



Tutorials and worked examples for simulation,
curve fitting, statistical analysis, and plotting.

<https://simfit.org.uk>

<https://simfit.silverfrost.com>

Hypothesis testing is based upon specifying a null hypothesis H_0 then testing to see if a statistic calculated from the data is sufficiently extreme to justify rejecting the null hypothesis. There are two possible errors.

- **Type I error**

The null hypothesis is rejected when it is true and the probability of this happening is α .

- **Type II error**

The null hypothesis is accepted when it is false and the probability of this happening is β .

The significance level is α while the power is $1 - \beta$, often expressed as a percentage. The situation can be summarized in the following table.

Decision	H_0 is true	H_0 is false
Reject H_0	Type I error α	Correct $1 - \beta$
Accept H_0	Correct $1 - \alpha$	Type II error β

Calculations related to power as a function of sample size can be performed as long as the statistical distributions and parameters required for the null hypothesis are correct and specified. Unfortunately, while calculation of α is straightforward, calculation of β requires that an alternative hypothesis H_A be specified and can be much more difficult.

In any given situation it may be necessary to estimate the sample size n required given α and β , or to estimate β given α and n , as it is not possible to simultaneously minimize α and β . If n is fixed, then increasing α decreases β , while decreasing α increases β . The following cases are discussed here.

1. One binomial sample
2. Two binomial samples
3. One normal sample
4. Two normal samples
5. Multiple normal samples
6. One and two variances
7. One and two correlations
8. The chi-squared test

For each of these cases, the minimum essential theoretical details are given followed by typical examples. Plots of power as a function of sample size can also be created. Finally, a more comprehensive description of the underlying theory is given.

1. Power calculations for 1 binomial sample

The calculations are based on the binomial test, the binomial distribution, and the normal approximation to it for large samples and p not close to 0 or 1.

If the theoretical binomial parameters p_0 and $q_0 = 1 - p_0$ are not too close to 0 or 1 and it is wished to estimate this with an error of at most δ , then the sample size required is

$$n = \frac{Z_{\alpha/2}^2 p_0 q_0}{\delta^2},$$

where $P(Z > Z_{\alpha/2}) = \alpha/2$,

or $\Phi(Z_{\alpha/2}) = 1 - \alpha/2$,

which, for many purposes, can be approximated by $n \approx 1/\delta^2$. The power in a binomial or sign test can be approximated, again if the sample estimates p_1 and $q_1 = 1 - p_1$ are not too close to 0 or 1, by

$$1 - \beta = P\left(Z < \frac{p_1 - p_0}{\sqrt{p_0 q_0/n}} - Z_{\alpha/2} \sqrt{\frac{p_1 q_1}{p_0 q_0}}\right) + P\left(Z > \frac{p_1 - p_0}{\sqrt{p_0 q_0/n}} + Z_{\alpha/2} \sqrt{\frac{p_1 q_1}{p_0 q_0}}\right).$$

Example 1

This demonstrates calculations that are possible when, in a sample of size n , x successes are used to estimate the binomial parameter $\hat{p} = x/n$. Given a binomial distribution with H_0 : parameter $p = p_0$, H_A : parameter $p = p_1$, and α specified, then three calculations are possible, namely

1. Calculate $n(\delta)$, i.e. n giving an error at most δ
2. Calculate $n(\beta)$, i.e. n given β
3. Calculate $\beta(n)$, i.e. β given n

as summarized in the next table.

Example 1: Power analysis for 1 binomial sample

For $n(\delta)$	$\alpha = 0.05$	$p = 0.5$	$\delta = 0.1$		$n = 96$
For $n(\beta)$	$\alpha = 0.05$	$\beta = 0.2$	$p_0 = 0.5$	$p_1 = 0.6$	$n = 192$
For $\beta(n)$	$\alpha = 0.05$	$p_0 = 0.5$	$p_1 = 0.6$	power = 80%	$n = 192$

The conclusion is that a sample size of 96 is required to ensure that the parameter estimated is within 0.1 of the true parameter $p = 0.5$ for 95% of the results from repeated samples, while to confirm that the parameter is distinct from $p = 0.6$ with 80% power requires a sample size of 192.

