# THE TESTS

## Test 1 *Z*-test for a population mean (variance known)

### Object

To investigate the significance of the difference between an assumed population mean $\mu_0$ and a sample mean $\bar{x}$.

### Limitations

1. It is necessary that the population variance $\sigma^2$ is known. (If $\sigma^2$ is not known, see the $t$-test for a population mean (Test 7).)
2. The test is accurate if the population is normally distributed. If the population is not normal, the test will still give an approximate guide.

### Method

From a population with assumed mean $\mu_0$ and known variance $\sigma^2$, a random sample of size $n$ is taken and the sample mean $\bar{x}$ calculated. The test statistic

$$Z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

may be compared with the standard normal distribution using either a one- or two-tailed test, with critical region of size $\alpha$.

### Example

For a particular range of cosmetics a filling process is set to fill tubs of face powder with 4 gm on average and standard deviation 1 gm. A quality inspector takes a random sample of nine tubs and weighs the powder in each. The average weight of powder is 4.6 gm. What can be said about the filling process?

A two-tailed test is used if we are concerned about over- and under-filling.

In this $Z = 1.8$ and our acceptance range is $-1.96 < Z < 1.96$, so we do not reject the null hypothesis. That is, there is no reason to suggest, for this sample, that the filling process is not running on target.

On the other hand if we are only concerned about over-filling of the cosmetic then a one-tailed test is appropriate. The acceptance region is now $Z < 1.645$. Notice that we have fixed our probability, which determines our acceptance or rejection of the null hypothesis, at 0.05 (or 10 per cent) whether the test is one- or two-tailed. So now we reject the null hypothesis and can reasonably suspect that we are over-filling the tubs with cosmetic.

Quality control inspectors would normally take regular small samples to detect the departure of a process from its target, but the basis of this process is essentially that suggested above.

### Numerical calculation

$\mu_0 = 4.0, n = 9, \bar{x} = 4.6, \sigma = 1.0$

$Z = 1.8$

Critical value $Z_{0.05} = 1.96$ [Table 1].

$H_0$: $\mu = \mu_0, H_1$: $\mu \neq \mu_0$. (Do not reject the null hypothesis $H_0$.)

$H_0$: $\mu = \mu_0, H_1$: $\mu > \mu_0$. (Reject $H_0$.)

## Test 2 $Z$-test for two population means (variances known and equal)

### Object

To investigate the significance of the difference between the means of two populations.

### Limitations

1. Both populations must have equal variances and this variance $\sigma^2$ must be known. (If $\sigma^2$ is not known, see the $t$-test for two population means (Test 8).)
2. The test is accurate if the populations are normally distributed. If not normal, the test may be regarded as approximate.

### Method

Consider two populations with means $\mu_1$ and $\mu_2$. Independent random samples of size $n_1$ and $n_2$ are taken which give sample means $\bar{x}_1$ and $\bar{x}_2$. The test statistic

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sigma \left( \dfrac{1}{n_1} + \dfrac{1}{n_2} \right)^{\frac{1}{2}}}$$

may be compared with the standard normal distribution using either a one- or two-tailed test.

### Example

Two teams of financial sales persons are compared to see if it is likely that the instruction each has received could have led to differing success rates. A sample of nine transactions (which involves the whole team) yields an average success rate of 1.2. Similarly a sample of 16 transactions for the second team yields a success rate of 1.7. The variances for both teams are equal to 2.0750 (standard deviation 1.4405). The success rate is calculated using a range of output measures for a transaction.

If we are only interested to know of a difference between the two teams then a two-tailed test is appropriate. In this case we accept the null hypothesis and can assume that both teams are equally successful. This is because our acceptance region is $-1.96 < Z < 1.96$ and we have computed a $Z$ value, for this sample, of $-0.833$.

On the other hand, if we suspect that the first team had received better training than the second team we would use a one-tailed test.

For our example, here, this is certainly not the case since our $Z$ value is negative. Our acceptance region is $Z < 1.645$. Since the performance is in the wrong direction we don't even need to perform a calculation. Notice that we are not doing all possible combination of tests so that we can find a significant result. Our test is based on our design of the 'experiment' or survey planned before we collect any data. Our data do not have a bearing on the form of the testing.

### Numerical calculation

$n_1 = 9, n_2 = 16, \bar{x}_1 = 1.2, \bar{x}_2 = 1.7, \sigma = 1.4405, \sigma^2 = 2.0750$

$Z = -0.833$

Critical value $Z_{0.05} = 1.96$ [Table 1].

$H_0$: $\mu_1 - \mu_2 = 0, H_1$: $\mu_1 - \mu_2 \neq 0$. (Do not reject $H_0$.)

$H_1$: $\mu_1 - \mu_2 = 0, H_1$: $\mu_1 - \mu_2 > 0$. (Do not reject $H_0$.)

# Test 3 $Z$-test for two population means (variances known and unequal)

## Object

To investigate the significance of the difference between the means of two populations.

## Limitations

1. It is necessary that the two population variances be known. (If they are not known, see the $t$-test for two population means (Test 9).)
2. The test is accurate if the populations are normally distributed. If not normal, the test may be regarded as approximate.

## Method

Consider two populations with means $\mu_1$ and $\mu_2$ and variances $\sigma_1^2$ and $\sigma_2^2$. Independent random samples of size $n_1$ and $n_2$ are taken and sample means $\bar{x}_1$ and $\bar{x}_2$ are calculated. The test statistic

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)^{\frac{1}{2}}}$$

may be compared with the standard normal distribution using either a one- or two-tailed test.

## Example

Brand $A$ of a jumbo-sized pack of potato crisp is known to have a more variable weight than brand $B$ of potato crisp. Population variances are 0.000576 gm$^2$ and 0.001089 gm$^2$, respectively. The respective means for samples of size 13 and 8 are 80.02 gm and 79.98 gm.

Is there a difference between the two brands in terms of the weights of the jumbo packs? We do not have any pre-conceived notion of which brand might be ‘heavier’ so we use a two-tailed test. Our acceptance region is $-1.96 < Z < 1.96$ and our calculated $Z$ value of 2.98. We therefore reject our null hypothesis and can conclude that there is a difference with brand $B$ yielding a heavier pack of crisps.

## Numerical calculation

$n_1 = 13, n_2 = 8, \bar{x}_1 = 80.02, \bar{x}_2 = 79.98, \sigma_1^2 = 0.000576, \sigma_2^2 = 0.001089$

$Z = 2.98$

Critical value $Z_{0.05} = 1.96$ [Table 1].

Reject the null hypothesis of no difference between means.

# Test 4 $Z$-test for a proportion (binomial distribution)

## Object

To investigate the significance of the difference between an assumed proportion $p_0$ and an observed proportion $p$.

## Limitations

The test is approximate and assumes that the number of observations in the sample is sufficiently large (i.e. $n \geqslant 30$) to justify the normal approximation to the binomial.

## Method

A random sample of $n$ elements is taken from a population in which it is assumed that a proportion $p_0$ belongs to a specified class. The proportion $p$ of elements in the sample belonging to this class is calculated. The test statistic is

$$Z = \frac{|p - p_0| - 1/2n}{\left\{\frac{p_0(1-p_0)}{n}\right\}^{\frac{1}{2}}}.$$

This may be compared with a standard normal distribution using either a one- or two-tailed test.

## Example

The pass rate for a national statistics test has been 0.5, or 50 per cent for some years. A random sample of 100 papers from independent (or non-college based) students yields a pass rate of 40 per cent. Does this show a significant difference? Our computed $Z$ is $-2.0$ and our acceptance region is $-1.96 < Z < 1.96$. So we reject the null hypothesis and conclude that there is a difference in pass rates. In this case, the independent students fare worse than those attending college. While we might have expected this, there are other possible factors that could point to either an increase or decrease in the pass rate. Our two-tailed test affirms our ignorance of the possible direction of a difference, if one exists.

## Numerical calculation

$n = 100, p = 0.4, p_0 = 0.5$

$Z = -2.1$

Critical value $Z_{0.05} = \pm 1.96$ [Table 1].

Reject the null hypothesis of no difference in proportions.

# Test 5 *Z*-test for the equality of two proportions (binomial distribution)

## Object

To investigate the assumption that the proportions $\pi_1$ and $\pi_2$ of elements from two populations are equal, based on two samples, one from each population.

