Tutorials and worked examples for simulation, curve fitting, statistical analysis, and plotting.
https://simfit.org.uk
https://simfit.silverfrost.com

1 Standard data types

Once a data set has been collected it is useful to see what statistical tests could be done to characterize a possible theoretical statistical distribution underlying the observations. Clearly success in this endeavor depends on having a sufficiently large sample with maximum possible signal to noise ratio, as well as a sensible presumed distribution. First we consider some standard data types.

- **A single one-dimensional data set.**
  Such a sample would ordinarily consist of n observations x_i, such as estimates of blood pressure in n subjects, and it might be sensible in this case to consider an underlying normal distribution.
  Such a sample will be referred to in SimFIT as a n dimensional vector X that is

  $$X = x_1, x_2, \ldots, x_n.$$

  However, in SimFIT this data set would have to be submitted for analysis as a single vertical column of numbers, either from the clipboard or else from a file.

- **Two independent one-dimensional data sets, e.g. unpaired data.**
  For instance a set of n blood pressure measurements on one group, say X, and another set of m blood pressure measurements on an independent group, say Y. That is

  $$X = x_1, x_2, \ldots, x_n$$
  $$Y = y_1, y_2, \ldots, y_m.$$

  Here the question could be if the two sample estimates for the means and variances suggest a common distribution, and the data would have to be presented to SimFIT as two separate vertical columns of measurements, unless $n = m$ when a two-dimensional matrix would be a possible data format.

- **Two dependent one-dimensional data sets, e.g. paired data.**
  For instance a set of n blood pressure measurements on a group before medication, say X_b, and another set of n blood pressure measurements on the same group, say X_a, after medication. That is

  $$X_a = x_{a1}, x_{a2}, \ldots, x_{an}$$
  $$X_b = x_{b1}, x_{b2}, \ldots, x_{bn}.$$

  Here a more searching test for equality of means, that is for the presence or absence of a treatment effect, could be conducted because in such cases the obvious correlation between the groups can be exploited.
  In this instance the two samples could also be submitted to SimFIT as a matrix with n rows and 2 columns, or as two columns selected from a n by m matrix if $m > 2$.

2 Types of data

Before summarizing the simple statistical tests options in SimFIT it must be clear that there are essentially two types of data.

1. **Continuous**
   These would be observations that can take values in a range and are collected using apparatus.
   Examples could be where X is temperature measured using a thermometer, or time measured by a

clock. Of course continuous variables by definition can have an infinite number of values between any two limits but, given the limits of measurement accuracy, they could well be collected as integers. For example temperature to the nearest degree, or time to the nearest second, would still be analyzed using continuous statistical distributions even though the values are expressed as integers.

2. **Discrete**
   These would be observation that are definitely integers and are often presented as counts. Examples could be where X is the number of observations in a limited number of defined categories, such as the number responding to treatment or not responding to treatment in a group of subjects. A special case is dichotomous data where each experiment has only one of two possible outcomes, for example improvement or deterioration, say 0 or 1. Such categorical data are analyzed using discrete statistical distributions, but often the test statistics are continuous random variables, such as a chi-square statistic resulting from a contingency table.

3 Types of test

Another distinction that can be made separates statistical tests into one of two categories.

A) Parametric tests.
These depend upon choosing a defined statistical distribution to model the data. If the model is correct these are the most powerful tests. However, if the model is wrong, or the parameters assumed for the model are incorrect or are estimated from the sample, then performance is degraded. In extreme cases the test may not just be compromised but could lead to completely erroneous conclusions.

B) Nonparametric tests.
These do not depend on an assumed model and frequently use ranks instead of just measured values. They are much weaker than parameteric tests but have the advantage that they seldom lead to false conclusions. That is, for instance, why the Mann-Whitney U test for equal medians is frequently preferred to the Student t test for equal means. It should be noted however that nonparametric tests often lead to test statistics that are only asymptotic to known continuous distributions, so that they can require large samples for reliability.

4 Statistics

Any function evaluated using a data set can be called a statistic, and some listed below are used as test statistics, that is, numbers that can tested for extreme values given a data set and a theoretical distribution.

a) The sample mean
Given a sample $X = x_1, x_2, \ldots, x_n$ of size n, the sample mean $\bar{x}$ defined as

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

estimates the center of the distribution, as opposed to the median which is the point where half of the sample is below the median and half above. The sample mean is frequently used in parametric tests, and the sample median in nonparametric tests.

b) The sample variance
The sample variance s^2 is defined as

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

and its square root s is the sample standard deviation. It is also known as the dispersion and sums up the spread of the data.

c) The sample cumulative distribution

Suppose the sample is rearranged into a non-decreasing order, then the sample cumulative distribution function $C(x_i)$ is a step function which is zero for values below x_1, one for values greater than or equal to x_n, and increases in steps of $1/n$ at each consecutive value of x_i.

d) Rank

The rank of an observation is the position it would take if the sample was to be arranged into non-decreasing order, and ranks are used in many nonparametric tests.