2. Power calculations for 2 binomial samples

For two sample proportions p_1 and p_2 that are similar and not too close to 0 or 1, as in a 2 by 2 contingency table, the sample size n and power $1 - \beta$ associated with a binomial test for $H_0 : p_{01} = p_{02}$ can be estimated using one of numerous methods based upon normal approximations. For example

$$n = \frac{(p_1 q_1 + p_2 q_2)(Z_{\alpha/2} + Z_{\beta})^2}{(p_1 - p_2)^2},$$

$$Z_{\beta} = \sqrt{\frac{n(p_1 - p_2)^2}{p_1 q_1 + p_2 q_2}} - Z_{\alpha/2},$$

$$\beta = P(Z \geq Z_{\beta}),$$

$$1 - \beta = \Phi(Z_{\beta}).$$

Example 2a

This deals with the situation where two samples of size n are analyzed to determine if the estimates $\hat{p}_1 = x_1/n$ and $\hat{p}_2 = x_2/n$ differ significantly. Parameters can also be input as log odds ratios.

Example 2a. Power analysis for 2 binomial samples

$\beta(n)$	$\alpha = 0.05$	$p_1 = 0.6$	$p_2 = 0.7$	power = 32%	$n = 100$
$\beta(n)$	$\alpha = 0.05$	$p_1 = 0.6$	$p_2 = 0.7$	power = 56%	$n = 200$
$\beta(n)$	$\alpha = 0.05$	$p_1 = 0.6$	$p_2 = 0.7$	power = 73%	$n = 300$
$n(\beta)$	$\alpha = 0.05$	$\beta = 0.2$	$p_1 = 0.6$	$p_2 = 0.7$	$n = 353$

Note that, for $p_1 = 0.6$ and $p_2 = 0.7$ the power increases from 32%, to 56%, to 73% as n increases from 100, to 200, to 300, while a sample of size $n = 353$ is required to achieve 80% power.

Example 2b

Power for the Fisher exact test with sample size n used to estimate both p_1 and p_2 , as for the binomial test, can be calculated using

$$1 - \beta = 1 - \sum_{r=0}^{2n} \sum_{C_r} \binom{n}{x} \binom{n}{r-x},$$

where $r =$ total successes,

$x =$ number of successes in the group,

and $C_r =$ the critical region.

This can be inverted by SIMFIT to estimate n , but unfortunately the sample sizes required may be too large to implement by the normal procedure of enumerating probabilities for all 2 by 2 contingency tables with consistent marginals.

Example 2b. Power analysis for the Fisher Exact Test

$\beta(n)$	$\alpha = 0.05$	$p_1 = 0.6$	$p_2 = 0.7$	power = 37%	$n = 100$
$\beta(n)$	$\alpha = 0.05$	$p_1 = 0.6$	$p_2 = 0.7$	power = 64%	$n = 200$
$n(\beta)$	$\alpha = 0.05$	$\beta = 0.2$	$p_1 = 0.6$	$p_2 = 0.7$	$n = 304$

For sample sizes of 100 and 200 the power is 37% and 64% but sample sizes of 304 are required for 80% power.

3. Power calculations for 1 normal sample

The calculations for a 1 sample t test are based upon the confidence limit formula for the population mean μ from a sample of size n , using the sample mean \bar{x} , sample variance s^2 and the t distribution, as follows

$$P\left(\bar{x} - t_{\alpha/2, n-1} \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{\alpha/2, n-1} \frac{s}{\sqrt{n}}\right) = 1 - \alpha,$$

$$\text{where } \bar{x} = \sum_{i=1}^n x_i / n,$$

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

$$P(t \leq t_{\alpha/2, \nu}) = 1 - \alpha/2,$$

$$\text{and } \nu = n - 1.$$

You input the sample variance, which should be calculated using a sample size comparable to those predicted above. Power calculations can be done using the half width $h = t_{\alpha/2, n-1} s / \sqrt{n}$, or using the absolute difference δ between the population mean and the null hypothesis mean as argument. The following options are available:

- $n(h)$, i.e. to calculate the sample size necessary to estimate the true mean within a half width h

$$n = \frac{s^2 t_{\alpha/2, n-1}^2}{h^2};$$

□ $n(\delta)$, i.e. to calculate the sample size necessary for an absolute difference δ

$$n = \frac{s^2}{\delta^2} (t_{\alpha/2, n-1} + t_{\beta, n-1})^2; \text{ or}$$

□ $\beta(n)$, i.e. to estimate the power

$$t_{\beta, n-1} = \frac{\delta}{\sqrt{s^2/n}} - t_{\alpha/2, n-1}.$$

It should be noted that the sample size occurs in the degrees of freedom for the t distribution, necessitating an iterative solution to estimate n .

