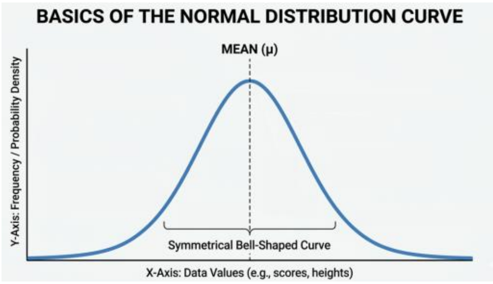
Method Sheet 45

How to perform a Student's T-test

Overview

This method sheet explains how to perform a Student's T-test using Microsoft Excel. This type of statistical test is helpful when analysing data from experiments where only two different treatments, such as a control condition and one other treatment, were analysed in the same experiment. The output of a Student's T-test is a statistical measure, called a p-value. The p-value is an estimate of how likely it is that any differences observed between the two treatments are genuine, and not just due to chance. If the p-value is below a particular threshold that we choose (typically 0.05), we say that the difference between the two means is "statistically significant". If the p-value is greater than the threshold, the difference is not significant, which means it likely occurred due to chance as a result of variability in the data.

Testing for data normality in drug discovery data



T-tests are designed to analyse data that are **normally distributed**. That means, if you were to plot a curve showing how frequently each value appears in a large data set, it would appear as a bell shape curve. This curve is what we call the normal distribution (see image at right).

Although most forms of data from biological assays yield data that are normally distributed, we prefer not to assume that, and typically perform a statistical test, such as the **Shapiro-Wilk test**, to ensure the data are distributed in this way before choosing a statistical test. The problem with drug discovery data, and bioassay results generally, is that the number of independent experiments is typically small, perhaps three or four repeats of the same experiment (i.e. n=3 or 4).

This small number presents a difficulty for the Shapiro-Wilk test, as it can only work effectively with much larger 'n' values. Therefore, when results of *in vitro* bioassays with a small number of repeat experiments, we typically assume that the data are close enough to normal distribution for use in a Student's T-test on the basis of historical use of the assay and visual inspection to ensure there are no very large outliers within replicate measurements of the same treatments.

Performing a Student's T-test in Microsoft Excel

1) Microsoft Excel is quite flexible in terms of the arrangement of the data within the worksheet, so they can be in rows or in columns as you prefer.
2) You should aim to have two separate sets of numbers, each reflecting the result from a whole experiment.
3) For example, if you have completed 4 separate experiments, you would have 4 values for the control condition, and 4 values for the treatment, as shown in the example below:

	A	B	C	D	E
1					
2		Exp 1	Exp 2	Exp 3	Exp 4
3	Control	98%	109%	101%	117%
4	Treatment	82%	78%	71%	58%

4) In an empty cell near these blocks of values, type “T-test p-value”, then in another cell nearby, type the following formula:

```
=TTEST(B3:E3,B4:E4,2,2)
```

5) In this example, the formula tells Excel to compare the values between cells `B3` and `E3` as the first condition, and cell between `B4` and `E4` in the second condition.

C6 | fx | =TTEST(B3:E3,B4:E4,2,3)

	A	B	C	D	E
1					
2		Exp 1	Exp 2	Exp 3	Exp 4
3	Control	98%	109%	101%	117%
4	Treatment	82%	78%	71%	58%
5					
6	T-test p-value		0.0027		

6) Make sure you modify these cell references to correctly indicate where your own data are located in the worksheet.

7) Note that there are two values after the cell references - the first is set to 2, and the second is set to 3.

8) The first value sets the ‘Tails’ of the test, a value of one tells Excel to perform a one tailed T-test, and a value of 2 tells Excel to use a two tailed T-test.

9) A one-tailed T-test means we assume the difference will go in only one direction from the control.

10) We can’t assume that, as our treatments could either increase or decrease the readout in comparison to control, so we require a 2 tailed T-test.

11) The second value tells Excel which ‘Type’ of T-test to use, a value of 1 selects a paired T-test and a value of 2 selects a Two-sample equal variance T-test.

12) We choose a value of 3 to perform a Two-sample unequal variance T-test, since although the measurements are from within the same bioassay and are therefore likely to show similar levels of variance between the two treatments, it is still possible that the variance could be greater in the presence of the treatment.

13) Press enter after typing the formula, and Excel will automatically calculate the result.

14) In this example, the result is 0.0027, which is highly significant if we have set the threshold at 0.05.

Notes

- Remember, you can only use the Student's T-test to compare two groups of numbers, reflecting two different experimental conditions.
- If you have to compare more than two means, you should either perform a statistical test that can adjust for multiple testing, such as ANOVA, or you should apply an adjustment for multiple testing manually, such as the Bonferroni correction in which you multiply your p-values by the number of separate T-tests you have performed.
- If the p-value is below 0.05, it means there is less than a 5% probability that the observed difference occurred by chance.
- If the p-value is below 0.01, it means there is less than a 1% probability that the observed difference occurred by chance.
- You should conduct your T-test, and other statistical tests, using the results from three or more independent experiments.
- Remember that it is not appropriate to perform a T-test, or other statistical tests, using results arising from multiple replicates from within only one experiment.
- For example, you should not perform statistical tests on triplicate measurements from within the same plate - these are called technical replicates, which are not the same as experimental replicates.