## Limitations

The test is approximate and assumes that the number of observations in the two samples is sufficiently large (i.e. $n_1, n_2 \geqslant 30$) to justify the normal approximation to the binomial.

## Method

It is assumed that the populations have proportions $\pi_1$ and $\pi_2$ with the same characteristic. Random samples of size $n_1$ and $n_2$ are taken and respective proportions $p_1$ and $p_2$ calculated. The test statistic is

$$Z = \frac{(p_1 - p_2)}{\left\{P(1-P)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right\}^{\frac{1}{2}}}$$

where

$$P = \frac{p_1 n_1 + p_2 n_2}{n_1 + n_2}.$$

Under the null hypothesis that $\pi_1 = \pi_2$, $Z$ is approximately distributed as a standard normal deviate and the resulting test may be either one- or two-tailed.

## Example

Two random samples are taken from two populations, which are two makes of clock mechanism produced in different factories. The first sample of size 952 yielded the proportion of clock mechanisms, giving accuracy not within fixed acceptable limits over a period of time, to be 0.325 per cent. The second sample of size 1168 yielded 5.73 per cent. What can be said about the two populations of clock mechanisms, are they significantly different? Again, we do not have any pre-conceived notion of whether one mechanism is better than the other, so a two-tailed test is employed.

With a $Z$ value of $-6.93$ and an acceptance region of $-1.96 < Z < 1.96$, we reject the null hypothesis and conclude that there is significant difference between the mechanisms in terms of accuracy. The second mechanism is significantly less accurate than the first.

## Numerical calculation

$n_1 = 952, n_2 = 1168, p_1 = 0.00325, p_2 = 0.0573$
$Z = -6.93$
Critical value $Z_{0.05} = \pm 1.96$ [Table 1].
Reject the null hypothesis.

# Test 6 $Z$-test for comparing two counts (Poisson distribution)

## Object

To investigate the significance of the difference between two counts.

## Limitations

The test is approximate and assumes that the counts are large enough for the normal approximation to the Poisson to apply.

## Method

Let $n_1$ and $n_2$ be the two counts taken over times $t_1$ and $t_2$, respectively. Then the two average frequencies are $R_1 = n_1/t_1$ and $R_2 = n_2/t_2$. To test the assumption of equal average frequencies we use the test statistic

$$Z = \frac{(R_1 - R_2)}{\left(\dfrac{R_1}{t_1} + \dfrac{R_2}{t_2}\right)^{\frac{1}{2}}}.$$

This may be compared with a standard normal distribution using either a one-tailed or two-tailed test.

## Example

Two traffic roundabouts are compared for intensity of traffic at non-peak times with steady conditions. Roundabout one has 952/2 arrivals over 11 minutes and roundabout two has 1168/2 arrivals over 15 minutes. The arrival rates, per minute, are therefore 476/11 (43.27) and 584/15 (38.93) respectively.

What do these results say about the two arrival rates or frequency taken over the two time intervals? We calculate a $Z$ value of 2.4 and have an acceptance region of $-1.96 < Z < 1.96$. So we reject the null hypothesis of no difference between the two rates. Roundabout one has an intensity of arrival significantly higher than roundabout two.

## Numerical calculation

$n_1 = 952, n_2 = 1168, R_1 = \dfrac{n_1}{t_1} = 43.27, R_2 = \dfrac{n_2}{t_2} = 38.93$

$t_1 = 22, t_2 = 30$

$$Z = \frac{(R_1 - R_2)}{\left(\dfrac{R_1}{t_1} + \dfrac{R_2}{t_2}\right)^{\frac{1}{2}}} = \frac{4.34}{(3.26)^{\frac{1}{2}}} = \frac{4.34}{1.81} = 2.40$$

Critical value $Z_{0.05} = 1.96$ [Table 1].

Reject the null hypothesis of no difference between the counts.

# Test 7 *t*-test for a population mean (variance unknown)

### Object

To investigate the significance of the difference between an assumed population mean $\mu_0$ and a sample mean $\bar{x}$.

### Limitations

1. If the variance of the population $\sigma^2$ is known, a more powerful test is available: the $Z$-test for a population mean (Test 1).
2. The test is accurate if the population is normally distributed. If the population is not normal, the test will give an approximate guide.

### Method

From a population with assumed mean $\mu_0$ and unknown variance, a random sample of size $n$ is taken and the sample mean $\bar{x}$ calculated as well as the sample standard deviation using the formula

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

The test statistic is

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

which may be compared with Student's $t$-distribution with $n - 1$ degrees of freedom. The test may be either one-tailed or two-tailed.

### Example

A sample of nine plastic nuts yielded an average diameter of 3.1 cm with estimated standard deviation of 1.0 cm. It is assumed from design and manufacturing requirements that the population mean of nuts is 4.0 cm. What does this say about the mean diameter of plastic nuts being produced? Since we are concerned about both under- and over-sized nuts (for different reasons) a two-tailed test is appropriate.

Our computed $t$ value is $-2.7$ and acceptance region $-2.3 < t < 2.3$. We reject the null hypothesis and accept the alternative hypothesis of a difference between the sample and population means. There is a significant difference (a drop in fact) in the mean diameters of plastic nuts (i.e. between the sample and population).

### Numerical calculation

$\mu_0 = 4, n = 9, \bar{x} = 3.1, s = 1.0, \nu = n - 1$

$t = -2.7$

Critical value $t_{8;\,0.025} = \pm 2.3$ [Table 2].

$H_0$: $\mu = \mu_0, H_1$: $\mu \neq \mu_0$. (Reject $H_0$.)

# Test 8 *t*-test for two population means (variances unknown but equal)

## Object

To investigate the significance of the difference between the means of two populations.

## Limitations

1. If the variance of the populations is known, a more powerful test is available: the $Z$-test for two population means (Test 2).
2. The test is accurate if the populations are normally distributed. If the populations are not normal, the test will give an approximate guide.

## Method

Consider two populations with means $\mu_1$ and $\mu_2$. Independent random samples of size $n_1$ and $n_2$ are taken from which sample means $\bar{x}_1$ and $\bar{x}_2$ together with sums of squares

$$s_1^2 = \sum_{i=1}^{n_1} (x_i - \bar{x}_1)^2$$

and

$$s_2^2 = \sum_{i=1}^{n_2} (x_i - \bar{x}_2)^2$$

are calculated. The best estimate of the population variance is found as $s^2 = [(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2]/(n_1 + n_2 - 2)$. The test statistic is

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)^{\frac{1}{2}}}$$

which may be compared with Student's $t$-distribution with $n_1 + n_2 - 2$ degrees of freedom. The test may be either one-tailed or two-tailed.

## Example

Two snack foods are made and sold in 30 gm packets. Random samples of size 12 are taken from the production line of each snack food and means and variances obtained viz.: mean1 31.75 gm, variance1 112.25 gm$^2$; mean2 28.67 gm, variance2 66.64 gm$^2$. What can be said about the two production processes in relation to the weight of packets?

We use a two-tailed test and find that $t$ is 0.798. Our acceptance region is $-2.07 < t < 2.07$ and so we accept our null hypothesis. So we can conclude that the mean weight of packs from the two production lines is the same.

### Numerical calculation

$n_1 = 12, n_2 = 12, \bar{x}_1 = 31.75, \bar{x}_2 = 28.67, \nu = n_1 + n_2 - 2$

$s_1^2 = 112.25, s_2^2 = 66.64$

$s^2 = 89.445$

$t = 0.798, \nu = 12 + 12 - 2 = 22$

Critical value $t_{22;\,0.025} = 2.07$ [Table 2].

Reject the alternative hypothesis.

# Test 9 *t*-test for two population means (variances unknown and unequal)

## Object

To investigate the significance of the difference between the means of two populations.

## Limitations

1. If the variances of the populations are known, a more powerful test is available: the $Z$-test for two population means (Test 3).
2. The test is approximate if the populations are normally distributed or if the sample sizes are sufficiently large.
3. The test should only be used to test the hypothesis $\mu_1 = \mu_2$.