Example 3

Example 3. Power analysis for 1 sample t test

$n(h)$	$h = 1$	$\alpha = 0.05$		$s^2 = 1$	$n = 7$
$n(\delta)$	$\delta = 1$	$\alpha = 0.05$	$\beta = 0.2$	$s^2 = 1$	$n = 10$
$\delta(n, \beta)$	$n = 10$	$\alpha = 0.05$	$\beta = 0.2$	$s^2 = 1$	$\delta = 0.9947$
$\beta(n)$	$n = 10$	$\delta = 1$	$\alpha = 0.05$	$s^2 = 1$	$\beta = 0.1958$

4. Power calculations for 2 normal samples

These calculations are based upon the same type of t test approach as just described for 1 normal sample, except that the pooled variance s_p^2 should be input as the estimate for the common variance σ^2 , i.e.,

$$s_p^2 = \frac{\sum_{i=1}^{n_x} (x_i - \bar{x})^2 + \sum_{j=1}^{n_y} (y_j - \bar{y})^2}{n_x + n_y - 2}$$

where X has sample size n_x and Y has sample size n_y . The following options are available:

- To calculate the sample size necessary to estimate the difference between the two population means within a half width h

$$n = \frac{2s_p^2 t_{\alpha/2, 2n-2}^2}{h^2};$$

- To calculate the sample size necessary to detect an absolute difference δ between population means

$$n = \frac{2s_p^2}{\delta^2} (t_{\alpha/2, 2n-2} + t_{\beta, 2n-2})^2; \text{ or}$$

- To estimate the power

$$t_{\beta, 2n-2} = \frac{\delta}{\sqrt{2s_p^2/n}} - t_{\alpha/2, 2n-2}.$$

The t test has maximum power when $n_x = n_y$ but, if the two sample sizes are unequal, calculations based on the the harmonic mean n_h should be used, i.e.,

$$n_h = \frac{2n_x n_y}{n_x + n_y},$$

so that $n_y = \frac{n_h n_x}{2n_x - n_h}.$

Example 4

In order to perform calculations it is necessary to assume that both samples are from normal distributions with the same variance and then input those of the following parameters as required.

- The significance level α
- An accurate estimate for the common variance s^2
- Choice of a 2-tail test or 1-tail test
- The half width h to determine a 95% confidence range $2h$ for the difference between the sample means
- The minimum absolute difference δ between the sample means that can be detected
- The power $100(1 - \beta)\%$
- The sample size n

The following table was created using the analysis of power and sample size option from the statistical calculations procedure available from the [Statistics] menu on the SIMFIT main menu, or by using the [A/Z] option to open program **simstat**.

Example 4. Power analysis for the t test

$n(h)$	$h = 1$	$\alpha = 0.05$		$s^2 = 1$	$n = 9$
$n(\delta)$	$\delta = 1$	$\alpha = 0.05$	$\beta = 0.2$	$s^2 = 1$	$n = 17$
$\delta(n)$	$n = 17$	$\alpha = 0.05$	$\beta = 0.2$	$s^2 = 1$	$\delta = 0.9912$
$\beta(n)$	$n = 17$	$\delta = 1$	$\alpha = 0.05$	$s^2 = 1$	$\beta = 0.1931$
$n(h)$	$h = 0.5$	$\alpha = 0.05$		$s^2 = 0.5193$	$n = 18$
$n(\delta)$	$\delta = 0.5$	$\alpha = 0.05$	$\beta = 0.1$	$s^2 = 0.5193$	$n = 45$
$\beta(n)$	$n = 15$	$\delta = 1$	$\alpha = 0.05$	$s^2 = 0.5193$	$\beta = 0.0454$

The last three entries in the above table would be typical. They are for two samples of size $n = 15$ with pooled variance $s^2 = 0.5193$, and the results would be interpreted as follows.

- $n(h)$ shows that a sample size of $n = 18$ would be required to have a 95% confidence interval for the difference between the true means no larger than 1, that is with $h = 0.5$.
- $n(\delta)$ illustrates that a sample size of $n = 45$ is necessary in order for a 90% chance of detecting a difference δ between the true means as small as 0.5.
- $\beta(n)$ demonstrate that the power for detecting a difference of $\delta = 1$ between the true means has a power of 95.46%.

