contingency table data by chance if HO is true. As P=0.0329 is less than

0.05 (if testing at this level of significance) then the null hypothesis (HO) should be rejected.

1.27 Description and script operation...

· Friedman analysis of variance by ranks.
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

The parametric equivalent to this test is the two-way ANOVA with no replication test.

· Script operation.
~~~~~~~~~~~~~~~~

This tool operates in much the same way as most of the others with no specific departures from the usual methods needed.

Click here for information about general script usage.

Raw sample data must be entered as multiple samples, the data being arranged in columns. Note that sample titles may be included within the output by inclusion within the input range.

```
Raw data:           Spreadsheet output:

   I    II   III    IV     Friedman 2-Way ANOVA by ranks
   4     5     9     3     Table of ranks
   8     9    14     7              I    II   III    IV
   7    13    14     6              2     3     4     1
  16    12    14    10              2     3     4     1
   2     4     6              2     3     4     1
   1     4     3              4     2     3     1
   2     6     9              1     2     3     4
   5     7     9              1     3     2     4
                  1     2     3     4
                  1     2     3     4

         Sum of Ranks:       14    20    26    20
         Chi-Square:        5.4
         d.f.(v):             3
         P(CHI<=chi):    0.8553
         Chi-Critical (95%):   7.8147
         Chi-Critical (99%):  11.3449
```

· Interpretation.
~~~~~~~~~~~~~~
~~~~~~~~~~~~~~

1.28 Description and script operation...

· Spearman’s rank correlation coefficient.
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

Investigation of the degree of correlation between two variables in effect determines the strength of association between them. There are two commonly used correlation techniques. Spearman’s rank correlation coefficient is calculated for data variables that do not exhibit a normal distribution, and is therefore a non-parametric form of analysis.

Related tools:

The parametric equivalent to this test: Pearson product-moment correlation test.

A further non-parametric equivalent: Kendall’s rank correlation.

The Pearson product-moment correlation coefficient is more sensitive and therefore preferable but assumes that sample variables are normally distributed.

The Spearman’s rank correlation coefficient was one of the earliest developed statistics based on ranks. The statistic generated, sometimes known as ’rho’ is usually represented as ’r(s)’.

· Script operation.
~~~~~~~~~~~~~~~~

This tool operates in much the same way as most of the others with no specific departures from the usual methods needed.

Click here for information about general script usage.

Raw sample data must be entered as two equal sized samples, the data being arranged in columns. Note that sample titles may be included within the output by inclusion within the input range.

```
Raw data:            Spreadsheet output:

    X       Y        Spearmans Rho - Rank Correlation Non-Parametric Test
 10.4     7.4          Rank_X     Rank_Y
 10.8     7.6               4    5
 11.1     7.9             8.5    7
 10.2     7.2              10   11
 10.3     7.4             1.5        2.5
 10.2     7.1               3    5
 10.7     7.4             1.5    1
 10.5     7.2               7    5
 10.8     7.8               5        2.5
 11.2     7.7             8.5        9.5
 10.6     7.8              11    8
 11.4     8.3               6        9.5
               12   12

         Ties:              5
         Rho(no ties):            0.8531
```

```
Rho(corrected for ties):       0.8511
P(Rho<=rho):           0.000323
d.f.:           10
t:           5.1261
P(T<=t) one-tail:      0.000223
T-Critical (95%):        1.8125
T-Critical (99%):        2.7638
P(T<=t) two-tail:      0.000447
T-Critical (95%):        2.2281
T-Critical (99%):        3.1693
```

· Interpretation.
~~~~~~~~~~~~~~

In the example above the correlation coefficient calculated (0.8531) indicates that there is a fairly strong degree of positive correlation between the two data variables.

If the correlation coefficient is near to -1, correlation is known to be negative; a value of +1 means positive correlation; a value of 0 means that there is no correlation.

Although we can see that the positive correlation between the two example variables is reasonably strong it is orthodox practice to express this result in the usual statistical terms of probability.

There are two ways in which this may be carried out using the output provided:

The first method will require access to a table of significance levels of the value of 'r(s)'. This table is widely available in the appendices of most text books.

The first method: using the r(s) distribution.

In the example above, r(s)=0.8531 based upon 12 pairs of observations. The n=12 value is used to enter the table. Reference to the r(s) correlation coefficient table will indicate that the value of 0.8531 exceeds the critical value (0.587) for a two-tailed test at the 0.05 level of significance. Use of such a table will indicate that the calculated value of r(s) is significant beyond the 0.001 level of significance and this is demonstrated in the probability value in the output (i.e., 0.000323). In other words P<0.05 and we can be more than 95% certain that there is a statistically significant correlation between the two variables.

Note that a further value of r(s), or rho, is calculated. This statistic should be used in preference if the number of ties (occurrences of identical data elements in a sample) is larger than approximately a quarter of the sample size.

The second method: using the t-distribution.

If large sample sizes are available the significance of the r(s) statistic (corrected for ties) can also be tested by comparison of a calculated t-statistic with a critical value at a given level of significance. If the critical value at the 0.05 or 0.01 level of significance is exceeded by the
~~~~~~~~~~~~~~

t-statistic then the null hypothesis, that there is no correlation, should be rejected.