## Method

Consider two populations with means $\mu_1$ and $\mu_2$. Independent random samples of size $n_1$ and $n_2$ are taken from which sample means $\bar{x}_1$ and $\bar{x}_2$ and variances

$$s_1^2 = \frac{\sum_{i=1}^{n_1}(x_i - \bar{x}_1)^2}{n_1 - 1} \quad \text{and} \quad s_2^2 = \frac{\sum_{i=1}^{n_2}(x_i - \bar{x}_2)^2}{n_2 - 1}$$

are calculated. The test statistic is

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\left(\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}\right)^{\frac{1}{2}}}$$

which may be compared with Student's $t$-distribution with degrees of freedom given by

$$\nu = \left\{ \frac{\left\{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}\right\}^2}{\dfrac{s_1^4}{n_1^2(n_1 - 1)} + \dfrac{s_2^4}{n_2^2(n_2 - 1)}} \right\}.$$

## Example

Two financial organizations are about to merge and, as part of the rationalization process, some consideration is to be made of service duplication. Two sales teams responsible for essentially identical products are compared by selecting samples from each and reviewing their respective profit contribution levels per employee over a period of two weeks. These are found to be 3166.00 and 2240.40 with estimated variance of 6328.27 and 221 661.3 respectively. How do the two teams compare on performance?

We compute a $t$ value of 5.72. Our acceptance region is $-2.26 < t < 2.26$ so we reject the null hypothesis and accept the alternative. There is a significant difference between the two teams. Team 1 is more productive than team 2.

## Numerical calculation

$n_1 = 4, n_2 = 9, \bar{x}_1 = 3166.0, \bar{x}_2 = 2240.4, s_1^2 = 6328.67, s_2^2 = 221\,661.3$

$t = 5.72, \nu = 9$ (rounded)

Critical value $t_{9;\,0.025} = 2.26$ [Table 2].

Reject the null hypothesis.

## Test 10 *t*-test for two population means (method of paired comparisons)

### Object

To investigate the significance of the difference between two population means, $\mu_1$ and $\mu_2$. No assumption is made about the population variances.

### Limitations

1. The observations for the two samples must be obtained in pairs. Apart from population differences, the observations in each pair should be carried out under identical, or almost identical, conditions.
2. The test is accurate if the populations are normally distributed. If not normal, the test may be regarded as approximate.

### Method

The differences $d_i$ are formed for each pair of observations. If there are $n$ such pairs of observations, we can calculate the variance of the differences by

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

Let the means of the samples from the two populations be denoted by $\bar{x}_1$ and $\bar{x}_2$. Then the test statistic becomes

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{s/n^{\frac{1}{2}}}$$

which follows Student's $t$-distribution with $n - 1$ degrees of freedom. The test may be either one-tailed or two-tailed.

### Example

To compare the efficacy of two treatments for a respiratory condition, ten patients are selected at random and the treatments are administered using an oral spray. The patients then perform a treadmill exercise until a maximum exercise rate is reached. The times for these are compared. A suitable period of time is ensured between treatments to prevent the effect of treatments to interact. Do the two treatments differ? In this case we do not expect one particular treatment to be superior to the other so a two-tailed test is used. We compute a $t$ value of $-0.11$ and have an acceptance region of $-2.26 < t < 2.26$. So we accept the null hypothesis of no difference between the two treatments. However, in such situations it is often the case that an improvement over an existing or original treatment is expected. Then a one-tailed test would be appropriate.

### Numerical calculation

$\bar{d} = \bar{x}_1 - \bar{x}_2 = -0.1, n = 10, \nu = n - 1, s = 2.9$

$t = -0.11, \nu = 9$

Critical value $t_{9;\,0.025} = 2.26$ [Table 2].

Do not reject the null hypothesis of no difference between means.

# Test 11 *t*-test of a regression coefficient

### Object

To investigate the significance of the regression coefficient of $y$ on $x$.

### Limitations

The usual assumptions for regression should apply, namely:

1. the variable $y$ follows a normal distribution for each value of $x$;
2. the variance among the $y$ values remains constant for any given values of $x$.

### Method

In order to estimate a linear regression of the form $y = A + B(x - \bar{x})$, a sample of $n$ pairs of points $(x_i, y_i)$ is required. $B$ is called the regression coefficient, and to test the null hypothesis that this is equal to zero we first calculate the sample estimate

$$b = \frac{\sum x_i y_i - \frac{1}{n}\sum x_i \sum y_i}{\sum x_i^2 - \frac{1}{n}\left(\sum x_i\right)^2}.$$

The variance of the $x$s and the variance of the $y$s about the regression line are calculated as follows:

$$s_x^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} \quad \text{and} \quad s_{y\cdot x}^2 = \frac{\sum \{y_i - \bar{y} - b(x_i - \bar{x})\}^2}{n-2}$$

where $\bar{x}$ and $\bar{y}$ are the means of the $x$s and $y$s, respectively. The test statistic becomes

$$t = \frac{bs_x}{s_{y\cdot x}}(n-1)^{-\frac{1}{2}}$$

which follows Student's $t$-distribution with $n - 2$ degrees of freedom. The test must be two-tailed since $b$ may be positive or negative. However, the test may be one-tailed if the alternative hypothesis is directional.

### Example

In an investigation of the relationship between a reaction test, on a vehicle simulator, and a composite test a sample of 12 male subjects is selected. The composite test of reactions is a much cheaper alternative to a vehicle simulator test. A regression relationship is computed with regression coefficient 5.029 and $t$ value 6.86. The acceptance region for the null hypothesis is $-2.23 < t < 2.23$. Since the computed $t$ value lies outside the acceptance region we conclude that slope (b coefficient) is significantly greater than zero and a significant regression exists.

Notice that the test does not tell us how good a predictor $x$ is of $y$, only that the regression is significant.

## Numerical calculation

$\sum x_i = 766, \sum y_i = 1700, \sum x_i^2 = 49\,068,$

$\sum y_i^2 = 246\,100, \sum x_i y_i = 109\,380$

$n = 12, \bar{x} = 68.83, \bar{y} = 141.67, \nu = n - 2$

$s_x^2 = 15.61, s_y^2 = 478.8, b = 5.029$

$s_{y\cdot x}^2 = 92.4$

$t = 6.86, \nu = 10$

Critical value $t_{10;\,0.025} = \pm 2.23$ [Table 2].

Reject the null hypothesis.

# Test 12 *t*-test of a correlation coefficient

## Object

To investigate whether the difference between the sample correlation coefficient and zero is statistically significant.

## Limitations

It is assumed that the $x$ and $y$ values originate from a bivariate normal distribution, and that the relationship is linear. To test an assumed value of the population coefficient other than zero, refer to the $Z$-test for a correlation coefficient (Test 13).

## Method

Given a sample of $n$ points $(x_i, y_i)$ the correlation coefficient $r$ is calculated from the formula

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\left[\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2\right]^{\frac{1}{2}}}.$$

To test the null hypothesis that the population value of $r$ is zero, the test statistic

$$t = \frac{r}{\sqrt{1 - r^2}} \cdot \sqrt{n-2}$$

is calculated and this follows Student's $t$-distribution with $n - 2$ degrees of freedom. The test will normally be two-tailed but in certain cases could be one-tailed.

## Example

In a study of the possible relationship between advertising on television and product preferences a panel of television viewers is selected. For two brands of toothpaste, one supermarket own brand and one popular brand, panel members were asked to score (on a scale from 1 to 20) their preference for each product. The correlation coefficient between brands was 0.32, which is modest but is it significantly greater than zero? The calculated $t$ value is 1.35. Our acceptance region is $-1.35 < t < 1.35$ so we accept the null hypothesis. So there is no association between the brands compared, which would suggest a clear preference for popular brand or own brand. Consumers are less likely to substitute own brand for popular brand when preferences appear not to be associated.

## Numerical calculation

$n = 18, r = 0.32, \nu = n - 2$

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0.32\sqrt{16}}{\sqrt{1-(0.32)^2}} = 1.35$$

Critical value $t_{16;\,0.05} = 1.75$ [Table 2].

Do not reject the null hypothesis. *NB*: In this case the $x$ and $y$ variables are independent.

# Test 13 *Z*-test of a correlation coefficient

## Object

To investigate the significance of the difference between a correlation coefficient and a specified value $\rho_0$.