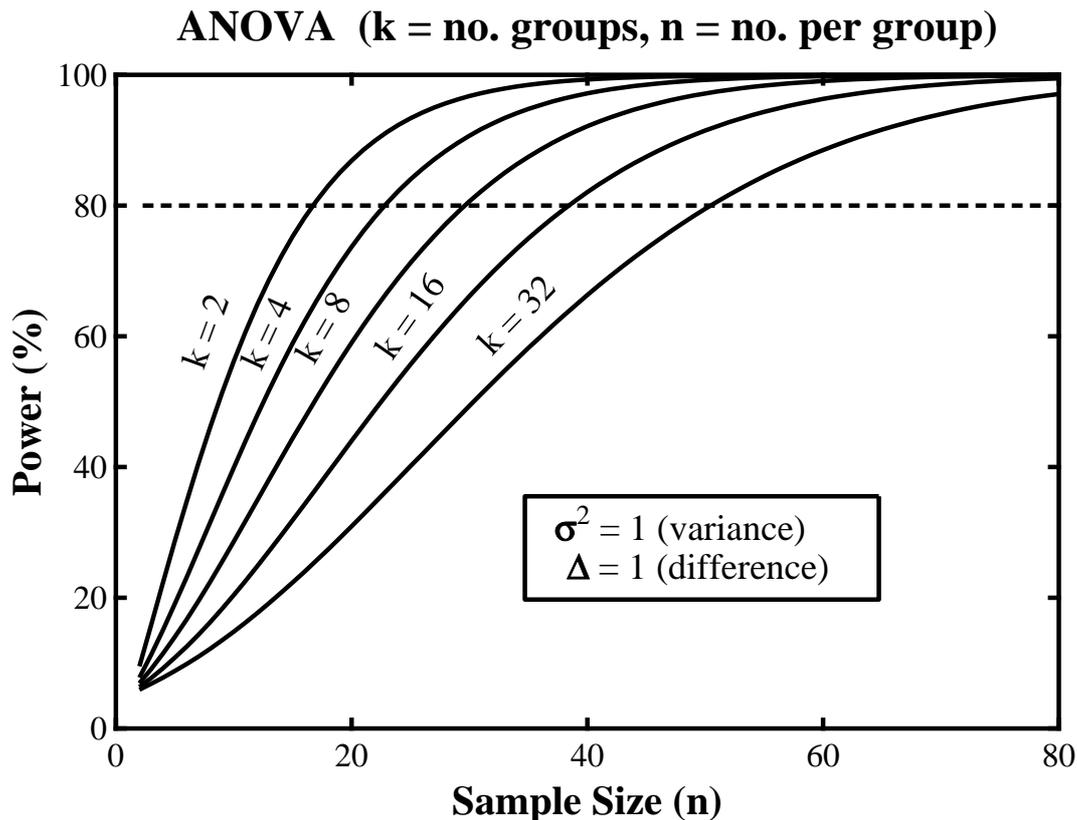
5. Power calculations for k normal samples

The calculations are based on the 1-way analysis of variance technique, i.e. ANOVA. Note that the SIMFIT power as a function of sample size procedure also allows you to plot power as a function of sample size, which is particularly useful with ANOVA designs where the number of columns k can be of interest, in addition to the number per sample n . The power calculation involves the F and non-central F distributions and you calculate the required n values by using graphical estimation to obtain starting estimates for the iteration. If you choose a n value that is sufficient to make the power as a function on n plot cross the critical power, the program then calculates the power for sample sizes adjacent to the intersection, which is of use when studying k and n for ANOVA.

Example 5

All the power procedures available in SIMFIT provide the ability to plot power as a function of sample size but this particularly useful with ANOVA, as will now be explained.

It is important in the design of experiments to be able to estimate the sample size needed to detect a significant effect. For such calculations you must specify all the parameters of interest except one, then calculate the unknown parameter using numerical techniques. For example, the problem of deciding whether one or more samples differ significantly is a problem in the Analysis of Variance, as long as the samples are all normally distributed and with the same variance. You specify the known variance, σ^2 , the minimum detectable difference between means, Δ , the number of groups, k , the significance level, α , and the sample size per group, n . Then, using nonlinear equations involving the F and noncentral F distributions, the power, $100(1 - \beta)$ can be calculated. It can be very confusing trying to understand the relationship between all of these parameters so, in order to obtain an impression of how these factors alter the power, a graphical technique is very useful, as in this figure.



simstat was used to create this graph. The variance, significance level, minimum detectable difference and number of groups were fixed, then power was plotted as a function of sample size. The ASCII text coordinate files from several such plots were collected together into a library file to compose the joint plot using **simplot**. Note that, if a power plot reaches the current power level of interest, the critical power level is plotted (80% in the above plot) and the n values either side of the intersection point are displayed.