Note that this procedure assumes that the null hypothesis is HO:r(s)=0. If another value other than '0' is hypothesised then the correlation coefficient cannot be predicted to have come from a distribution approximated by the normal distribution, and this in turn invalidates the calculation of a t-statistic.

Using the example above, at the 0.05 level of significance the two-tailed t-critical value with d.f.=10 is 2.2281. As this is exceeded by the computed value of 't' this also indicates that the null hypothesis should be rejected.

1.29 Description and script operation...

- Kendall rank correlation coefficient (tau).
  ~~~~~~

Investigation of the degree of correlation between two variables in effect determines the strength of association between them. There are two commonly used correlation techniques. Spearman Rank correlation coefficient is calculated for data variables that are ranked and do not exhibit a normal distribution, and is therefore a non-parametric form of analysis. The Kendall rank correlation is a further alternative form of non-parametric analysis.

Related tools:

The parametric equivalent to this test: Pearson product-moment correlation test.

A further non-parametric equivalent: Spearman's rank correlation.

The Pearson product-moment correlation coefficient is more sensitive and is therefore the preferable tool to be used for correlation analysis but assumes that sample variables are normally distributed.
The Kendall rank correlation coefficient is a very similar alternative to the Spearman rank test. It has the disadvantages that it is not as powerful when large sample sizes are tested and can be cumbersome to calculate manually with large sample sizes.

- Script operation.
  ~~~~~~~~~~~~~~~~

This tool operates in much the same way as most of the others with no specific departures from the usual methods needed.

Click here for information about general script usage.

Raw sample data must be entered as two equal sized samples, the data being arranged in columns. Note that sample titles may be included in the output by including these within the data input range.

Raw data: Spreadsheet output:

```
  X      Y        Kendall’s tau: Nonparametric Rank Correlation
  4      4
  1      2        Pairs Sorted on Col.1
  6      5              X      Y
  5      6              1      2
  3      1              2      3
  2      3              3      1
  7      7              4      4
          5      6
          6      5
          7      7

    Concordant Pairs:       18
    Discordant pairs:        3
    tau:          0.7143
```

```
· Interpretation.
  ~~~~~~~~~~~~~~
 Although this test and the Spearman’s rank test for correlation are similar
 they will not necessarily result in similar values for their respective
 test statistics. However, just as in the case of the test statistic
 determined by using Pearson’s correlation method, the test statistic (or
 degree of correlation) can range from -1 to +1. In the output above it can
 be seen that there is a reasonably strong correlation evident.
```

1.30 Description and script operation...

```
· Durbin Test.
  ~~~~~~~~~~
 This test reduces to the Friedman Test if the number of treatments
 equals the number of experimental units per block.
 Use this test rather than parametric tests if "normality" assumptions are not met ↩
     .

 The test is called "balanced" because:

 * Every block contains k experimental units.
 * Every treatment appears in r blocks.
 * Every treatment appears with every other treatment an equal number of times.

 The test is "incomplete" because not all treatments are applied to each block.

· Script operation.
  ~~~~~~~~~~~~~~~~
 This tool operates in much the same way as most of the others with no specific
 departures from the usual methods needed.

 Note that labels are not used in the output, however the column titles can be
 included in the input range with no harm. Row labels should not be included in
 the range.

 Click here for information about general script usage.

 In the following example, seven people test three of seven samples of a product
```

```
 and rank the three samples 1,2,3.

Raw data:

  Product
 Person A B C D E F G
 1  2 3   1
 2    3 1   2
 3      2 1   3
 4        1 2   3
 5  3       1 2
 6    3       1 2
 7  3   1       2

Spreadsheet output:


 Durbin Test - Balanced Incomplete Block Design
 (There were no ties)
 Durbin T:  8
 df(1): 6
 df(2): 8
 P(F<=f) One-tail:  0.005
 F-Critical(95%): 3.5806
 F-Critical(99%): 6.3707

· Interpretation.
  ~~~~~~~~~~~~~~
 The null hypothesis is that no variety of product tends to be preferred over
 the other. The output provides the Durbin "T" statistic which is tested using
 the F  distribution.

 In this case, the null hypothesis is rejected as the P value for a one-tail
 test is less than .01.

 Note: Multiple comparisons between treatments is not (yet!) part of the output.
```

1.31 Description and script operation...

```
· Linear regression analysis.
  ~~~~~~~~~~~~~~~~~~~~~~~~~~
 Regression analysis is concerned with making predictions of the values of
 one variable (known as the dependent variable, usually 'y') based upon the
 values of another variable (known as the independent variable, usually
 'x'). It is effectively concerned with investigation of the nature of the
 relationship that may exist between two or more variables. As it is
 concerned with making predictions it has greater investigative potential
 than simple correlation analysis between two variables which, although
 useful, simply inform us of the precise relationship in quantitative
 terms.

 Related tools:

 Pearson Product Moment correlation:   Bivariate correlation analysis.
```

Spearman Rank correlation: Non-parametric correlation.

If statistically significant correlation has been observed between two variables then these can be related mathematically through regression analysis to enable a trend line, or line of best fit, to be fitted to the data when one variable is plotted against the other.

It is possible to calculate the equations of two regression lines for linear data relationships: 'the regression line of y on x' and the 'regression line of x on y'. The tool determines the equation for the former regression line (i.e., the regression line of the dependent variable 'y' on the independent variable 'x'). The reason for this is that it is orthodox to design experiments carefully so that the independent 'x' variable is the 'fixed' variable (i.e., concentration in $\mathrm{\mu}$g/ml, temperature in \textdegree{}C, or in the example data below IQ level). In effect, this then allows us to concentrate on deviations about a theoretical 'y' population mean and make according predictions of values of 'y'.

The regression line of 'y' on 'x' takes the form of:

```
     y = bx + a     (sometimes seen as: y = mx + C)

     where,

     'b' is the regression coefficient describing the slope of the
   regression line.

