

## Sample Size Determination (One Sample)

This procedure determines a suitable sample size for estimating or testing hypotheses concerning any of the following:

1. the **mean**  $\mu$  of a normal distribution.
2. the **standard deviation**  $\sigma$  of a normal distribution.
3. the **proportion** of successes  $\theta$  in a binomial distribution.
4. the **rate** of events  $\lambda$  in a Poisson distribution.

It finds a sample size that achieves either of two goals:

1. generates a *confidence interval* for the parameter of specified width.
2. yields the desired *power* in a test of hypotheses concerning the value of the parameter.

**Sample StatFolio:** *samsize1.sgp*

### Sample Data:

None.

### Data Input

The first dialog displayed by this procedure is used to specify the problem of interest to the analyst.

Sample-Size Determination

Parameter

Normal Mean

Normal Sigma

Binomial Proportion

Poisson Rate

OK

Cancel

Help

Hypothesized Mean:  
10.

Hypothesized Sigma:  
3.

Hypothesized Proportion:  
0.5

Hypothesized Rate:  
1.

- **Parameter:** the parameter to be estimated. It is assumed that a random sample of size  $n$  will be taken from the specified distribution and used to estimate or test the value of the indicated parameter. The procedure will determine a suitable value for  $n$ .
- **Hypothesized Mean:** the anticipated value of the mean  $\mu$ . If performing a hypothesis test, this value forms the null hypothesis. If constructing a confidence interval, this value is only used if the desired width of the interval is specified in relative (percentage) terms.
- **Hypothesized Sigma:** the anticipated value of the standard deviation  $\sigma$ . If performing a hypothesis test for  $\sigma$ , this value forms the null hypothesis. If constructing a confidence interval for  $\sigma$ , this value is only used if the desired width of the interval is specified in absolute (not percentage) terms. If performing a hypothesis test for  $\mu$ , then this value is a critical part of the calculation and should either be known exactly or be a reliable estimate from previous data.
- **Hypothesized Proportion:** the anticipated value of the binomial proportion  $\theta$ . If performing a hypothesis test, this value forms the null hypothesis. If constructing a confidence interval, it is used as a preliminary estimate in the calculations.
- **Hypothesized Rate:** the anticipated value of the Poisson rate  $\lambda$ . If performing a hypothesis test, this value forms the null hypothesis. If constructing a confidence interval, it is used as a preliminary estimate in the calculations.

For example, the above dialog box indicates a desire to estimate the mean of a normal distribution, thought to be around  $\mu = 10$  with a standard deviation of  $\sigma = 3$ .

The second dialog box elicits information about the goal of the analysis:

- **Control:** specifies the goal from among the following choices:
  1. **Absolute error:** a confidence interval for the parameter is to be constructed. That interval should not deviate from the point estimate of the parameter in either direction by more than the absolute distance  $W$  indicated.
  2. **Relative error:** a confidence interval for the parameter is to be constructed. That interval should not deviate from the point estimate of the parameter in either direction by more than the relative percentage  $P$  indicated. This is identical to the absolute error case with  $W$  set equal to  $P$  times the specified parameter.
  3. **Power:** a hypothesis test is to be performed. The power of the test  $100(1-\beta) \%$  should equal the percentage specified when the true value of the parameter deviates from the null hypothesis by the indicated  $\Delta = \text{Difference to Detect}$ . Power is defined as the probability of rejecting the null hypothesis when it is false. If a two-sided test is to be performed, then that power must be achieved both above and below the parameter value specified by the null hypothesis.
  4. **Sample Size:** the predetermined sample size  $n$ . This option is used to plot the power curve for a sample size that was not calculated by this procedure.
- **Confidence Level:** the level of confidence  $100(1-\alpha)\%$  used when constructing confidence intervals. The value  $\alpha$  is also used as the level of Type I error when testing hypotheses. A Type I error occurs when the null hypothesis is falsely rejected.
- **Alternative Hypothesis:** select *Not Equal* for a two-sided hypothesis test, *Less Than* if the alternative hypothesis is that the parameter is less than the value specified by the null hypothesis, or *Greater Than* if the alternative hypothesis is that the parameter is greater than the value specified by the null hypothesis.
- **Sigma:** when estimating or testing a normal mean, whether the standard deviation is assumed to be known (z test) or if it will be estimated from the data (t test).

For example, the dialog box above indicates the following two-sided test is to be performed:

Null hypothesis  $H_0: \mu = 10$   
 Alt. hypothesis  $H_A: \mu \neq 10$

The probability of a Type I error (rejecting a true null hypothesis) is set to  $\alpha = 5\%$ , while the probability of a Type II error (not rejecting a false null hypothesis) is set to  $\beta = 10\%$  when the true mean falls to 7 or rises to 13.

The following table may be helpful in remembering how to set the error probabilities.