6. Power calculations for 1 and 2 variances

The calculations depend on the fact that, for a sample of size n from a normal distribution with true variance σ_0^2 , the function χ^2 defined as

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

is distributed as a chi-square variable with $n-1$ degrees of freedom. Also, given variance estimates s_x^2 and s_y^2 obtained with sample sizes n_x and n_y from the same normal distribution, the variance ratio F defined as

$$F = \max\left(\frac{s_x^2}{s_y^2}, \frac{s_y^2}{s_x^2}\right)$$

is distributed as an F variable with either n_x, n_y or n_y, n_x degrees of freedom. If possible n_x should equal n_y , of course. The 1-tailed options available are:

□ $H_0 : \sigma^2 \leq \sigma_0^2$ against $H_1 : \sigma^2 > \sigma_0^2$

$$1 - \beta = P(\chi^2 \geq \chi_{\alpha, n-1}^2 \sigma_0^2 / s^2);$$

□ $H_0 : \sigma^2 \geq \sigma_0^2$ against $H_1 : \sigma^2 < \sigma_0^2$

$$1 - \beta = P(\chi^2 \leq \chi_{1-\alpha, n-1}^2 \sigma_0^2 / s^2); \text{ or}$$

□ Rearranging the samples, if necessary, so that $s_x^2 > s_y^2$ then

$$H_0 : \sigma_x^2 = \sigma_y^2 \text{ against } H_1 : \sigma_x^2 \neq \sigma_y^2$$

$$Z_\beta = \sqrt{\frac{2m(n_y - 2)}{m + 1}} \log\left(\frac{s_x^2}{s_y^2}\right) - Z_\alpha$$

$$\text{where } m = \frac{n_x - 1}{n_y - 1}.$$

Example 6a

This example shows the results from performing a one-tail test on a variance estimate s_1^2 of 2.6898 obtained with a sample size of 8 compared to a theoretical value of σ_0^2 of 1.5. A confidence interval for the sample variance is calculated as well as the actual power, followed by the sample size needed for 90% power.

Example 6a: Test for $H_0 : \sigma^2 \leq 1.5, H_1 : \sigma^2 > 1.5$

Sample variance s_1^2	2.6898
Sample size used to calculate s_1^2	8
95% confidence interval for s_1^2	1.1758, 11.142
Test statistic (C)	12.552
$P(\chi^2 \geq C)$	0.0838
Power of this test	34.6486% (for $\alpha = 0.05$)
Minimal sample size	51 (for $\alpha = 0.05, \beta = 0.1$)

Example 6b

This example explores the power as a function of sample size for a two-tail variance ratio test using samples with variances $s_1^2 = 21.87 (n_1 = 11)$ and $s_2^2 = 15.36 (n_2 = 8)$, i.e. for the null hypothesis $H_0 : \sigma_1^2 = \sigma_2^2$. Details for two-tail tests are also given for the additional hypothetical cases with the same variances but estimated with sample sizes $n_1 = n_2 = 60$ and $n_1 = 20, n_2 = 30$.

Example 6b. Two-tailed variance ratio (F) test	
Sample variance s_1^2	21.87
Sample variance s_2^2	15.36
Sample size n_1	11
Sample size n_2	8
Variance ratio (VR)	1.4238
$P(F \geq VR)$	0.3284
Power of this actual test	15.359% (for $\alpha = 0.05$)
Minimal sample size	86 (for $\alpha = 0.05, \beta = 0.1$)
Hypothetical sample size n_1	60
Hypothetical sample size n_2	60
Variance ratio (VR)	1.4238
$P(F \geq VR)$	0.0889
Power of this hypothetical test	76.763% (for $\alpha = 0.05$)
Hypothetical sample size n_1	20
Hypothetical sample size n_2	30
Variance ratio (VR)	1.4238
$P(F \geq VR)$	0.1908
Power of this hypothetical test	38.348% (for $\alpha = 0.05$)

7. Power calculations for 1 and 2 correlations

The correlation coefficient r calculated from a normally distributed sample of size n has a standard error