     'a' is the intercept of the regression line where it cuts, or
   intercepts, the y-axis when the data are plotted.
```

Once both 'a' and 'b' have been obtained it is straightforward to predict values of 'y' by substituting values of 'x' into the expression. Make sure that you have not extrapolated the regression line beyond the range of the data used to derive it! At extremes of the data range the relationship may not necesarily still be of the same linear nature.

If the relationship between your 'x' and 'y' variables seems not to be linear then it may be possible to log, square root or arcsine transform it so that it is legitimate to perform the regression analysis. If this is not the case then more complex analysis may be performed which is beyond the scope of this package. If you need to fit complex lines to curvilinear data then 'PolyFit' by Camiel Rouweler (Aminet misc/sci directory) is recommended.

· Script operation.
 ~~~~~~~~~~~~~~~~

This tool operates in a similar way to the others with some small differences. Raw sample data is entered in the input requestor as two columns. During the calculations a further requestor will appear to obtain information about whether the dependent variable is to be found in the first data column. Answer 'Y' or 'N' to proceed.
 ~~~~~~~~~~~~~~~~

Click here for information about general script usage.

Note that column headings for either variable may be included in the input range but will not be used in the output. Typical spreadsheet output is printed below:

Raw data:

IQ (dependent variable)	Reading Scores
x	y
118	66
99	50
118	73
121	69
123	72
98	54
131	74
121	70
108	65
111	62
118	65
112	63
113	67
111	59
106	60
102	59
113	70
101	57

Spreadsheet output:

Least Squares Regression

Predicted Values	Regression Statistics	
67.8932	n:	18
55.1486	Pearson r:	0.8999
67.8932	r sq.:	0.8098
69.9055	Std.Err.of Est.:	146.8964
71.247	Intercept(a):	-11.2576
54.4778	Slope(b):	0.6708
76.6132		
69.9055	T-test	
61.1855		
63.1978	Std.Err.of Reg.Coef.:	0.0813
67.8932	t:	8.2548
63.8685	d.f.:	16
64.5393	P(T<=t) one-tail:	0
63.1978	T-Critical(95%):	1.7459
59.8439	T-Critical(99%):	2.5835
57.1609	P(T<=t) two-tail:	0
64.5393	T-Critical(95%):	2.1199
56.4901	T-Critical(99%):	2.9208

```
ANOVA
Source of     Sum of    Mean     Degrees of   F-ratio
Variation     Squares   Squares  Freedom

Regression:  625.6036  625.6036          1    68.141
Residual:    146.8964     9.181         16
Total:          772.5              17

P(F Sample<=f) one-tail:      0
F-Critical(95%):          4.494
F-Critical(99%):         8.5309
```

· Interpretation.
~~~~~~~~~~~~~~~

There are several components to the output provided by this tool. On the left-hand side is a column of 'Predicted Values'. These are the predicted 'y', or dependent variable, values based on the equation of the regression line. For example, the equation of the line in the example above is:

y = 0.6708x + -11.2576

and when the known values of 'x' are fed into this expression the predicted values of 'y' are obtained and output to the spreadsheet.

The t-test and ANOVA output is designed to test the statistical significance of the regression analysis. In order to do this it is necessary to set up a null hypothesis which may be tested. For example, it may be assumed that if there was no functional relationship between 'x' and 'y' then the slope (the regression coefficient) may be zero. More importantly, the slope of the sample (known as 'b') may be something other than zero but this sample may, or may not, be representative of the population slope (known as 'ß') from which the sample was derived.

A typical null hypothesis may be proposed as:

HO: ß = 0

This may be rejected only if the probability of obtaining the computed value of 'b', from sampling a population that actually has ß = 0, is significantly small (i.e., less than 0.05 or 0.01, etc.).

ANOVA analysis.