## Limitations

1. The $x$ and $y$ values originate from normal distributions.
2. The variance in the $y$ values is independent of the $x$ values.
3. The relationship is linear.

When these conditions cannot be met, the user should turn to the Kendall rank correlation test (Test 59).

## Method

With $r$ as defined in the $t$-test of a correlation coefficient (Test 12), using the Fisher $Z$-transformation we have

$$Z_1 = \frac{1}{2}\log_e\left(\frac{1+r}{1-r}\right) = 1.1513\log_{10}\left(\frac{1+r}{1-r}\right).$$

The distribution of $Z_1$ is approximately normal with mean $\mu_{Z_1}$ and standard deviation $\sigma_{Z_1}$ where

$$\mu_{Z_1} = \frac{1}{2}\log_e\left(\frac{1+\rho_0}{1-\rho_0}\right) = 1.1513\log_{10}\left(\frac{1+r}{1-r}\right)$$

$$\sigma_{Z_1} = \frac{1}{\sqrt{n-3}}.$$

The test statistic is now

$$Z = \frac{Z_1 - \mu_{Z_1}}{\sigma_{Z_1}}.$$

## Example

A market research company has assumed from previous research that the correlation between two brands in terms of consumer preference is 0.50. This value has a bearing on stocking levels in supermarkets since one brand will often substitute for another when the number on the shelves of one product runs out. A panel of 24 consumers produces a correlation on preference scores (based on a scale of 1 to 20) for the two brands of 0.75. Can we say that the correlation coefficient is at least 0.50? The Fisher $Z$-transformation value, calculated as 0.973, yields a test statistic of 1.94. The acceptance region is $Z < 1.64$. Since the calculated value is greater than the critical value we reject the null hypothesis. This means that the correlation coefficient is at least 0.50 and there is no need to re-evaluate supermarket stocking policy for these two products.

### Numerical calculation

$r = 0.75, \rho_0 = 0.50, n = 24$

$$\mu_{Z_1} = 1.1513 \log_{10} 3 = 0.5493, Z_1 = 1.1513 \log_{10} \left( \frac{1 + 0.75}{1 - 0.75} \right)$$

$$= 0.9730 \text{ [Table 4]}$$

$\sigma_{Z_1} = 0.2182$

$$Z = \frac{Z_1 - \mu_{Z_1}}{\sigma_{Z_1}} = 1.94.$$

The critical value at $\alpha = 0.10$ is 1.64 [Table 1].
The calculated value is greater than the critical value.
Reject the null hypothesis of no difference.

# Test 14 *Z*-test for two correlation coefficients

## Object

To investigate the significance of the difference between the correlation coefficients for a pair of variables occurring from two different samples and the difference between two specified values $\rho_1$ and $\rho_2$.

## Limitations

1. The $x$ and $y$ values originate from normal distributions.
2. The variance in the $y$ values is independent of the $x$ values.
3. The relationships are linear.

## Method

Using the notation of the $Z$-test of a correlation coefficient, we form for the first sample

$$Z_1 = \frac{1}{2}\log_e\left(\frac{1+r_1}{1-r_1}\right) = 1.1513\log_{10}\left(\frac{1+r_1}{1-r_1}\right)$$

which has mean $\mu_{Z_1} = \frac{1}{2}\log_e[(1+\rho_1)/(1-\rho_1)]$ and variance $\sigma_{Z_1} = 1/\sqrt{n_1-3}$, where $n_1$ is the size of the first sample; $Z_2$ is determined in a similar manner. The test statistic is now

$$Z = \frac{(Z_1 - Z_2) - (\mu_{Z_1} - \mu_{Z_2})}{\sigma}$$

where $\sigma = (\sigma_{Z_1}^2 + \sigma_{Z_2}^2)^{\frac{1}{2}}$. $Z$ is normally distributed with mean 0 and with variance 1.

## Example

A market research company is keen to categorize a variety of brands of potato crisp based on the correlation coefficients of consumer preferences. The market research company has found that if consumers' preferences for brands are similar then marketing programmes can be merged. Two brands of potato crisp are compared for two advertising regions. Panels are selected of sizes 28 and 35 for the two regions and correlation coefficients for brand preferences are 0.50 and 0.30 respectively. Are the two associations statistically different or can marketing programmes be merged? The calculated $Z$ value is 0.8985 and the acceptance region for the null hypothesis is $-1.96 < Z < 1.96$. So we accept the null hypothesis and conclude that we can go ahead and merge the marketing programmes. This, of course, assumes that the correlation coefficient is a good measure to use for grouping market research programmes.

## Numerical calculation

$n_1 = 28, n_2 = 35, r_1 = 0.50, r_2 = 0.30, \alpha = 0.05$

$$Z_1 = 1.1513\log_{10}\left(\frac{1+r_1}{1-r_1}\right) = 0.5493 \text{ [Table 4]}$$

$$Z_2 = 1.1513 \log_{10}\left(\frac{1+r_2}{1-r_2}\right) = 0.3095 \text{ [Table 4]}$$

$$\sigma = \left(\frac{1}{n_1-3} + \frac{1}{n_2-3}\right)^{\frac{1}{2}} = 0.2669$$

$$Z = \frac{0.5493 - 0.3095}{0.2669} = 0.8985$$

The critical value at $\alpha = 0.05$ is 1.96 [Table 1].
Do not reject the null hypothesis.

# Test 15 $\chi^2$-test for a population variance

## Object

To investigate the difference between a sample variance $s^2$ and an assumed population variance $\sigma_0^2$.

## Limitations

It is assumed that the population from which the sample is drawn follows a normal distribution.

## Method

Given a sample of $n$ values $x_1, x_2, \ldots, x_n$, the values of

$$\bar{x} = \frac{\sum x_i}{n} \quad \text{and} \quad s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$$

are calculated. To test the null hypothesis that the population variance is equal to $\sigma_0^2$ the test statistic $(n-1)s^2/\sigma_0^2$ will follow a $\chi^2$-distributkm with $n-1$ degrees of freedom. The test may be either one-tailed or two-tailed.

## Example

A manufacturing process produces a fixed fluid injection into micro-hydraulic systems. The variability of the volume of injected fluid is critical and is set at 9 sq ml. A sample of 25 hydraulic systems yields a sample variance of 12 sq ml. Has the variability of the volume of fluid injected changed? The calculated chi-squared value is 32.0 and the 5 per cent critical value is 36.42. So we do not reject the null hypothesis of no difference. This means that we can still consider the variability to be set as required.

## Numerical calculation

$\bar{x} = 70, \sigma_0^2 = 9, n = 25, s^2 = 12, \nu = 24$

$$\chi^2 = (n-1)s^2/\sigma_0^2 = \frac{24 \times 12}{9} = 32.0$$

Critical value $x^2_{24;\,0.05} = 36.42$ [Table 5].

Do not reject the null hypothesis. The difference between the variances is not significant.

# Test 16 *F*-test for two population variances (variance ratio test)

### Object

To investigate the significance of the difference between two population variances.

### Limitations

The two populations should both follow normal distributions. (It is not necessary that they should have the same means.)

### Method

Given samples of size $n_1$ with values $x_1, x_2, \ldots, x_{n_1}$ and size $n_2$ with values $y_1, y_2, \ldots, y_{n_2}$ from the two populations, the values of

$$\bar{x} = \frac{\sum x_i}{n_1}, \quad \bar{y} = \frac{\sum y_i}{n_2}$$

and

$$s_1^2 = \frac{\sum (x_i - \bar{x})^2}{n_1 - 1}, \quad s_2^2 = \frac{\sum (y_i - \bar{y})^2}{n_2 - 1}$$

can be calculated. Under the null hypothesis that the variances of the two populations are equal the test statistic $F = s_1^2/s_2^2$ follows the $F$-distribution with $(n_1 - 1, n_2 - 1)$ degrees of freedom. The test may be either one-tailed or two-tailed.

### Example

Two production lines for the manufacture of springs are compared. It is important that the variances of the compression resistance (in standard units) for the two production lines are the same. Two samples are taken, one from each production line and variances are calculated. What can be said about the two population variances from which the two samples have been taken? Is it likely that they differ? The variance ratio statistic $F$ is calculated as the ratio of the two variances and yields a value of $0.36/0.087 = 4.14$. The 5 per cent critical value for $F$ is 5.41. We do not reject our null hypothesis of no difference between the two population variances. There is no significant difference between population variances.