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

and is an estimator of the population correlation ρ . A test for zero correlation, i.e., $H_0 : \rho = 0$, can be based on the statistics

$$t = \frac{r}{s_r},$$

$$\text{or } F = \frac{1+|r|}{1-|r|},$$

where t has a t distribution with $n-2$ degrees of freedom, and F has an F distribution with $n-2$ and $n-2$ degrees of freedom. The Fisher z transform and standard error s_z , defined as

$$z = \tanh^{-1} r,$$

$$= \frac{1}{2} \log \left(\frac{1+r}{1-r} \right),$$

$$s_z = \sqrt{\frac{1}{n-3}},$$

are also used to test $H_0 : \rho = \rho_0$, by calculating the unit normal deviate

$$Z = \frac{z - \zeta_0}{s_z}$$

where $\zeta_0 = \tanh^{-1} \rho_0$. The power is calculated using the critical value

$$r_c = \sqrt{\frac{t_{\alpha/2, n-2}^2}{t_{\alpha/2, n-2}^2 + n - 2}}$$

which leads to the transform $z_c = \tanh^{-1} r_c$ and

$$Z_\beta = (z - z_c)\sqrt{n-3}$$

then the sample size required to reject $H_0 : \rho = 0$, when actually ρ is nonzero, can be calculated using

$$n = \left(\frac{Z_\beta + Z_{\alpha/2}}{\zeta_0} \right)^2 + 3.$$

For two samples, X of size n_x and Y of size n_y , where it is desired to test $H_0 : \rho_x = \rho_y$, the appropriate Z statistic is

$$Z = \frac{z_x - z_y}{s_{xy}}$$

$$\text{where } s_{xy} = \sqrt{\frac{1}{n_x - 3} + \frac{1}{n_y - 3}}$$

and the power and sample size are calculated from

$$Z_\beta = \frac{|z_x - z_y|}{s_{xy}} - Z_{\alpha/2},$$

$$\text{and } n = 2 \left(\frac{Z_{\alpha/2} + Z_\beta}{z_x - z_y} \right)^2 + 3.$$

Example 7a

A sample correlation coefficient $R_1 = 0.87$ was calculated from a sample of size 12 and a population correlation $R_0 = 0.5$ was assumed, resulting in the following calculations.

Example 7a. $H_0 : \rho = 0, H_1 : \rho > 0$ (or $> R_0$)	
Current α	0.05
Current β	0.01
Correlation coefficient R_1	0.87
Correlation coefficient R_0	0.5
Sample size	12
95% confidence limits	0.5893, 0.9633
Two tailed t -test, p	0.0002
Power of this test	97.882%
N for $H_1 : \rho > 0$	14 (R_1 given)
N for $H_1 : \rho > R_0$	64 (R_1 arbitrary)

These results indicate that a 95% confidence interval for R_1 is (0.5893, 0.9633) and that the power for this test was 97.882%. The sample size would have to be increased to 14 to obtain 99% power with the current sample correlation coefficient R_1 , while a sample size of 64 would be required to ensure 99% power in a test for $|\rho| \geq 0.5$ before a sample is taken in order to calculate R_1 .

Example 7b

Two samples were analyzed to investigate equality of correlation coefficients and estimate the power and sample size needed for specified power for hypothetical samples with the same correlation coefficients but with equal hypothetical sample sizes.

Example 7b. $H_0 : \rho_1 = \rho_2, H_1 : \rho_1 - \rho_2 > 0$	
Current α	0.05
Current β	0.2
R_1, N_1, Z_1	0.78, 98, 1.045
R_2, N_2, Z_2	0.84, 95, 1.221
Two-tailed Z -test, p	0.2294
Power of this test	22.42%
For a Z -diff $ z_1 - z_2 $	0.1758
sample size N required	511

These results show that, for samples with R_1 estimated from a sample of size 98, and R_2 estimated from a sample of size 95 the power was 22.42%. For a hypothetical sample with the same correlation coefficients but estimated from samples with size N then, for 80% power, N would have to be at least 511.

8. Power calculations for a chi-square test

The calculations are based on the chi-square test for either a contingency table, or sets of observed and expected frequencies. However, irrespective of whether the test is to be performed on a contingency table or on samples of observed and expected frequencies, the null hypotheses can be stated in terms of k probabilities as

H_0 : the probabilities are $p_0(i)$, for $i = 1, 2, \dots, k$,

H_1 : the probabilities are $p_1(i)$, for $i = 1, 2, \dots, k$.

The power can then be estimated using the non-central chi-square distribution with non-centrality parameter λ and ν degrees of freedom given by

$$\lambda = nQ,$$

$$\text{where } Q = \sum_{i=1}^k \frac{(p_0(i) - p_1(i))^2}{p_0(i)},$$

$$n = \text{total sample size,}$$

$$\text{and } \nu = k - 1 - \text{no. of parameters estimated.}$$

You can either input the Q values directly, or read in vectors of observed and expected frequencies. If you do input frequencies $f_i \geq 0$ they will be transformed internally into probabilities, i.e., the frequencies only have to be positive integers as they are normalized to sum unity using

$$p_i = f_i / \sum_{i=1}^k f_i.$$

In the case of contingency table data with r rows and c columns, the probabilities are calculated from the marginals $p_{ij} = p(i)p(j)$ in the usual way, so you must input $k = rc$, and the number of parameters estimated as $r + c - 2$, so that $\nu = (r - 1)(c - 1)$.