	Do Not Reject $H_0$	Reject $H_0$
$H_0$ is true	Correct decision	Type I error risk = $\alpha$
$H_0$ is false	Type II error risk = $\beta$	Correct decision power = $1-\beta$

## Analysis Summary

The *Analysis Summary* displays the desired goal and the sample size that will achieve it:

### Sample-Size Determination

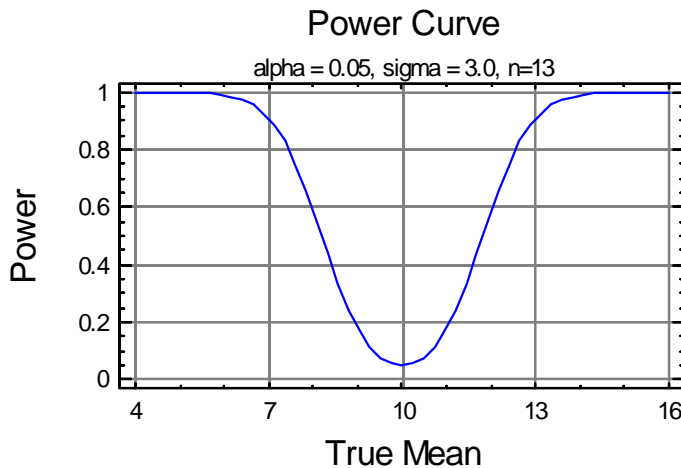
Parameter to be estimated: normal mean  
 Desired power: 90.0% for mean = 10.0 versus mean = 13.0  
 Type of alternative: not equal  
 Alpha risk: 5.0%  
 Sigma: 3.0 (to be estimated)

The required sample size is  $n=13$  observations.

In the current example, a sample of  $n = 13$  observations is required to achieve the power requested.

## Power Curve

The *Power Curve* shows the power of the specified test of hypotheses when the sample size is set to  $n$ .



It can be seen that the power (probability of rejecting the null) is only around  $\alpha$  when the true mean is close to that specified by the null hypothesis, by it rises to  $1-\beta$  when the mean varies in either direction by the specified *Difference to Detect*.

Calculations

- **Normal Mean – Confidence Interval**

If  $\sigma$  is assumed to be known, find the smallest  $n$  such that

$$z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq W \quad (1)$$

If  $\sigma$  will be estimated from the data, find the smallest  $n$  such that

$$t_{\alpha/2, n-1} \frac{\sigma}{\sqrt{n}} \leq W \quad (2)$$

- **Normal Mean – Hypothesis Test**

If  $\sigma$  is assumed to be known, find the smallest  $n$  such that

$$\left[ \frac{(z_{\alpha/2} + z_{\beta})\sigma}{\Delta} \right]^2 \leq n \quad (3)$$

If  $\sigma$  is to be estimated from the data, find the smallest  $n$  such that

$$\left[ \frac{(t_{\alpha/2, n-1} + t_{\beta, n-1})\sigma}{\Delta} \right]^2 \leq n \quad (4)$$

- **Normal Sigma – Confidence Interval**

Find the smallest  $n$  such that

$$\sigma \left( 1 - \sqrt{\frac{(n-1)}{\chi_{\alpha/2, n-1}^2}} \right) \leq W \quad (5)$$

and

$$\sigma \left( \sqrt{\frac{(n-1)}{\chi_{1-\alpha/2, n-1}^2}} - 1 \right) \leq W \quad (6)$$

• **Normal Sigma – Hypothesis Test**

Find the smallest  $n$  such that

$$1.5 + 0.5 \left[ \frac{(\lambda z_{\alpha/2} + z_{\beta})}{(\lambda - 1)} \right]^2 \leq n \tag{7}$$

where

$$\lambda = \frac{\sigma}{\sigma + \Delta} \quad \text{if } \Delta > 0 \tag{8}$$

or

$$\lambda = \frac{\sigma + \Delta}{\sigma} \quad \text{if } \Delta < 0 \tag{9}$$

• **Binomial Proportion – Confidence Intervals**

Find the smallest  $n$  such that

$$\theta - \left( \frac{v_1 F_{1-\alpha/2, v_1, v_2}}{v_2 + v_1 F_{1-\alpha/2, v_1, v_2}} \right) \leq W \tag{10}$$

and

$$\left( \frac{v_3 F_{1-\alpha/2, v_3, v_4}}{v_4 + v_3 F_{1-\alpha/2, v_3, v_4}} \right) - \theta \leq W \tag{11}$$

where

$$v_1 = 2n\theta \tag{12}$$

$$v_2 = 2(n(1-\theta) + 1) \tag{13}$$

$$v_3 = 2(n\theta + 1) \tag{14}$$

$$v_4 = 2n(1-\theta) \tag{15}$$

• **Binomial Proportion – Hypothesis Tests**

Find the smallest  $n$  such that the power of the test is no less than that specified, using exact probabilities for the binomial distribution.

- **Poisson Rate – Confidence Intervals**

Find the smallest  $n$  such that

$$\lambda - \frac{\chi_{1-\alpha/2, 2n\lambda}}{2n} \leq W \quad (16)$$

and

$$\frac{\chi_{\alpha/2, 2(n\lambda+1)}}{2n} - \lambda \leq W \quad (17)$$

- **Poisson Rate – Hypothesis Tests**

Find the smallest  $n$  such that the power of the test is no less than that specified, using exact probabilities for the Poisson distribution.

**Note:** for all one-sided tests, replace  $\alpha/2$  by  $\alpha$  in the equations for the hypothesis tests.