### Numerical calculation

$n_1 = 4, n_2 = 6, \sum x = 0.4, \sum x^2 = 0.30, s_1^2 = 0.087$

$\sum y = 0.06, \sum y^2 = 1.78, s_2^2 = 0.36$

$$F_{3;5} = \frac{0.36}{0.087} = 4.14$$

Critical value $F_{3.5;\,0.05} = 5.41$ [Table 3].

Do not reject the null hypothesis. The two population variances are not significantly different from each other.

# Test 17 *F*-test for two population variances (with correlated observations)

## Object

To investigate the difference between two population variances when there is correlation between the pairs of observations.

## Limitations

It is assumed that the observations have been performed in pairs and that correlation exists between the paired observations. The populations are normally distributed.

## Method

A random sample of size $n$ yields the following pairs of observations $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$. The variance ratio $F$ is calculated as in Test 16. Also the sample correlation $r$ is found from

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\left[\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2\right]^{\frac{1}{2}}}.$$

The quotient

$$\gamma_F = \frac{F - 1}{[(F + 1)^2 - 4r^2 F]^{\frac{1}{2}}}$$

provides a test statistic with degrees of freedom $\nu = n - 2$. The critical values for this test can be found in Table 6. Here the null hypothesis is $\sigma_1^2 = \sigma_2^2$, when the population correlation is not zero. Here $F$ is greater than 1.

## Example

A researcher tests a sample panel of television viewers on their support for a particular issue prior to a focus group, during which the issue is discussed in some detail. The panel members are then asked the same questions after the discussion. The pre-discussion view is $x$ and the post-discussion view is $y$. The question, here, is ‘has the focus group altered the variability of responses?’

We find the test statistic, $F$, is 0.796. Table 6 gives us a 5 per cent critical value of 0.811. For this test, since the calculated value is greater than the critical value, we do not reject the null hypothesis of no difference between variances. Hence the focus group has not altered the variability of responses.

**Numerical calculation**

$n_1 = n_2 = 6, \sum x = 0.4, \sum x^2 = 0.30, s_1^2 = 0.087$

$$\sum y = 0.06, \sum y^2 = 1.78, s_2^2 = 0.36, F = \frac{s_2^2}{s_1^2} = 4.14, r = 0.811$$

$$\begin{aligned}\gamma_F &= \frac{F-1}{[(F+1)^2 - 4r^2F]^{\frac{1}{2}}} = \frac{4.14-1}{[(5.14)^2 - 4r^2.4.14]^{\frac{1}{2}}} \\ &= \frac{3.14}{[26.42 - 16.56 \times 0.658]^{\frac{1}{2}}} = 0.796\end{aligned}$$

$\alpha = 0.05, \nu = n - 2 = 4, r = 0.811$ [Table 6].

Hence do not reject the hypothesis of no difference between variances.
The null hypothesis $\sigma_1^2 = \sigma_2^2$ has to be reflected when the value of the test-statistic equals or exceeds the critical value.

# Test 18 Hotelling's $T^2$-test for two series of population means

## Object

To compare the results of two experiments, each of which yields a multivariate result. In other words, we wish to know if the mean pattern obtained from the first experiment agrees with the mean pattern obtained for the second.

## Limitations

All the variables can be assumed to be independent of each other and all variables follow a multivariate normal distribution. (The variables are usually correlated.)

## Method

Denote the results of the two experiments by subscripts A and B. For ease of description we shall limit the number of variables to three and we shall call these $x$, $y$ and $z$. The number of observations is denoted by $n_\mathrm{A}$ and $n_\mathrm{B}$ for the two experiments. It is necessary to solve the following three equations to find the statistics $a, b$ and $c$:

$$\begin{aligned}
&a[(xx)_\mathrm{A} + (xx)_\mathrm{B}] + b[(xy)_\mathrm{A} + (xy)_\mathrm{B}] + c[(xz)_\mathrm{A} + (xz)_\mathrm{B}] \\
&\quad = (n_\mathrm{A} + n_\mathrm{B} - 2)(\bar{x}_\mathrm{A} - \bar{x}_\mathrm{B}) \\
&a[(xy)_\mathrm{A} + (xy)_\mathrm{B}] + b[(yy)_\mathrm{A} + (yy)_\mathrm{B}] + c[(yz)_\mathrm{A} + (yz)_\mathrm{B}] \\
&\quad = (n_\mathrm{A} + n_\mathrm{B} - 2)(\bar{y}_\mathrm{A} - \bar{y}_\mathrm{B}) \\
&a[(xz)_\mathrm{A} + (xz)_\mathrm{B}] + b[(yz)_\mathrm{A} + (yz)_\mathrm{B}] + c[(zz)_\mathrm{A} + (zz)_\mathrm{B}] \\
&\quad = (n_\mathrm{A} + n_\mathrm{B} - 2)(\bar{z}_\mathrm{A} - \bar{z}_\mathrm{B})
\end{aligned}$$

where $(xx)_\mathrm{A} = \sum(x_\mathrm{A} - \bar{x}_\mathrm{A})^2$, $(xy)_\mathrm{A} = \sum(x_\mathrm{A} - \bar{x}_\mathrm{A})(y_\mathrm{A} - \bar{y}_\mathrm{A})$, and similar definitions exist for other terms.

Hotelling's $T^2$ is defined as

$$T^2 = \frac{n_\mathrm{A} n_\mathrm{B}}{n_\mathrm{A} + n_\mathrm{B}} \cdot \{a(\bar{x}_\mathrm{A} - \bar{x}_\mathrm{B}) + b(\bar{y}_\mathrm{A} - \bar{y}_\mathrm{B}) + c(\bar{z}_\mathrm{A} - \bar{z}_\mathrm{B})\}$$

and the test statistic is

$$F = \frac{n_\mathrm{A} + n_\mathrm{B} - p - 1}{p(n_\mathrm{A} + n_\mathrm{B} - 2)} T^2$$

which follows an $F$-distribution with $(p, n_\mathrm{A} + n_\mathrm{B} - p - 1)$ degrees of freedom. Here $p$ is the number of variables.

## Example

Two batteries of visual stimulus are applied in two experiments on young male and female volunteer students. A researcher wishes to know if the multivariate pattern of responses is the same for males and females. The appropriate $F$ statistic is computed as 3.60 and compared with the tabulated value of 4.76 [Table 3]. Since the computed $F$ value is less than the critical $F$ value the null hypothesis is of no difference between the two multivariate patterns of stimulus. So the males and females do not differ in their responses on the stimuli.

### Numerical calculation

$n_{\mathrm{A}} = 6, n_{\mathrm{B}} = 4, DF = \nu = 6 + 4 - 4 = 6, \alpha = 0.05$

$(xx) = (xx)_{\mathrm{A}} + (xx)_{\mathrm{B}} = 19, (yy) = 30, (zz) = 18, (xy) = -6, \nu_1 = p = 3$

$(xz) = 1, (yz) = -7, \bar{x}_{\mathrm{A}} = +7, \bar{x}_{\mathrm{B}} = 4.5, \bar{y}_{\mathrm{A}} = 8, \bar{y}_{\mathrm{B}} = 6, \bar{z}_{\mathrm{A}} = 6, \bar{z}_{\mathrm{B}} = 5$

The equations

$$\begin{aligned} 19a - 6b + c &= 20 \\ -6a + 30b - 7c &= 16 \\ a - 7b + 18c &= 8 \end{aligned}$$

are satisfied by $a = 1.320$, $b = 0.972$, $c = 0.749$. Thus

$$T^2 = \frac{6 \times 4}{10} \cdot (1.320 \times 2.5 + 0.972 \times 2 + 0.749 \times 1) = 14.38$$

$$F = \frac{6}{3 \times 8} \times 14.38 = 3.60$$

Critical value $F_{3.6;\,0.0} = 4.76$ [Table 3].
Do not reject the null hypothesis.

# Test 19 Discriminant test for the origin of a *p*-fold sample

### Object

To investigate the origin of one series of values for $p$ random variates, when one of two markedly different populations may have produced that particular series.

### Limitations

This test provides a decision rule which is closely related to Hotelling's $T^2$-test (Test 18), hence is subject to the same limitations.