In the ANOVA section of the output several intermediate statistics are computed and output to the spreadsheet. The total SS (sum of squares) is a measure of the overall variability of the dependent variable and the regression SS is a measure of the amount of variability among the 'y' values resulting from a linear regression. These two values will be identical only if all data points fall exactly on the regression line. The degrees of freedom associated with the total variability of 'y' values is n-1 and that associated with the variability of 'y' values due to regression is always 1 in simple linear regression. The MS (mean squares) statistics are calculated from this information by MS = SS/d.f. and the F-ratio is determined by F = regression MS/residual MS.
~~~~~~~~~~~~~~~

At the 0.05 level of significance the one-tailed F-critical value for d.f. = 16 can be seen to be 4.494. In this particular case it is found that the null hypothesis mentioned above should be rejected. The reason for this is that the computed F-statistic has exceeded the critical value and the probability of obtaining a value of 'b' that is derived from a population where ß = 0 is lower than 5%.

For further details about ANOVA see the one-way ANOVA section of this guide.

t-test analysis.

In the t-test section of the output summary statistics are also generated. Here the t-test procedure has equivalence to the ANOVA test when the general null hypothesis is HO: ß = 0. Although most significance testing of simple linear regression analysis will employ this hypothesis the t-test, unlike the ANOVA test, will allow directional null hypotheses to be tested such as ß < 0 and ß > 0.

Where HO: ß = 0, it can be seen in the output that for a two-tailed test at the 0.05 level of significance the null hypothesis should also be rejected as the calculated value of 't' (i.e., 8.2548) exceeds the critical value of 2.1199.

For further details about t-tests see the relevant sections beginning with the t-test of independent sample means.

1.32 Description and script operation...

- Exponential smoothing of data.
  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

The exponential smoothing technique may be used for making a prediction of the next value in a set of data, particularly when this is unknown due to being in the future. For example if the average quantity of a product sold is known for each month of a year then we can then use this information to predict sales for a subsequent month. This adjustment technique uses calculation of weighted averages to provide forecast values which are exponentially weighted equivalents of the original data. Exponential smoothing, as a weighted moving average technique, places more emphasis on the observations which are nearest to the time period being forecast than does simple moving average calculations which weight all original data elements equally.

Related tool:

Moving average calculations.

- Script operation.
  ~~~~~~~~~~~~~~~~

This tool operates in much the same way as most of the others. However, only column of data may be entered in the input range. A further requestor will also demand the damping factor to be used (default is 0.3) which may

```
range from 0 to 1.

Click here for information about general script usage.

Typical input data and output information is shown below. Note that any
title in the data column may be included within the output by inclusion
within the input range.

Raw data:          Spreadsheet output:

 Sample A           Sample_A
  1          Forecast Values    Standard Error
  3                  0     0
  6                  1     0
  8                2.4     0
  3               4.92     0
  8              7.076      2.9685
  9             4.2228      3.1179
  7             6.8668      2.9971
  3             8.3601      2.8445
  8              7.408      3.0579
  9             4.3224      3.1254
        12            6.8967      3.0467
        14             8.369      3.0988
        15           10.9107      3.0988
        13.0732      3.0302

     Smoothing
     Constant:              0.7
```

· Interpretation.
~~~~~~~~~~~~~~

1.33 Description and script operation...

· Moving Average

Time series analysis is based on a set of observations taken at specific times, usually at equal intervals. The data may (and usually does) express cyclic, seasonal and irregular movements that often can mask a long term trend.

'Moving averages' is one way of smoothing the time series to expose the trend. Given a set of numbers X1,X2,X3,..., we define a moving average of order 'N' to be the sequence of arithmetic means. Thus, for example, for the sequence 4,6,8,7,5 with N=3 the moving average is expressed by the sequence (4+6+8)/3, (6+8+7)/3, (8+7+5)/3 (or 6,7,6.66).

Be aware that there are disadvantages to using this method:
* The data at the start and end of the series are lost.
* The moving average may generate its own cycles or other movements not found in the original data.
* The moving average is strongly affected by extreme values, although this
~~~~~~~~~~~~~~

could be overcome with the use of appropriate weighting techniques.

A related method of estimating trends is the Least Squares Regression method.

· Script operation.
~~~~~~~~~~~~~~~~

This tool operates in a similar way to others but there are some exceptions which need to be taken into consideration. The first of these is that this tool will only accept raw data arranged in a single column. The second is that during the script operation a further requestor will appear to demand the 'number of intervals' (the default is 3).

Click here for information about general script usage.

Note that any heading for your column data may be included in the data input range and will then appear in the sheet output. Typical input and output is shown below:

```
Raw data:      Spreadsheet output:

 Sample 1       Moving Average
      420
      650       Sample_1
      800
     1420       Forecast    Std.Error
     1360              0      0
     1600              0      0
     2110         623.33      0
     2400         956.66      0
     1193.33      304.003
        1460      298.017
        1690      276.193
     2036.66      333.729

    Interval:          3
```

· Interpretation.
~~~~~~~~~~~~~~

To see how this analysis works, it would be best to plot both the original sample values and the Forecast values. Note that the Forecast values should be plotted midway between the first and last value of each sequence, which in this example, is a sequence of 3 values. So the plot should have the pairs 650:623.33; 800:956.66;etc.

1.34 Description and script operation...

· Covariance of random variables.
 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Covariance is a statistical measure of the degree to which two or more random variables tend to vary together.

· Script operation.
 ~~~~~~~~~~~~~~~~
This tool operates in much the same way as most of the others with no specific departures from the usual methods needed.

Click here for information about general script usage.

Note that column data titles may be included in the input range so that they are then reproduced in the spreadheet output. Typical data input and output are shown below:

```
Raw data:              Spreadsheet output:

Trial A  Trial B  Trial C    Covariance: Ungrouped Data
      2  5       40
      3  4       30             Trial_A  Trial_B  Trial_C
      6       12      110    Trial_A      2.5      4.5       45
      5  8       70    Trial_B      4.5   9.6875   96.875
           Trial_C       45   96.875   968.75
```

· Interpretation.
 ~~~~~~~~~~~~~~

1.35 Description and script operation...

· Shannon-Wiener biodiversity index.
 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Biodiversity indices, for which the Shannon-Wiener test is often employed, can be useful in any ecological assessment of a particular environment as they take in to account species richness (no. of species) and species equitability (relative abundance of species).

