

# Testing a Hypothesis – Part 3

Compiled by Henry Posters, PE, CQE, CRE, CQM

References: Part One and Part Two @ <http://www.asqsandiego.org/library.htm>

## Determining the sample size required for testing a hypothesis

(Normal Distributions)

The sample size required to test a hypothesis will depend on:

- A) The sample risk desired ( $\alpha$  and  $\beta$ ). Increasing the sample size will decrease the chance of making a type I error (rejecting the null hypothesis when it is true).
- B) The size of the smallest true difference that is to be detected. Detecting small differences will usually require large sample sizes.
- C) The variation in the characteristic being measured. Less variation is better. Data with little variation require smaller test samples to detect a departure from  $H_0$ .

The following questions should be asked when determining the sample size for estimation or hypotheses testing:

1. Type of data: Is the chosen response variable discrete or continuous? Can a normal distribution be assumed? Determining the required sample size depends on the type of data that is being generated.
2. Amount of expected variance: How much variability is expected between individual response variable measurements? If this quantity is unknown, what is the **range** of measurements?
3. What type of inference will be made: point estimation or hypothesis testing?
4. How great an effect is to be shown? That is, how accurate an estimate, how large a difference between levels of a test factor, etc.?
5. What is the required confidence coefficient for estimation testing? What are the  $\alpha$  and  $\beta$  for the hypothesis-testing inference?
6. From a practical/economical viewpoint, how large a sample can be taken? What is the cost of taking data relative to expected value gained?
7. Data integrity: How accurate is the data? Could there be bias?

### Example 1 (Continuous Data)

Suppose it were important to detect the fact that the average life of the battery model (A) cited in the previous example (part two) decreased to 0.78 Ah. We want to be 90% sure of detecting this change ( $\beta = 0.1$ ). If the true average capacity is 0.8 Ah, there should be only be a 5% risk of rejecting the hypothesis ( $\alpha = 0.05$ ) when it is true.

#### Analysis:

To detect a 0.2 Ah decrease in average battery life with a test power of 90 percent and a five percent risk of drawing the wrong conclusion (no decrease if there actually is one) we need a sample size of:

$$n = \frac{(Z_{1-\alpha} + Z_{1-\beta})^2}{\Delta^2} \quad \text{where} \quad \Delta = \frac{|\mu_1 - \mu|}{\sigma} = \frac{|0.78 - 0.80|}{0.067} = 0.3$$
$$n = \frac{(Z_{1-\alpha} + Z_{1-\beta})^2}{\Delta^2} = \frac{(Z_{0.95} + Z_{0.9})^2}{\Delta^2} = \frac{(1.645 + 1.28)^2}{(0.3)^2} = \frac{8.556}{0.09} = 95 \text{ units}$$

When the variance is estimated from the observed test values, some corrections in the calculated sample size must be made. If  $\alpha = 0.05$  **two units** are added to the calculated sample size. So the adjusted sample size is 97 units.

### Example 2 (Discrete Data)

Historically, a manufacturing process has been known to produce 10 percent defective items. It is estimated that a process change will cut the defect rate in half and reduce it to five percent. If we want to achieve a 95 confidence level and 80 percent test power, how many items need to be examined to detect this level of improvement?

#### Analysis:

$$p_0 = 0.1 \text{ and } p' = 0.05 \quad \beta = 0.2$$

To ensure that the power of the test ( $1 - \beta$ ) is 80 percent with  $\alpha = 0.05$  we need a sample size of:

$$n = \left[ \frac{z_{\alpha/2} \sqrt{p_0(1-p_0)} + z_{\beta} \sqrt{p'(1-p')}}{p' - p_0} \right]^2 = \left[ \frac{1.96 \sqrt{(0.1)(0.9)} + 0.84 \sqrt{(0.05)(0.95)}}{0.05} \right]^2 =$$
$$\left[ \frac{0.588 + 0.1831}{0.05} \right]^2 = \left[ \frac{0.7711}{0.05} \right]^2 = (15.42)^2 \approx 238 \text{ items}$$

We need to inspect 238 items to detect, with 95 percent confidence, the result of the process improvement.