### Method

Using the notation of Hotelling's $T^2$-test, we may take samples from the two populations and obtain two quantities

$$D_A = a\bar{x}_A + b\bar{y}_A + c\bar{z}_A$$
$$D_B = a\bar{x}_B + b\bar{y}_B + c\bar{z}_B$$

for the two populations. From the series for which the origin has to be traced we can obtain a third quantity

$$D_S = a\bar{x}_S + b\bar{y}_S + c\bar{z}_S.$$

If $D_A - D_S < D_B - D_S$ we say that the series belongs to population A, but if $D_A - D_S > D_B - D_S$ we conclude that population B produced the series under consideration.

### Example

A discriminant function is produced for a collection of pre-historic dog bones. A new relic is found and the appropriate measurements are taken. There are two ancient populations of dog A or B to which the new bones could belong. To which population do the new bones belong? This procedure is normally performed by statistical computer software. The $D_A$ and $D_B$ values as well as the $D_S$ value are computed. The $D_S$ value is closer to $D_A$ and so the new dog bone relic belongs to population A.

### Numerical calculation

$a = 1.320, b = 0.972, c = 0.749$

$\bar{x}_A = 7, \bar{y}_A = 8, \bar{z}_A = 6, \bar{x}_B = 4.5, \bar{y}_B = 6, \bar{z}_B = 5$

$D_A = 1.320 \times 7 + 0.972 \times 8 + 0.749 \times 6 = 21.510$

$D_B = 1.320 \times 4.5 + 0.972 \times 6 + 0.749 \times 5 = 15.517$

If $\bar{x}_S = 6, \bar{y}_S = 6$ and $\bar{z}_S = 7$, then

$D_S = 1.320 \times 6 + 0.972 \times 6 + 0.749 \times 7 = 18.995$

$D_A - D_S = 21.510 - 18.995 = 2.515$

$D_B - D_S = 15.517 - 18.995 = -3.478$

$D_S$ lies closer to $D_A$. $D_S$ belongs to population A.

# Test 21 Dixon's test for outliers

## Object

To investigate the significance of the difference between a suspicious extreme value and other values in the sample.

## Limitations

1. The sample size should be greater than 3.
2. The population which is being sampled is assumed normal.

## Method

Consider a sample of size $n$, where the sample is arranged with the suspect value in front, its nearest neighbour next and then the following values arranged in ascending (or descending) order. The order is determined by whether the suspect value is the largest or the smallest. Denoting the ordered series by $x_1, x_2, \ldots, x_n$, the test statistic $r$ where

$$\begin{aligned}
r &= (x_2 - x_1)/(x_n - x_1) &\quad \text{if } 3 < n \leqslant 7,\\
r &= (x_2 - x_1)/(x_{n-1} - x_1) &\quad \text{if } 8 \leqslant n \leqslant 10,\\
r &= (x_3 - x_1)/(x_{n-1} - x_1) &\quad \text{if } 11 \leqslant n \leqslant 13,\\
r &= (x_3 - x_1)/(x_{n-2} - x_1) &\quad \text{if } 14 \leqslant n \leqslant 25.
\end{aligned}$$

Critical values for $r$ can be obtained from Table 8. The null hypothesis that the outlier belongs to the sample is rejected if the observed value of $r$ exceeds the critical value.

## Example

As part of a quality control programmed/implementation small samples are taken, at regular intervals, for a number of processes. On several of these processes there is the potential for inaccuracies occurring in the measurements that are taken due to the complexity of the measuring process and the inexperience of the process workers. One such sample of size 4 is tested for potential outliers and the following are produced: $x_1 = 326$, $x_2 = 177$, $x_3 = 176$, $x_4 = 157$.

Dixon's ratio yields $r = 0.882$.

The critical value at the 5 per cent level from Table 8 is 0.765, so the calculated value exceeds the critical value. We thus reject the null hypothesis that the outlier belongs to the sample. Thus we need to re-sample and measure again or only use three sample values in this case.

## Numerical calculation

$x_1, = 326,\ x_2 = 177,\ x_3 = 176,\ x_4 = 157,\ n = 4$

Here $r = \dfrac{x_2 - x_1}{x_n - x_1} = \dfrac{177 - 326}{157 - 326} = 0.882$

The critical value at $\alpha = 0.05$ is 0.765 [Table 8].
The calculated value exceeds the critical value.
Hence reject the null hypothesis that the value $x_1$ comes from the same population.

# Test 23 The $Z$-test for correlated proportions

## Object

To investigate the significance of the difference between two correlated proportions in opinion surveys. It can also be used for more general applications.

## Limitations

1. The same people are questioned both times on a yes–no basis.
2. The sample size must be quite large.

## Method

$N$ people respond to a yes–no question both before and after a certain stimulus. The following two-way table can then be built up:

| | | First poll | | |
|---|---|---|---|---|
| | | Yes | No | |
| Second poll | Yes | $a$ | $b$ | |
| | No | $c$ | $d$ | |
| | | | | $N$ |

To decide whether the stimulus has produced a significant change in the proportion answering 'yes', we calculate the test statistic

$$Z = \frac{b - c}{N\sigma}$$

where

$$\sigma = \sqrt{\frac{(b + c) - (b - c)^2/N}{N(N - 1)}}.$$

## Example

Sampled panels of potential buyers of a financial product are asked if they might buy the product. They are then shown a product advertisement of 30 seconds duration and asked again if they would buy the product. Has the advertising stimulus produced a significant change in the proportion of the panel responding 'yes'?

We have

| | | First poll | |
|---|---|---|---|
| | | Yes | No |
| Second poll | Yes | 30 | 15 |
| | No | 9 | 51 |

which yields the test statistic $Z = 1.23$. The 5 per cent critical value from the normal distribution is 1.96. Since 1.23 is less than 1.96 we do not reject the null hypothesis of no difference. The advertisement does not increase the proportion saying 'yes'. Notice that we have used a one-tailed test, here, because we are only interested in an increase, i.e. a positive effect of advertising.

### Numerical calculation

$a = 30, b = 15, c = 9, d = 51, N = 105$

The null hypothesis is that there is no apparent change due to the stimulus. The difference in proportion is

$$\frac{b-c}{N} = \frac{15}{105} - \frac{9}{105} = \frac{6}{105} = 0.0571$$

$$\sigma = \sqrt{\frac{(15+9)-(15-9)^2/105}{105 \times 104}} = 0.0465$$

$$Z = \frac{0.0571}{0.0465} = 1.23$$

The critical value at $\alpha = 0.05$ is 1.96 [Table 1].
The calculated value is less than the critical value.
Do not reject the null hypothesis.

# Test 24 $\chi^2$-test for an assumed population variance

### Object

To investigate the significance of the difference between a population variance $\sigma^2$ and an assumed value $\sigma_0^2$.

### Limitations

It is assumed that the sample is taken from a normal population.

### Method

The sample variance

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

is calculated. The test statistic is then

$$\chi^2 = \frac{s^2}{\sigma_0^2}(n-1)$$

which follows a $\chi^2$-distribution with $n-1$ degrees of freedom.

### Example

An engineering process has specified variance for a machined component of 9 square cm. A sample of 25 components is selected at random from the production and the mean value for a critical dimension on the component is measured at 71 cm with sample variance of 12 square cm. Is there a difference between variances? A calculated chi-squared value of 32 is less than the tabulated value of 36.4 suggesting no difference between variances.

### Numerical calculation

$n = 25, \bar{x} = 71, s^2 = 12, \sigma_0^2 = 9$

$H_0\colon \sigma^2 = \sigma_0^2, H_1\colon \sigma^2 \neq \sigma_0^2$

$\chi^2 = 24 \times \frac{12}{9} = 32$

Critical value $\chi^2_{24;\,0.05} = 36.4$ [Table 5].
Do not reject the null hypothesis. The difference between the variances is not significant.

# Test 25 *F*-test for two counts (Poisson distribution)

### Object

To investigate the significance of the difference between two counted results (based on a Poisson distribution).

### Limitations

It is assumed that the counts satisfy a Poisson distribution and that the two samples were obtained under similar conditions.

