

## Sample Size Determination for the Test of One Proportion

By Keith M. Bower, M.S.

### Abstract

It is frequently necessary for manufacturers to assess whether a batch of product meets a certain level of quality (e.g. less than  $x\%$  of items are defective). It is possible to use MINITAB™ to assist in determining an appropriate sample size to assess the true quality level. Such a sampling scheme may be required, for example, when the cost of inspection is very high, or where destructive testing is necessary. In such instances, 100% inspection may not be feasible or not desired.

The statistical test used in this paper is the one-sample proportion test, whereby the total number of defective items from a random sample will be used to estimate the true defect rate for the population.

### Power and Sample Size

When performing a statistical hypothesis test, consider two types of mistakes the analyst could ultimately make, namely:

Type I: Reject the null hypothesis, though the null hypothesis is true, and

Type II: Fail to reject the null hypothesis, though the null hypothesis is false.

We shall define the type I error rate as  $\alpha$ , and the type II error rate as  $\beta$ . In many instances, practitioners will be satisfied with setting  $\alpha$  at 0.05, and are typically more liberal with  $\beta$  as being around 0.10 to 0.20. It is important to note that these are rough guidelines and that the practitioner should assess the cost associated with making each type of error. For example, regarding the frequently used  $\alpha$  rate of 0.05, note the comment by Fisher<sup>1</sup> (1926) in that:

*“If one in twenty does not seem high enough odds, we may, if we prefer it, draw the line at one in fifty (the two per cent. point), or one in a hundred (the 1 per cent. point). Personally, the writer prefers to set a low standard of significance at the 5 per cent. point...”*

The power of a statistical test is defined as the value of 1 minus  $\beta$  (i.e. the probability of *correctly* rejecting the null hypothesis). Practitioners, therefore, frequently look for the power of their test as being greater than around 0.80. To emphasize again, such guidelines are, of course, highly dependent upon the experiment itself.

It should be noted at this point that if the power of the statistical test is very high, then the test might be considered over-sensitive from an economic perspective. Imagine a sample size that is very large indeed. We may be able to detect a very small difference from the hypothesized value, though that difference may not be of any *practical* significance, especially in relation to the cost of sampling such a large amount of product. As is noted by Lenth<sup>2</sup> (2000):

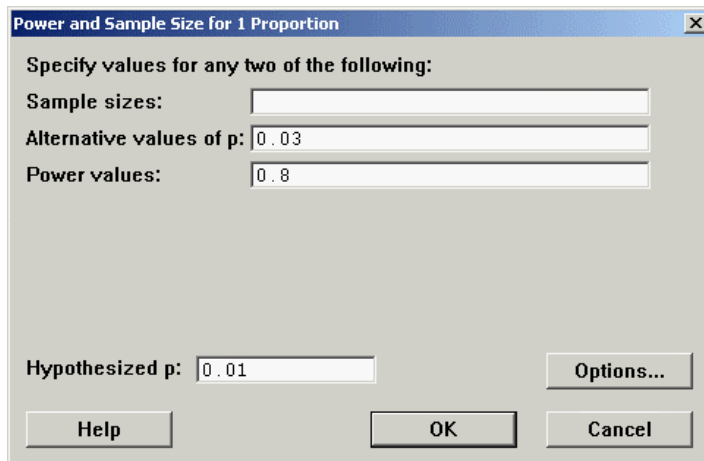
*“I recommend consulting with subject-matter experts before the data are collected to determine absolute effect-size goals that are consistent with the scientific goals of the study.”*

### Example

Consider a manufacturing process that is operating around a 1% defect rate for a particular type of machine part. If the rate of defectives increases to above 3%, this will have serious cost implications to the organization. They are required to determine an appropriate sample size such that the type I error rate will be 0.05, and for the test to have a power of 0.80 to detect an increase in the rate of defects from 1% to 3% or above.

Formally, we shall test the null hypothesis,  $H_0: P = 0.01$ , vs. the alternative hypothesis,  $H_1: P > 0.01$ , where  $P$  is the true proportion nonconforming.

Figure 1



MINITAB can be used to assist in this sample size determination by selecting *Stat>Power and Sample Size>1 Proportion*. One would complete the dialog box as shown in Figure 1, then click on the *Options* button and choose *Greater than* to make the test sensitive to the change in the positive direction.

As the output in Figure 2 shows, a sample size of 236 would meet the requirements for this test. Note that 3.00E-02 denotes that three is multiplied by ten raised to the power of negative two (i.e. 0.03).

Figure 2

### Power and Sample Size

Test for One Proportion

Testing proportion = 0.01 (versus > 0.01)

Alpha = 0.05

Alternative Proportion	Sample Size	Target Power	Actual Power
3.00E-02	236	0.8000	0.8000

To obtain a wider set of results through the *Power and Sample Size* functionality, it is possible to include more than one number in some of the boxes. For example, in the *Power values* box, entering “0.8:0.9/0.02” will result in power computations from 0.80 to 0.90 in increasing units of 0.02, as shown in the third column in Figure 3. Obviously, for the test to become more sensitive, a larger sample size is required, ceteris paribus. This is reflected in the output for Figure 3.

Figure 3

### Power and Sample Size

Test for One Proportion

Testing proportion = 0.01 (versus > 0.01)

Alpha = 0.05

Alternative Proportion	Sample Size	Target Power	Actual Power
3.00E-02	236	0.8000	0.8000
3.00E-02	256	0.8200	0.8203
3.00E-02	278	0.8400	0.8402
3.00E-02	303	0.8600	0.8602
3.00E-02	332	0.8800	0.8804
3.00E-02	366	0.9000	0.9004

### Summary

It is possible to use the *Power and Sample Size* functionality in MINITAB to determine sample sizes to perform statistical tests. It is wise to ensure that adequate resources are devoted to obtain an appropriately large sample for a test. The size and direction of the difference from a hypothesized value, the hypothesized value itself, and the  $\alpha$  and  $\beta$  rates are key factors when choosing sample sizes using the one-sample proportion test. For more information on this statistical procedure, see Hogg and Tanis<sup>3</sup> (1997) and Box, Hunter and Hunter<sup>4</sup> (1978).

### Reference

1. Fisher, R.A. (1926). “The Arrangement of Field Experiments” *Journal of the Ministry of Agriculture, Vol. 33*.
2. Lenth, R.V. (2000). “Two Sample-Size Practices That I Don’t Recommend,” *American Statistical Association: Proceedings of the Section on Physical and Engineering Sciences*.

3. Hogg, R.V. and Tanis, E.A. (1997). *Probability and Statistical Inference*, 5<sup>th</sup> Edition, Prentice-Hall.
4. Box, G.E.P., Hunter, W.G., and Hunter, J.S. (1978). *Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building*, Wiley.

***Keith M. Bower has an M.S. in Quality Management and Productivity from the University of Iowa, and is a Technical Training Specialist with Minitab Inc.***