### Determining the sample size in product life testing

In product life testing the Normal distribution is not of major importance. The most common distributions in life testing are the Exponential, Weibull, and Log-Normal distributions. As the probability of failure, at normal use conditions, is very low for products such as aerospace products, medical devices and modern electronics, it is often necessary to apply accelerated tests at high-stress conditions. This to obtain some failed units so that an estimate of the average failure rate can be made. To determine the minimum sample size required to demonstrate that a product does not exceed a maximum average failure rate, a statistical distribution must be assumed for product life testing. Conditions for successful application of accelerated life test models are:

- The form of the lifetime distribution is known (Exponential, Weibull, or Log-Normal).
- The average failure rate requirement has been stated for a well defined time period.
- With the exception of the Exponential distribution, the shape factor ( $\beta$ ) of the life distribution must be known, estimated, or assumed under sound engineering judgment.
- The acceleration factor ( $A_F$ ) must be known or estimated.
- The confidence coefficient or confidence level must be given.
- For sample size determination the feasible test time and a failure demonstration number must be defined.

### Determining the sample size - Exponential Distributions

The exponential distribution is one of the most common life distribution used in reliability testing. It is the only distribution with a constant failure rate function. The exponential distribution represents a good distribution model for testing units/products that have no early failures (units have passed the burn-in period) and no significant wear-out mechanism. For the test model to be valid, the test time should be confined to the constant-failure-rate region of the 'bathtub' curve. The sample size is calculated as follows:

$$n = \frac{\chi^2_{(2r+2),\gamma}}{2F_{(t_n)}} \quad \text{in which} \quad F_{(t_n)} = 1 - e^{-\lambda_R t_n} \quad \text{and} \quad t_n = t_s \times A_F$$

Notation:  $\lambda_R$  denotes the specified average failure rate

<b>T</b>	denotes the time period related to the specified average failure rate
<b>t<sub>n</sub></b>	denotes the test time under normal operating conditions
<b>A<sub>F</sub></b>	denotes the acceleration factor between test and normal use conditions
<b>t<sub>s</sub></b>	denotes the accelerated test duration
<b>γ</b>	denotes the confidence coefficient
<b>f<sub>r</sub></b>	denotes the failure demonstration number

$\chi_{(2r+2),\gamma}$  denotes the  $\gamma$  cutoff point of the chi-squared distribution with  $2f_r+2$  degrees of freedom. Degrees of freedom indicate the amount of independence in a data set.  $2(f_r+1)$  df means that only  $2(f_r+1)$  of the quantities on which  $\chi$  is based are freely determined (are free to vary).

### Example

The reliability of a complex printed circuit board, used in an aeronautical instrument, must have a reliability of at least 85% at five year of continuous operation. The circuit board has 250 non-redundant components, each making a small contribution to the total failure rate. Hence, the exponential distribution is considered a valid model. A 2000 hour accelerated test will be conducted at elevated temperature, for which an acceleration factor of six has been calculated. Choosing a demonstration number of two failures (failed boards will be replaced), what should be the sample size (# of boards) to demonstrate, with 80% confidence, that the five year reliability requirement has been satisfied?

Analysis:  $t_n = 6(2000) = 12000$  hours                      Reliability (<sub>5 year</sub>) = 0.85 =  $e^{-\lambda T}$

$T = 5(24)(365) = 43800$  hours                       $\ln 0.85 = -43800 \lambda_R$                        $\lambda_R = 3.7 (10^{-6})$

$F_{tn} = 1 - R_{tn} = 1 - e^{-3.7(0.012)} = 1 - 0.9566 = 0.0434$                        $n = \frac{\chi_{6,0.8}^2}{2(0.0434)} = \frac{8.558}{0.0868} = 98$  units.

We need to test 98 printed circuit boards. To demonstrate, with 80 % confidence, that the circuit board design meets its reliability requirement, only two or fewer failures are allowed during the accelerated test.

This concludes a three part series on hypothesis testing. Some general references on how to calculate sample sizes in product life testing are listed below.

### References:

- Quality Control Handbook by J. M. Juran
- Experimental Statistics by Mary Gibbons Natrella
- Applied Reliability by Paul Tobias and David Trindade
- Accelerated Testing, Statistical Models, Test Plans, and Data Analyses by Wayne Nelson
- Determine Sample Size and Estimate Failure Rate in Life Testing by Eduardo Moura