### Method

Let $\mu_1$ and $\mu_2$ denote the means of the two populations and $N_1$ and $N_2$ the two counts. To test the hypothesis $\mu_1 = \mu_2$ we calculate the test statistic

$$F = \frac{N_1}{N_2 + 1}$$

which follows the $F$-distribution with $(2(N_2 + 1), 2N_1)$ degrees of freedom. When the counts are obtained over different periods of time $t_1$ and $t_2$, it is necessary to compare the counting rates $N_1/t_1$ and $N_2/t_2$. Hence the appropriate test statistic is

$$F = \frac{\dfrac{1}{t_1}(N_1 + 0.5)}{\dfrac{1}{t_2}(N_1 + 0.5)}$$

which follows the $F$-distribution with $(2N_1 + 1, 2N_2 + 1)$ degrees of freedom.

### Example

Two automated kiln processes (producing baked plant pots) are compared over their standard cycle times, i.e. 4 hours. Kiln 1 produced 13 triggered process corrections and kiln 2 produced 3 corrections. What can we say about the two kiln mean correction rates, are they the same? The calculated $F$ statistic is 3.25 and the critical value from Table 3 is 2.32. Since the calculated value exceeds the critical value we conclude that there is a statistical difference between the two counts. Kiln 1 has a higher error rate than kiln 2.

### Numerical calculation

$N_1 = 13, N_2 = 3, t_1 = t_2$

$f_1 = 2(N_2 + 1) = 2(3 + 1) = 8, f_2 = 2N_1 = 2 \times 13 = 26$

$$F = \frac{N_1}{N_2 + 1} = \frac{13}{3 + 1} = 3.25$$

Critical value $F_{8,26;\,0.05} = 2.32$ [Table 3].
The calculated value exceeds the table value.
Hence reject the null hypothesis.

# Test 33 The $w/s$-test for normality of a population

## Object

To investigate the significance of the difference between a frequency distribution based on a given sample and a normal frequency distribution.

## Limitations

This test is applicable if the sample is taken from a population with continuous distribution.

## Method

This is a much simpler test than Fisher's cumulant test (Test 20). The sample standard deviation ($s$) and the range ($w$) are first determined. Then the Studentized range $q = w/s$ is found.

The test statistic is $q$ and critical values are available for $q$ from Table 14. If the observed value of $q$ lies outside the two critical values, the sample distribution cannot be considered as a normal distribution.

## Example

For this test of normality we produce the ratio of sample range divided by sample standard deviation and compare with critical values from Table 14.

We have two samples for consideration. They are taken from two fluid injection processes. The two test statistics, $q_1$ and $q_2$, are both within their critical values. Hence we accept the null hypothesis that both samples could have been taken from normal distributions. Such tests are particularly relevant to quality control situations.

## Numerical calculation

$$n_1 = 4,\ n_2 = 9,\ \bar{x}_1 = 3166,\ \bar{x}_2 = 2240.4,\ \alpha = 0.025$$

$$s_1^2 = 6328.67,\ s_2^2 = 221\,661.3,\ s_1 = 79.6,\ s_2 = 471$$

$$w_1 = 171,\ w_2 = 1333$$

$$q_1 = \frac{w_1}{s_1} = 2.15,\ q_2 = \frac{w_2}{s_2} = 2.83$$

Critical values for this test are:
for $n_1 = 4$, 1.93 and 2.44 [Table 14];
for $n_2 = 9$, 2.51 and 3.63 [Table 14].
Hence the null hypothesis cannot be rejected.

# Test 34 Cochran's test for variance outliers

### Object

To investigate the significance of the difference between one rather large variance and $K - 1$ other variances.

### Limitations

1. It is assumed that the $K$ samples are taken from normally distributed populations.
2. Each sample is of equal size.

### Method

The test statistic is

$$C = \frac{\text{largest of the } s_i^2}{\text{sum of all } s_i^2}$$

where $s_i^2$ denotes the variance of the $i$th sample. Critical values of $C$ are available from Table 15. The null hypothesis that the large variance does not differ significantly from the others is rejected if the observed value of $C$ exceeds the critical value.

### Example

In a test for the equality of $k$ means (analysis of variance) it is assumed that the $k$ populations have equal variances. In this situation a quality control inspector suspects that errors in data recording have led to one variance being larger than expected. She performs this test to see if her suspicions are well founded and, therefore, if she needs to repeat sampling for this population (a machine process line). Her test statistic, $C = 0.302$ and the 5 per cent critical value from Table 15 is 0.4241. Since the test statistic is less than the critical value she has no need to suspect data collection error since the largest variance is not statistically different from the others.

### Numerical calculation

$s_1^2 = 26,\ s_2^2 = 51,\ s_3^2 = 40,\ s_4^2 = 24,\ s_5^2 = 28$

$n_1 = n_2 = n_3 = n_4 = n_5 = 10,\ K = 5, \nu = n - 1 = 9$

$$C = \frac{51}{26 + 51 + 40 + 24 + 28} = 0.302$$

Critical value $C_{9;0.05} = 0.4241$ [Table 15].
The calculated value is less than the critical value.
Do not reject the null hypothesis.

# Test 49 The Wilcoxon inversion test (*U*-test)

### Object

To test if two random samples could have come from two populations with the same frequency distribution.

### Limitations

It is assumed that the two frequency distributions are continuous and that the two samples are random and independent.

### Method

Samples of size $n_1$ and $n_2$ are taken from the two populations. When the two samples are merged and arranged in ascending order, there will be a number of jumps (or inversions) from one series to the other. The smaller of the number of inversions and the number of non-inversions forms the test statistic, $U$. The null hypothesis of the same frequency distribution is rejected if $U$ exceeds the critical value obtained from Table 20.

### Example

An educational researcher has two sets of adjusted reading scores for two sets of five pupils who have been taught by different methods. It is possible that the two samples could have come from the same population frequency distribution.

The collected data produce a calculated $U$ value of 4. Since the sample $U$ value equals the tabulated critical value the educational researcher rejects the null hypothesis of no difference. The data suggest that the two reading teaching methods produce different results.

### Numerical calculation

$n_1 = 5, n_2 = 5, \alpha = 0.05$

| | | | | | |
|---|---|---|---|---|---|
| $x_i$ | 11.79 | 11.21 | 13.20 | 12.66 | 13.37 |
| $y_i$ | 10.34 | 11.40 | 10.19 | 12.10 | 11.46 |

Rearrangement gives the following series

$$10.19,\ 10.34,\ \frac{11.21}{(2)},\ 11.40,\ 11.46,\ \frac{11.79}{(4)},\ 12.10,\ \frac{12.66}{(5)},\ \frac{13.20}{(5)},\ \frac{13.37}{(5)}$$

where underlined values come from the first row ($x_i$). Below these underlines, the corresponding number of inversions, i.e. the number of times a $y$-value comes after an $x$-value, is given in parentheses.

The number of inversions is $2 + 4 + 5 + 5 + 5 = 21$.

The number of non-inversions is $n_1 n_2 - 21 = 25 - 21 = 4$.

The critical value at $\alpha = 0.05$ is 4 [Table 20].

The sample value of $U$ is equal to the critical value.

The null hypothesis may be rejected; alternatively, the experiment could be repeated by collecting a second set of data.

# Test 52 The Wilcoxon–Mann–Whitney rank sum test of two populations

## Object

To test if two random samples could have come from two populations with the same mean.

## Limitations

It is assumed that the two populations have continuous frequency distributions with the same shape and spread.

## Method

The results of the two samples $x$ and $y$ are combined and arranged in order of increasing size and given a rank number. In cases where equal results occur the mean of the available rank numbers is assigned. The rank sum $R$ of the smaller sample is now found. Let $N$ denote the size of the combined samples and $n$ denote the size of the smaller sample.

A second quantity

$$R^1 = n(N + 1) - R$$

is now calculated. The values $R$ and $R^1$ are compared with critical values obtained from Table 21. If either $R$ or $R^1$ is less than the critical value the null hypothesis of equal means would be rejected.

*Note* If the samples are of equal size, then the rank sum $R$ is taken as the smaller of the two rank sums which occur.

## Example

A tax inspector wishes to compare the means of two samples of expenses claims taken from the same company but separated by a period of time (the values have been adjusted to account for inflation). Are the mean expenses for the two periods the same? He calculates a test statistic, $R^1$ of 103 and compares this with the tabulated value of 69. Since the calculated value is greater than the tabulated critical value he concludes that the mean expenses have not changed.