Whilst the Shannon-Wiener index is frequently used in ecology (there are others) two immediate problems associated with its use spring to mind. Firstly, the ecological importance of a species in the structuring of a community is not taken into consideration. For example, a large predatory organism will often be the least abundant organism in the community but its removal or absence may cause the community structure to change dramatically. Secondly, the comparison of index values from different sources (i.e., different laboratories) is potentially dangerous. The reason for this is that different scientists may exert different levels of taxonomic, or species identification, error. Furthermore, comparisons between results from different sources become meaningless when different log bases are used to obtain proportions of species.

The analysis tool bases its calculations on the use of log10 but other bases (i.e. log2, etc.) could be used. In this respect no problem should occur if the same log base is used. Many scientists do not actually state which log base they have used!

Note that you can also conduct a t-test to detect any significant difference in biological diversity at two sampling sites.

· Script operation.
~~~~~~~~~~~~~~~~

This tool operates in much the same way as most of the others with no specific departures from the usual methods needed.

Click here for information about general script usage.

Raw sample data is entered in one or two columns as shown in the printed example below. Note that this tool will not accept sample data with unequal sample sizes (i.e., where species abundance data is available for two sites). If two sets of sample data are analysed then the script will also conduct the t-test.

Note that summary statistics are computed for the two samples.

```
Raw data:       Spreadsheet output:

 Site 1   Site 2  Shannon-Wiener Index of Bio-diversity
     47       48  Using Log10 as Base
     35       23
      7       11            Site_1      Site_2
      5       13  Total Count:        99         105
      3        8  No. of Classes:      6           6
      2        2
      Index (H’):           0.5403      0.6328

      Variance:            0.00138     0.00097

      Std. dev.:            0.0484
      t:              -1.909
      d.f.(v):      195.9277
      P(T<=t) one-tail:   0.028863
      T-Critical (95%):     1.6527
      T-Critical (99%):     2.3456
      P(T<=t) two-tail:   0.057726
      T-Critical (95%):     1.9722
      T-Critical (99%):     2.6013
```
~~~~~~~~~~~~~~~~

· Interpretation.
~~~~~~~~~~~~~~~
The biodiversity index figures calculated do not have any units of measurement. In this respect they are only really useful when used to compare sites or habitats (i.e., two different freshwater streams, etc.).

In a two-tailed t-test we are interested in whether there is any statistically significant difference in biodiversity at each site; either site could exhibit higher or lower diversity than the next. In this example, the t-critical value for a two-tailed test at the 0.05 level of significance is not exceeded by the t-statistic computed. In other words, there is no statistically significant difference in the biodiversity at both sites.

1.36 Description and script operation...

· Random Normal Deviates Generator.
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
In case this is useful to anyone, there is a script supplied that generates columns of random normal deviates. These are (almost) random numbers generated by the computer that resemble a random sample from a normal distribution.

· Script operation.
~~~~~~~~~~~~~~~~~
The script will ask two questions: How many columns do you want and how many rows. It will then ask for the output cell.

Here is an example output from a request for 6 columns and six rows:

```
Table of Random Normal Deviates with zero Mean and unit Variance
-0.601788064 -0.231919926  -0.260229439  -0.218680699  -1.00921777 -0.473751528
0.109081557  -0.495747202  -0.627486387  -0.191639674  -1.72554642 -0.836677485
-1.38433659  -0.616409871  -1.16683239 0.861950629 -0.110619413  -0.702969623
0.132517283  -0.959023285  -0.256663751  -0.277253721  0.122101176 -2.03197675
-0.227154427 -0.512849246  0.433628894 0.749493874 0.0529511388  -0.850548665
0.0941952952 -0.461132784  -0.517924642  -0.19387795 -1.14086672 -1.04041071
```

Calculated using the ratio of uniforms method
of A.J. Kinderman and J.F. Monahan
augmented with quadratic bounding curves.
ADAPTED FROM WORK PUBLISHED IN TRANSACTIONS ON MATHEMATICAL SOFTWARE,
VOL. 18, NO. 4, DECEMBER, 1992, PP. 434-435.

1.37 System requirements...

· Required programs and libraries etc.
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
· A copy of Turbocalc v5.x installed on your system.

· Rexxreqtools.library v37.71 copyright 1992-1994 Rafael D'Halleweyn.

- Rexxmathlib.library v38.1 copyright 1998 Thomas Richter.

  The complete archives for these libraries are available from Aminet. Neither library may be used for commercial purposes without the express consent of the copyright holder.

- The program "RexxMast" running. Make sure that you load this program in your user-startup file (you should find this in S:) by inserting the following command:

```
;BEGIN ARexx
RexxMast >NIL:
;END ARexx
```

  Note that the user-startup will also be written to by the installation script. See the installation notes for further details. This is done automatically and simply sets up an assign to the TurboCalc directory (i.e., wherever you have installed TurboCalc and its associated files).

- A web browser to view the html documentation and/or a GIF datatype installed on your system plus "Multiview" installed in the "Sys:Utilities" path in order to view all files associated with this AmigaGuide file.