5 Summary of available tests

- **1-sample *t* test.**
  This is a parametric test used to check if a single sample of observations can be considered as arising from a defined normal distribution with a known mean which users can input interactively.

- **1-sample Kolmogorov-Smirnov test.**
  This is a nonparametric test to see if a single sample is consistent with one of a known set of standard distributions. The test statistic calculated is the maximum vertical distance between the sample cumulative distribution and the theoretical distribution. It is only really useful when the assumed distribution is a known continuous distribution with specified parameters, and is much weaker with discrete distributions, or where the parameters are estimated from the sample.

- **1-sample Shapiro-Wilks test for normality.**
  This is a recommended test to see if a sample is consistent with a normal distribution. It is based on the correlation between the sample scores and the expected normal scores.

- **1-sample Dispersion and Fisher exact Poisson test.**
  This tests if a sample of non-negative integers is consistent with a Poisson distribution. It is based on examining the sample variance to see if the data suggest over-dispersion due to clumping or an over-uniform dispersion, both of which can suggest departure from the behavior expected for a Poisson distribution. The Fisher exact test is only performed for small samples.

- **2-sample unpaired *t* and variance ratio tests.**
  This is the most frequently used test to see if two samples have the same mean. It relies on the samples being normally distributed with the same variance and, with large samples, these two assumptions can be tested interactively. It is essentially analysis of variance (ANOVA) but with only two columns.

- **2-sample paired *t* test.**
  This depends on the same assumptions as the 2-sample unpaired t test for equality of means, but in the additional circumstance that pairs of corresponding values are necessarily correlated. For instance, when the column vectors are body temperatures measured with the same subjects but before and after treatment. The correlation allows the use of a test statistic that is more searching than with the 2-sample unpaired t test.

- **2-sample Kolmogorov-Smirnov test.**
  This is a nonparameteric test for equality of distribution. It is based on the maximum vertical distance between the two sample cumulative distributions but is rather weak and requires large samples.

- **2-sample Wilcoxon-Mann-Whitney *U* test.**
  This is the most widely used nonparametric test to see if two samples can be regarded as having the same but unspecified distribution. It is based on the extent to which one of the samples dominates the other, that is, has a larger median.

- **2-sample Wilcoxon signed-ranks test.**
  Just as the Wilcoxon-Mann-Whitney U test is the nonparametric equivalent of the unpaired t test, this is the nonparametric equivalent of the paired t test and tests for equality of medians in two paired samples.

- **Chi-square test on observed and expected frequencies.**
  Given a set of n actual observed frequencies this tests if the observed frequencies are consistent with those generated as expected frequencies by some assumed distribution. It requires users to supply the expected frequencies along with the observed frequencies.

- **Chi-square and Fisher-exact contingency table tests.**
  The chi-square test can always be done on an arbitrary n by m contingency table, but the Fisher-exact test is only useful with small contingency tables. Yates's continuity correction is used for two by two tables.

- **McNemar test.**
  This requires paired samples of dichotomous data (i.e. values 0 or 1) as a two by two contingency table.

- **Cochran Q repeated measures test.**
  This also requires dichotomous data, but in a repeated measures design.

- **Binomial test.**
  This test checks to see if a set of dichotomous observations are consistent with a binomial distribution. The binomial p value and number of trials N are input along with the number of successes or failures.

- **Sign test.**
  Checks if a set of outcomes such as success or failure are consistent with a binomial distribution with $p = 0.5$, that is, equally likely outcomes.

- **Run test.**
  This is used to test if a set of residuals have a succession of signs that is consistent with equally likely positive and negative values occurring randomly along the sequence and not clustering to suggest a biased fit. Whereas the sign test just examines the overall number of positive and negative signs, the run test also depends on the order of occurrence. For instance, if a model fitted to 20 data points resulted in 10 positive and 10 negative residuals the sign test would report this as perfectly normal. If there were 10 positive residuals followed by 10 negative residuals, that is the sign pattern ++++++++++----------, the run test would draw attention to a badly fitting model.

- **F test for excess variance.**
  This test is performed automatically when SimFIT fits a nested set of models to some data but can also be done interactively using this option. It is based on the fact that adding extra parameters to a model, for example higher order terms in a polynomial, will generally improve the fit, i.e. decrease $WSSQ$, the weighted sum of squared residuals resulting from fitting. This test examines if the increased number of parameters required to decrease the $WSSQ$ is justified on statistical grounds.

- **Runs up and down test for randomness.**
  This is used to test for correlation in a set of numbers, for example pseudo random numbers from a random number generator. The runs up test is based on counting the lengths of runs of numbers increasing in magnitude within in a sequence, and the runs down test is done by simply changing the signs of all the numbers and doing the runs up test.

- **Mood and David tests for equal dispersion.**
  These is a nonparametric test for variance equality in two samples. They examines the ranks of observations in a pooled sample.

- **Kendall coefficient of concordance.**
  This measures the degree of agreement between k comparisons of n objects.

- **Bartlett and Levene tests for homogeneity of variance**
  Analysis of variance is based on the presumption that the samples under investigation are all normally distributed with the same variance. This test examines if all the variances are consistent with this assumption.