Example 8

We demonstrate this procedure using the example of a weighted die discussed by William C Guenther in The American Statistician 31 (1977) pp 83–85 using table look-up. Here the sum of squares was $Q = 0.05$ with a sample size of $n = 120$ so that the degrees of freedom were $\lambda = 6$ and β was calculated to be approximately 0.5671. Further calculations showed that a sample size of $n > 330$ was required to achieve $\beta = 0.1$, while the results calculated by SIMFIT were as follows.

Example 8. Power analysis for a chi-square test $H_0 : p_0(i) = p_1(i)$					
$\alpha = 0.05$	$\beta = 0.5761$	$Q = 0.05$	$\lambda = 6$	$\chi^2 = 12.59$	$N = 128$
$\alpha = 0.05$	$\beta = 0.1000$	$Q = 0.05$	$\lambda = 6$	$\chi^2 = 12.59$	$N = 352$

Theory

Experiments often generate random samples from a population so that parameters estimated from the samples can be used to test hypotheses about the population parameters. So it is natural to investigate the relationship between sample size and the absolute precision of the estimates, given the expectation $E(X)$ and variance $\sigma^2(X)$ of the random variable. For a single observation, i.e., $n = 1$, the Chebyshev inequality

$$P(|X - E(X)| < \epsilon) \geq 1 - \frac{\sigma^2(X)}{\epsilon^2}$$

with $\epsilon > 0$, indicates that, for an unspecified distribution,

$$P(|X - E(X)| < 4.5\sigma(X)) \geq 0.95,$$

and $P(|X - E(X)| < 10\sigma(X)) \geq 0.99$,

but, for an assumed normal distribution,

$$P(|X - E(X)| < 1.96\sigma(X)) \geq 0.95,$$

and $P(|X - E(X)| < 2.58\sigma(X)) \geq 0.99$.

However, provided that $E(X) \neq 0$, it is more useful to formulate the Chebyshev inequality in terms of the relative precision, that is, for $\delta > 0$

$$P\left(\left|\frac{X - E(X)}{E(X)}\right| < \delta\right) \geq 1 - \frac{1}{\delta^2} \frac{\sigma^2(X)}{E^2(X)}.$$

Now, for an unspecified distribution,

$$P\left(\left|\frac{X - E(X)}{E(X)}\right| < 4.5 \frac{\sigma(X)}{|E(X)|}\right) \geq 0.95,$$

and $P\left(\left|\frac{X - E(X)}{E(X)}\right| < 10 \frac{\sigma(X)}{|E(X)|}\right) \geq 0.99$,

but, for an assumed normal distribution,

$$P\left(\left|\frac{X - E(X)}{E(X)}\right| < 1.96 \frac{\sigma(X)}{|E(X)|}\right) \geq 0.95,$$

and $P\left(\left|\frac{X - E(X)}{E(X)}\right| < 2.58 \frac{\sigma(X)}{|E(X)|}\right) \geq 0.99$.

So, for high precision, the coefficient of variation $cv\%$

$$cv\% = 100 \frac{\sigma(X)}{|E(X)|}$$

must be as small as possible, while the signal-to-noise ratio $SN(X)$

$$SN(X) = \frac{|E(X)|}{\sigma(X)}$$

must be as large as possible. For instance, for the single measurement to be within 10% of the mean 95% of the time requires $SN \geq 45$ for an arbitrary distribution, or $SN \geq 20$ for a normal distribution. A particularly

valuable application of these results concerns the way that the signal-to-noise ratio of sample means depends on the sample size n . From

$$\begin{aligned}\bar{X} &= \frac{1}{n} \sum_{i=1}^n x_i, \\ \text{Var}(\bar{X}) &= \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X) \\ &= \frac{1}{n} \sigma^2(X),\end{aligned}$$

it follows that, for arbitrary distributions, the signal-to-noise ratio of the sample mean $SN(\bar{X})$ is given by $SN(\bar{X}) = \sqrt{n}SN(X)$, that is

$$SN(\bar{X}) = \sqrt{n} \frac{E(X)}{\sigma(X)}.$$

This result, known as the law of \sqrt{n} , implies that the signal-to-noise ratio of the sample mean as an estimate of the population mean increases as \sqrt{n} , so that the relative error in estimating the mean decreases like $1/\sqrt{n}$.