1.38 Installation procedure...

- How to install this package.
  ~~~~~~~~~~~~~~~~~~~~~~~~~~~

Recommended method

- Double-click on the "Install_TCalcStats2" icon. This installs all the necessary files to your hard-disk and adds an assign to your s:user-startup file. The installation script uses the standard Amiga installer program.

Manual installation

- Copy all Arexx scripts included in this archive to the TurboCalc/ARexx directory on your hard disk.

- Copy the "Stats_Macros.TCD" file anywhere where you will be able to find it via TurboCalc. For example, if you have set up a default path for loading/saving spreadsheets (under Turbocalc's "Global settings") which is a directory named "Sheets" then you may wish to copy the file here for easy access. Bear in mind that this macro file "Stats_Macros.TCD" must be loaded as a sheet within TurboCalc for the scripts to function. See here for further details on use of this package.

- Copy Rexxreqtools.library and Rexxmathlib.library to Libs:
  ~~~~~~~~~~~~~~~~~~~~~~~~~~~

- Set up an assign to your TurboCalc directory. In order that this assign will be made each time you boot your system, the command will need to be written to your user-startup file.
  It may take the form of:

```
;BEGIN TurboCalc
Assign TurboCalc: Work:TurboCalc
;END TurboCalc
```

  Change the directory path for the TurboCalc program to suit the location on your system.

- Copy the drawer containing all text files related to the HTML documentation to the following directory path:

  "TurboCalc:Help/Stats_Help/HTML"

  and all image files associated with the HTML documentation to the following directory path:

  "TurboCalc:Help/Stats_Help/HTML/Images"

  Under normal circumstances it would not matter where you install these files but in this case it is necessary to maintain this directory structure in order that the AmigaGuide file is able to find it's associated image files which are shared. Copy the AmigaGuide anywhere.

1.39 General script operation...

- How to use this package.
  ~~~~~~~~~~~~~~~~~~~~~~~

All the tools work more or less in the same fashion. Begin by entering the data you wish to analyse into a TurboCalc spreadsheet. In general, data will need to be entered in columns. Any deviations from this, and specific instructions for each test are outlined elsewhere.

The Arexx macros can be started either from a Shell or from within TurboCalc:

If started from a Shell, each file will check to see that TurboCalc is running and, if it is not, will start it and present the user with a "File/Open" requester. Load in the file you want to work on.

To start from within TurboCalc, open the spreadsheet you have the data in, then open the "STATS_Macros.TCD" file. The macro sheet contains macros to start each of the Arexx files. Their names appear in the "Play..." macro requester. Make sure your data spreadsheet is active before choosing the macro to play.

If the file selection is okay, the macro will be run. You will then be asked to supply the program with some data. The most common is a request to enter both the cell range for data input and an output cell
  ~~~~~~~~~~~~~~~~~~~~~~~

reference. This (and other program requests) will appear in normal requester windows. Enter the requested information and then press the <Return> key. The program will then proceed.

Note: Data ranges should be entered in the form eg. B2:E25 - the macros will parse that input to find how many rows and columns there are, and to extract the data from the spreadsheet for internal calculations. It is very important that you include blank cells that are at the end of columns to ensure that the total range is included. That is, if you have two or more columns with unequal numbers of cells with data, the range has to include a rectangle big enough to include the column with the most cells.

Labels at the top of columns will be used where they are included in the data range selection. Otherwise, the programs will generate their own labels. Note that if you need the script output to contain labels these may contain spaces when entered into a cell but the spaces themselves will be replaced by '_' characters.
i.e.,

The title 'Female Heights' will become 'Female_Heights'.

Please note that when you give the program a cell reference for output, the program uses that reference as the top left cell in the output. All cells in the output area will be overwritten, so be careful to select cells that are empty.

Once the ARexx file is running, the user is locked out to avoid any problems with stray mouse clicks or keyboard presses when the program stops. The screen display will not be updated until the calculations are finished - so be patient when generating output derived from large sample sizes! For each test a console window provides progress details.

1.40 Problems which may be encountered...

- Operational errors.
  ~~~~~~~~~~~~~~~~~~

No analysis tools available in 'Start Macro' window:

- Ensure that the 'Stats_Macros.TCD' is open behind your active spreadsheet.

- Check that the Arexx files have been successfully installed in the Arexx sub-directory of the TurboCalc: assigned directory. Make sure that the TurboCalc: assign has been made.

No output of results:

- Check that Arexx is running by typing 'RexxMast' in a shell or by double-clicking on the Rexxmast icon in your Sys:System directory. If you do not obtain the message 'REXX server already active' then refer to the instructions for making sure that Arexx starts each time you boot your system (recommended).
  ~~~~~~~~~~~~~~~~~~

- Perform the test again making sure that there are no problems with input and output ranges.

- If all else fails get in touch!

No picture files in AmigaGuide links:

- Ensure that the associated image files are in the following directory path: TurboCalc:Help/Stats_Help/HTML/Images

- Make sure that the program MultiView is in your SYS:Utilities directory.

- Make sure that you have a GIF picture datatype installed on your system.