## Numerical calculation

| | | | | | | | | | | | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $x$ | 50.5 | 37.5 | 49.8 | 56.0 | 42.0 | 56.0 | 50.0 | 54.0 | 48.0 | | |
| Rank | 9 | 1 | 7 | 15.5 | 2 | 15.5 | 8 | 13 | 6 | | 77 |
| $y$ | 57.0 | 52.0 | 51.0 | 44.2 | 55.0 | 62.0 | 59.0 | 45.2 | 53.5 | 44.4 | |
| Rank | 17 | 11 | 10 | 3 | 14 | 19 | 18 | 5 | 12 | 4 | 113 |

$R = 77$, $n_1 = 9$, $n_2 = 10$, $N = 19$, $R^1 = 9 \times 20 - 77 = 103$
The critical value at $\alpha = 0.05$ is 69 [Table 21].
Hence there is no difference between the two means.

# Test 58 The Spearman rank correlation test (paired observations)

## Object

To investigate the significance of the correlation between two series of observations obtained in pairs.

## Limitations

It is assumed that the two population distributions are continuous and that the observations $x_i$ and $y_i$ have been obtained in pairs.

## Method

The $x_i$ observations are assigned the rank numbers $1, 2, \ldots, n$ in order of increasing magnitude. A similar procedure is carried out for all the $y_i$ observations. For each pair of observations, the difference in the ranks, $d_i$, can be determined. The quantity $R = \sum_{i=1}^{n} d_i^2$ is now calculated.

For large samples ($n > 10$) the test statistic is

$$Z = \frac{6R - n(n^2 - 1)}{n(n + 1)\sqrt{(n - 1)}}$$

which may be compared with tables of the standard normal distribution. For small samples, the test statistic

$$r_S = 1 - \frac{6R}{n(n^2 - 1)}$$

must be compared with critical values obtained from Table 26. In both cases, if the experimental value lies in the critical region one has to reject the null hypothesis of no correlation between the two series.

## Example

A panel of consumers is asked to rate two brands of vegetarian sausage. It is hoped that advertising can be combined in a mail out to potential consumers. A small sample is taken and panel members are asked to rate each brand. The results produce a $Z$ value of $-2.82$. The critical value for $Z$ is 1.64 so the null hypothesis of zero correlation is rejected. Consumers tend to report similar preferences for the two brands of sausage.

## Numerical calculation

$d_i$: $0, -1, -2, 0, +3, -1, -1, +2, 0, 0, 2$
Hence $R = 24$, $n = 11$

$$Z = \frac{6 \times 24 - 11(11^2 - 1)}{11 \times 12\sqrt{10}} = \frac{144 - 1320}{132 \times 3.1623} = \frac{-1176}{417.42} = -2.82$$

The critical value of $Z$ at $\alpha$ = 0.05 is 1.64 [Table 1].
Hence reject the null hypothesis.

# Test 59 The Kendall rank correlation test (paired observations)

## Object

To investigate the significance of the correlation between two series of observations obtained in pairs.

## Limitations

It is assumed that the two population distributions are continuous and that the observations $x_i$, and $y_i$, have been obtained in pairs.

## Method

The $x_i$ observations are assigned the rank numbers $1, 2, \ldots, n$ in order of increasing magnitude. A similar procedure is carried out for all the $y_i$ observations. Each of the possible pairs of rank numbers (there will be $\frac{1}{2}n(n-1)$ of these) is now examined. Each pair $(x_i, y_i)$ will be compared successively and systematically with each other pair $(x_j, y_j)$. When $x_i - x_j$ and $y_i - y_j$ have the same sign a score of $+1$ is obtained. When they have opposite signs a score of $-1$ is obtained. When there is a difference of zero, no score is obtained. These scores are summed together and this sum is denoted $S$. In this manner we can work with observational results without having determined the rank numbers.

For large $n$ ($n > 10$), $Z$ follows a normal distribution and hence the test statistic

$$Z = \frac{S}{\{n(n-1)(2n+5)/18\}^{\frac{1}{2}}}$$

may be compared with tables of the standard normal distribution. For small samples, critical values of $S$ may be obtained from Table 27.

In both cases, if the experimental value lies in the critical region one has to reject the null hypothesis of no correlation between the two series.

## Example

A tax inspector wishes to investigate whether there is any correlation between total investment incomes (£00's), obs1 and total additional income (£00's), obs 2. He has collected a sample of 10 tax forms and calculates an $S$ value of 33. He compares this with the critical value of 21 obtained from Table 27. Since the calculated value is greater than the tabulated value he concludes that there is a significant correlation.

## Numerical calculation

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Observation 1 | 7.1 | 8.3 | 10.7 | 9.4 | 12.6 | 11.1 | 10.3 | 13.1 | 9.6 | 12.4 |
| Observation 2 | 62 | 66 | 74 | 74 | 82 | 76 | 72 | 79 | 68 | 74 |
| Plus scores | 9 | 8 | 5 | 3 | 4 | 3 | 3 | 2 | 1 | 0 |
| Minus scores | 0 | 0 | 0 | 2 | 1 | 1 | 1 | 0 | 0 | 0 |

Total plus scores = 38, total minus scores = 5
$S = 38 - 5 = 33$, $n = 10$
Critical value $S_{10;005} = 21$ [Table 27].
The calculated value is greater than the critical value.
Reject the null hypothesis.

# Test 82 *F*-test for testing linearity of regression

## Object

To test the linearity of regression between an $X$ variable and a $Y$ variable.

## Limitations

For given $X$, the $Y$s are normally and independently distributed. The error terms are normally and independently distributed with mean zero.

## Method

Once the relationship between $X$ and $Y$ is established using Test 81, we would further like to know whether the regression is linear or not. Under the same set-up as Test 81, we are interested in testing:

$$H_0: \mu_i = \alpha + \beta X_i, \quad i = 1, 2, \ldots, n,$$

Under $H_0$,

$$s_{\mathrm{E}}^2 = \sum_i (y_i - \bar{y}) - b^2 \sum_i n_i (x_i - \bar{x})^2,$$

with $n - 2$ degrees of freedom, and the sum of squares due to regression

$$s_{\mathrm{R}}^2 = b^2 \sum_i n_i (x_i - \bar{x})^2,$$

with 1 degree of freedom. The ratio of mean squares
$F = \bar{s}_{\mathrm{R}}^2 / \bar{s}_{\mathrm{E}}^2$ is used to test $H_0$ with $(1, n - 2)$ degrees of freedom.

## Example

In a chemical reaction the quantity of plastic polymer ($Y$) is measured at each of four levels of an enzyme additive ($X$). The experiment is repeated three times at each level of $X$ to enable a test of linearity of regression to be performed.

The data produce an $F$ value of 105.80 and this is compared with the critical $F$ value of 4.96 from Table 3. Since the critical value is exceeded we conclude that there is a significant regression.

## Numerical calculation

| $i$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $x_i$ | 150 | 150 | 150 | 200 | 200 | 200 | 250 | 250 | 250 | 300 | 300 | 300 |
| $y_i$ | 77.4 | 76.7 | 78.2 | 84.1 | 84.5 | 83.7 | 88.9 | 89.2 | 89.7 | 94.8 | 94.7 | 95.9 |

$n = 12, n - 2 = 10$. For $\beta = 0$, test $H_0$: $\beta = 0$ against $H_1$: $\beta \neq 0$

The total sum of squares is

$$\sum y_i^2 - \left(\sum y_i\right)^2 \Big/ n = 513.1167,$$

and

$$s_{\mathrm{R}}^2 = b\frac{\left(\sum x_i y_i - \dfrac{1}{n}\sum x_i y_i\right)^2}{\sum x_i^2 - \dfrac{1}{n}(x_i)^2} = 509.10,$$

$s_{\mathrm{E}}^2 = 4.0117$, $\bar{s}_{\mathrm{R}}^2 = 42.425$, $\bar{s}_{\mathrm{E}}^2 = 0.401$, $F = 105.80$
Critical value $F_{1,10;\,0.05} = 4.96$ [Table 3].
Hence reject the null hypothesis and conclude that $\beta \neq 0$.