If $f(x)$ is the density function for a random variable X , then the null and alternative hypotheses can sometimes be expressed as

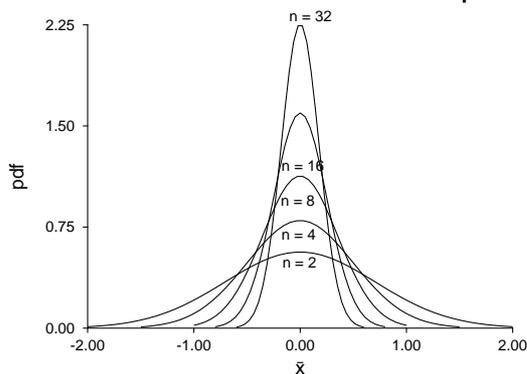
$$\begin{aligned}H_0 &: f(x) = f_0(x) \\ H_1 &: f(x) = f_1(x)\end{aligned}$$

while the error sizes, given a critical region C , are

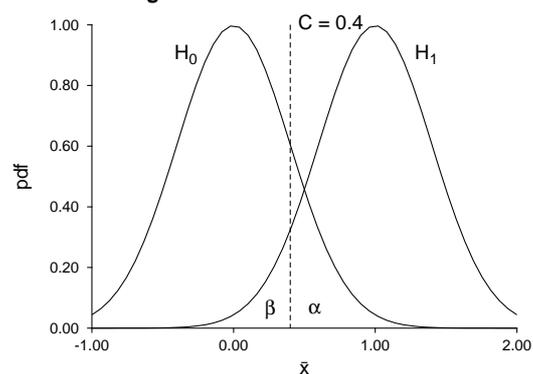
$$\begin{aligned}\alpha &= P_{H_0}(\text{reject } H_0) \text{ (i.e., the Type I error)} \\ &= \int_C f_0(x) dx \\ \beta &= P_{H_1}(\text{accept } H_0) \text{ (i.e., the Type II error)} \\ &= 1 - \int_C f_1(x) dx.\end{aligned}$$

Usually α is referred to as the significance level, β is the operating characteristic, while $1 - \beta$ is the power, frequently expressed as a percentage, i.e., $100(1 - \beta)\%$, and these will both alter as the critical region is changed.

Distribution of the mean as a function of sample size



Significance Level and Power



This figure illustrates the concepts of signal-to-noise ratio, significance level, and power. The family of curves on the left are the probability density functions for the distribution of the sample mean \bar{x} from a normal

distribution with mean $\mu = 0$ and variance $\sigma^2 = 1$. The curves on the right illustrate the significance level α , and operating characteristic β for the null and alternative hypotheses

$$H_0 : \mu = 0, \sigma^2 = 4$$

$$H_1 : \mu = 1, \sigma^2 = 4$$

for a test using the sample mean from a sample of size $n = 25$ from a normal distribution, with a critical point $C = 0.4$. The significance level is the area under the curve for H_0 to the right of the critical point, while the operating characteristic is the area under the curve for H_1 to the left of the critical point. Clearly, increasing the critical value C will decrease α and increase β , while increasing the sample size n will decrease both α and β .

Often it is wished to predict power as a function of sample size, which can sometimes be done if distributions $f_0(x)$ and $f_1(x)$ are assumed, necessary parameters are provided, the critical level is specified, and the test procedure is defined. Essentially, given an implicit expression in k unknowns, this option solves for one given the other $k - 1$, using iterative techniques. For instance, you might set α and β , then calculate the sample size n required, or you could input α and n and estimate the power. Note that 1-tail tests can sometimes be selected instead of 2-tail tests (e.g., by replacing $Z_{\alpha/2}$ by Z_α in the appropriate formula) and also be very careful to make the correct choice for supplying proportions, half-widths, absolute differences, theoretical parameters or sample estimates, etc.

A word of warning is required on the subject of calculating n required for a given power. The values of n will usually prove to be very large, probably much larger than can be used. So, for pilot studies and typical probing investigations, the sample sizes should be chosen according to cost, time, availability of materials, past experience, and so on. Sample size calculations are only called for when Type II errors may have serious consequences, as in clinical trials, so that large samples are justified.

Of course, the temptation to choose 1-tail instead of 2-tail tests, or use variance estimates that are too small, in order to decrease the n values should be avoided, but it happens.