· Statistical errors.
~~~~~~~~~~~~~~~~~~

Never trust the statistics generated unless you are sure you have conducted the test in question correctly! All of the supplied statistical tools have been designed to be as 'generic' as possible - one of the reasons why many calculate probability and critical values at more than one level of statistical significance and for both one-tailed and two-tailed tests for maximum flexibility. Bear in mind that many statistical tests are very flexible and are often modified for specific applications.

Case example

In this case a one-tailed paired sample t-test can result in problems which are not directly caused by operation of the script:

Consider first a two-tailed test:

HO: mean difference = 0; HA: mean difference is not = 0

```
   Raw data:        T-Test: Two Means for Correlated Samples
     x     y
   142   138        Mean of Diff.:            3.3
   140   136        Variance:              9.3444
   144   147        Std. Dev.:             3.0569
   144   139        Std. Err.:             0.9667
   142   143        t:                3.4137
   146   141        Count:                10
   149   143        d.f.:                  9
   150   145        P(T<=t) one-tail:         0.003852
   142   136        T-Critical (95%):           1.8331
   148   146        T-Critical (99%):           2.8214
        P(T<=t) two-tail:         0.007703
        T-Critical (95%):           2.2622
        T-Critical (99%):           3.2498
```
~~~~~~~~~~~~~~~~~~

This t-test is fine. The calculation for the t-statistic took the form of:

t = Mean of differences / standard error

Therefore reject HO. There is a significant difference between the two samples.
i.e.,

t:	3.4137	
d.f.:	9	
P:	<0.01	(i.e., 0.007703)

Now consider a specific one-tailed test:

HO: mean difference <= 250; HA: mean difference > 250

Raw data: x	y	T-Test: Two Means for Correlated Samples	
2250	1920	Mean of Diff.:	295.5556
2410	2020	Variance:	6502.7777
2260	2060	Std. Dev.:	80.6398
2200	1960	Std. Err.:	26.8799
2360	1960	t:	10.9953
2320	2140	Count:	9
2240	1980	d.f.:	8
2300	1940	P(T<=t) one-tail:	0.000002
2090	1790	T-Critical (95%):	1.8595
		T-Critical (99%):	2.8965
		P(T<=t) two-tail:	0.000004
		T-Critical (95%):	2.306
		T-Critical (99%):	3.3554

This t-test is not correct because we have customised the null hypothesis. This is tested on the basis of directional change with specific criteria (i.e., mean difference is equal to, or less than 250). The analysis tool calculated the t-test in the normal fashion without 'illegal' results. In effect, the output must then be modified by the user to accomodate the null hypothesis. The calculation for the t-test took the form of:

t = Mean of differences / standard error

This provided the following results and conclusion:

Reject HO. The mean difference is significantly greater than 250.
i.e.,

t:	10.9953
d.f.:	8
P:	<0.01

The modified t-test should take the form of:

```
       t = (Mean of differences - 250) / standard error


 To recalculate the value of the t-statistic simply use the necessary
 values provided by the output in a new cell calculation adjacent to the
 old t-statistic.

 This would then provide the following result:


 Retain HO. The mean difference is not significantly greater than 250.
 i.e.,

      t:       1.695
      d.f.:        8
      P:       >0.05
```

1.41 Archaelogical diggings...

```
· Previous versions.
  ~~~~~~~~~~~~~~~~~

  v1.0
  ~~~~
    - Statistical analysis tools for TurboCalc v3.5. Not compatible with
      later versions (untested with v4.x).

    - Uploaded to Aminet as "tcalc_stats.lha". Can be found in the
      util/rexx directory or on Aminet Set 4 disk A.

    - Featured the following analysis tools:

   - ANOVA_OneWay.rexx

   - ANOVA_TwoWay.rexx

   - ANOVA_TwoWay_no_rep.rexx

   - Chi-Sq_2_Ind_Sam.rexx

   - Correlation.rexx

   - Covariance.rexx

   - Descriptive_Stats.rexx

   - Exponential_Smooth.rexx

   - F_Ratio.rexx

   - Histogram.rexx

   - Moving_Average.rexx
```

```
  - Rank&percentile.rexx

  - Regression.rexx

  - T-test_Corr_Sample.rexx

  - T-test_IndSamples.rexx

  - T-test_IndSamples_UneqVar.rexx


 v2.0
 ~~~~
   - Statistical analysis tools for TurboCalc v5.x.

   - Redesigned installation procedure.

   - Features further analysis tools:

  - Friedman_Rank.rexx

  - Kruskal-Wallis.rexx

  - Mann-Whitney.rexx

  - Shannon-Wiener.rexx

  - Wilcoxon_Sign_Rank.rexx

  - Normality.rexx

  - Spearmans_rho.rexx


   - Hypothesis test scripts now include distribution critical values and
     the probability of obtaining test statistics (i.e., t, x$^2$, etc.) at
     0.05 and 0.01 levels of significance.

   - AmigaGuide and HTML documentation.
```

1.42 Development plans...

```
· Future versions.
  ~~~~~~~~~~~~~~~
 This package may still be in development. For those who are aware of
 Microsoft Excel's "Analysis Tools" add-in, this package is intended to be
 a (better) equivalent for TurboCalc.

 It is intended that the following additions may be made (if there's a few
 nods of approval!):

 - Probability and critical values for distributions which may still be
   missing. Anyone want to provide the algorithms?!

 - Multiple regression analysis.
```

- More non-parametric tests eg. Siegal-Tukey

- Analysis of circular distributions.

- Possible specialised test procedures (the Shannon-Wiener biodiversity test forms the first of these).

- Implementation of confidence intervals in tests.

- Possible translation of the documentation into other languages - if someone wants to offer! Moi même, je parle français 'un petit pois' (comme on dit en Angleterre..) mais c'est pas bonne... tu vois?¿?

