Hypothesis testing: comparison of two experimental means

One of the most important properties of an analytical method is that it should be free from bias. That is to say that the test result it gives for the amount of analyte is accurate, close to the true value. This property can be verified by applying the method to a certified reference material or spiked standard solution with known amount of analyte, as discussed in the previous blog article : https://consultglp.com/2018/12/15/7-practical-steps-of-hypothesis-testing/

Another way in which the results of a new analytical method may be tested is by comparing them with those obtained by using a second (perhaps a reference) method.

If the two methods tested on a given sample give two sample means $\bar{x}_1$ and $\bar{x}_2$ with respective repetitions, n_1 and n_2, and respective standard deviations, s_1 and s_2, then our hypothesis testing can be performed as below:

H_o: $\bar{x}_1 = \bar{x}_2$

H_1: $\bar{x}_1 \neq \bar{x}_2$

To start with, we must first establish that the two standard deviations from populations are not significantly different by carrying out a F-test for variance comparison. See article https://consultglp.com/2018/10/01/the-variance-ratio-test/

Once confirmed that these standard deviations are not significantly different, a pooled estimate, s, of the standard deviation can be calculated using the equation:

$$s_p{}^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 + n_2 - 2)} \qquad \text{Eq [1]}$$

To decide whether any significant difference between the two means, i.e., to test the null hypothesis H_o: $\bar{x}_1 = \bar{x}_2$, the statistic t is then calculated from:

$$t = \frac{|\bar{x}1 - \bar{x}2|}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \qquad \text{Eq [2]}$$

where t has $(n_1 + n_2 - 2)$ degrees of freedom.

Example:

In a comparison study of two methods (FAAS and ICP-OES) for the determination of chromium content in a waste water sample, the following results in mg/L were found after 5 determinations done for each method:

FAAS method: mean = 2.38; standard deviation 0.23

ICP-OES method: mean = 2.74; standard deviation 0.19

By calculations, we have:

Pooled standard deviation, s = 0.211 and t-value = 2.698.

There are 8 degrees of freedom and the critical value is t_8 = 2.306 at α = 0.05 for a two-tailed test.

Since the experimental value of *t* is greater than the critical value, it is concluded that the null hypothesis is rejected, indicating that the difference between the two results is significant at the 5% level.

One-tailed or two-tailed test?

In hypothesis testing, we can investigate the alternative hypothesis as a difference between two means in either direction when we have no idea, prior to the experiment, as to whether any difference between the experimental means of two methods will be positive or negative. In this case, the test used must cover either possibility, leading us to use a two-tailed (or two-sided) test. At α=0.05 probability level, we look for the critical value at α=0.025 on either side of the assumed symmetrical normal or Student's t-distribution curve.

In another scenario, we do, say a comparison study of the two analysts' performance and wish to know if Analyst A is more precise than that of Analyst B in terms of standard deviations obtained, we carry out a one-tailed (or one-sided) test, where at α=0.05 probability level, we look for the critical value on either positive or negative side of the distribution curve.

It may be noted that in general, two-tailed tests are much more commonly used than one-tailed ones. If the test is carried out on an Excel spreadsheet, it is necessary to indicate whether a one-tailed or a two-tailed is required.

In case of unequal variances

If the population standard deviation is unlikely to be equal, then we can no longer pool standard deviation like in equation [1], so as to get an overall standard deviation. Instead, we can use an approximate method outlined below:

In order to test H_o: $\bar{x}_1 = \bar{x}_2$, when it cannot be assumed that the two samples come from populations with equal variances or standard deviations (which are squared roots of variances), the statistic t is calculated where:

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \qquad \text{Eq [3]}$$

with the number of degrees of freedom

$$v = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{s_1^4}{n_1^2(n_1 - 1)} + \frac{s_2^4}{n_2^2(n_2 - 1)}} \qquad \text{Eq [4]}$$

and the value obtained is truncated to an integer.

You may note that there are several different equations suggested for the number of degrees of freedom for t when S_1 and S_2 differ. This has reflected the fact that the method is an approximate one. The equation [4] is being used by MS Excel which rounds the degrees of freedom number to the nearest integer.