Suggestions are welcome! For contact details see here.

1.43 References...

- Suggestions for further reading.
  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
  The following reading list is presented as a useful source of further information about statistical analysis. It is by no means exhaustive; there are many good texts available on the subject.

- Snedecor, G.W. & Cochran, W.G., 1980. Statistical methods, 7th Ed., 507pp. Iowa State University Press. ISBN 0-8138-1560-6

  A long-running treatise on statistical analysis. Advanced - requires some background knowledge.

- Zar, J.H., 1984. Biostatistical analysis, 2nd Ed., 718pp. Prentice-Hall International Inc. ISBN 0-13-077925-3

  Advanced - requires some background knowledge but highly recommended.

- Burt, J.E. & Barber, G.M., 1996. Elementary statistics for geographers, 2nd Ed., 640pp. The Guilford Press. ISBN 0-89862-999-3

  Undergraduate level. Lots of useful background information plus other information not always found in many other texts (i.e., explanation of probability value method of hypothesis testing, etc.).

- Elzey, F.F., 1967. A first reader in statistics, 71pp. Wadsworth Publishing Co., California.

  Very basic but nice little introductory text, if you can find it...

1.44 Further information...
  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

· Contacting the authors.
~~~~~~~~~~~~~~~~~~~~~~

If you would like to get in touch with the authors for any queries or to offer advice where you think improvements could be made, do so via the following addresses. Let us know if you find this software useful. Bug reports and/or error reports welcomed for future revisions! Please send these to the appropriate author:

Script programming and HTML documentation: Rudy Kohut.

Postal address: Melbourne
Victoria
Australia

E-mail address: kohutr@connexus.net.au

AmigaGuide documentation and some statistical analysis methods: Nick Bergquist.

Postal address: 175 Marlborough Road
Gillingham
Kent
ME7 5HP
England

E-mail address: nick@nab.u-net.com
ICQ: 23842984

· Disclaimer.
~~~~~~~~~~

This package is supplied 'as is' and no express or implied warranties are offered. Under no circumstance will the authors be liable for any direct, indirect, incidental, special or consequential damages (including, but not limited to, loss of data, profits or business interruption) however caused and under any theory of liability arising in any way from the use of this package.

Although every effort has been made to ensure the accuracy of results produced from analysis of data using this package, it is important that familiarisation with analysis techniques and correct application of included tests is made.
See the limitations section for further details.

· Source Material
~~~~~~~~~~~~~~~
The authors would like to acknowledge that the coding for some complex statistical procedures has been adapted from freely available sources on the internet (usually in C or Fortran programming language and adapted to ARexx). A major source has been "http://lib.stat.cmu.edu" (StatLib), a system for distributing statistical software, datasets, and information by electronic mail, FTP and WWW.

The apstat collection contains a nearly complete set of algorithms published in Applied Statistics. The collection is maintained by Alan Miller in Melbourne. Many of the algorithms came directly from the Royal Statistical Society.

Generally, though, most of the statistical algorithms in the ARexx scripts were adapted from formulas available in published text books.

· AmigaGuide information.
~~~~~~~~~~~~~~~~~~~~~~
This guide was written on an Amiga 1200 1230/IV 68882 FPU and 34Mb RAM. It was created using Blacks Editor v1.02 by Marco Negri.
Equations for statistical tests generated using MathScript v3.2 by Simon Ihmig. Other graphics developed using TurboCalc v.5.02 and PPaint v7.1

April 1999.

1.45 Information about TurboCalc by Michael Friedrich...

· TurboCalc v5.x © 1993-98 Michael Friedrich.
~~~~~~
TurboCalc is a spreadsheet program for the Amiga. the latest version (as of 20-01-98) is v.5.02.
Visit the TurboCalc Home Page at:

http://www.uni-karlsruhe.de/~ukrc/turbocalc.html

There are several ways of obtaining the latest copy:

U.K. distribution: Digita International Ltd.
Black Horse House
Exmouth
Devon
EX8 1JL
Tel: 01395 270273
Fax: 01395 268893
~~~~~~

```
    German distribution: Stefan Ossowskis Schatztruhe
       Gesellschaft für Software mbh
       Veronikastr. 33
       D-45131 Essen
        Tel: 0201-788778
        Fax: 0201-798447


    French distribution: Quartz Informatique
       2bis Avenue de Brogny
       74000 Annecy
        Tel: Int+50.52.83.31
        Fax: Int+50.52.83.31


    Italia distribution: NonSoLoSoft di Ferruccio Zamuner
       il distributore di software Amiga
       Casella Postale n. 63
       I-10023 Chieri (TO)
        Tel: 011 9415237
        Fax: 011 9415237
     E-mail: solo3@chierinet.it


 A demo version with some limitations is also available from the biz/demo
 directory of Aminet or on Aminet CDROM No.25.
```

1.46 Index...

```
· Looking for something specific?!
  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

     >>> Under construction <<<


  Authors                    ........ About - contact details.
  Development                ........ Future development plans.
  Disclaimer                 ........ Legal stuff.
  History                    ........ Previous versions.
  Operation                  ........ General script usage.
  Sampling                   ........ Methodology & considerations.
  References                 ........ Recommended texts.
  Regression                 ........ Linear regression analysis.
  TurboCalc                  ........ Program details.
```