Six Sigma:

Sample Size Determination and Simple Design of Experiments

Short Examples Series

using

Risk Simulator

Six Sigma – Sample Size Determination and Design of Experiments

File Names: *Six Sigma – Sample Size Correlation; Six Sigma – Sample Size DPU; Six Sigma – Sample Size Mean; Six Sigma – Sample Size Proportion; Six Sigma – Sample Size Sigma, Six Sigma – Delta Precision; and Six Sigma – Design of Experiments and Combinatorics*
Location: *Modeling Toolkit | Six Sigma*
Brief Description: *These five Six Sigma models illustrate how to obtain the required sample size in performing hypotheses testing from means to standard deviations and proportions*
Requirements: *Modeling Toolkit, Risk Simulator*
Modeling Toolkit Functions Used: *B2SixSigmaSampleSize, B2SixSigmaSampleSizeProportion, B2SixSigmaSampleSizeZeroCorrelTest, B2SixSigmaSampleSizeDPU, B2SixSigmaSampleSizeStdev, B2SixSigmaDeltaPrecision*

In performing quality controls and hypothesis testing, the size of the samples collected is of paramount importance. Theoretically, it would be impossible or too expensive and impractical to collect information and data on the entire population to be tested (e.g., all outputs from a large manufacturing facility). Therefore, statistical sampling is required. The question is: What size sample is sufficient? The size of a statistical sample is the number of repeated measurements that are collected. It is typically denoted *n*, a positive integer. Sample size determination is critical in Six Sigma and quality analysis as different sample sizes lead to different accuracies and precision of measurement. This can be seen in such statistical rules as the Law of Large Numbers and the Central Limit Theorem. With all else being equal, a larger sample size *n* leads to increased precision in estimates of various properties of the population. The question that always arises is: How many sample data points are required? The answer is it depends. It depends on the error tolerances and precision required in the analysis. This model is used to compute the minimum required sample size given the required error, precision and variance levels.

There are five different sample size determination models in the Modeling Toolkit. The first model (*Six Sigma – Sample Size Mean*) computes the minimum required sample size for a hypothesis test (one or two-tailed test), where the required errors and precisions are stated, as is the sample standard deviation (Figure 1). Alpha Error is the Type I error, also known as the significance level in a hypothesis test. It measures the probability of not having the true population mean included in the confidence interval of the sample. That is, it computes the probability of rejecting a true hypothesis. *1 – Alpha* is of course the confidence interval, or the probability that the true population mean resides in the sample confidence interval. Beta Error is the Type II error, or the probability of accepting a false hypothesis, or the probability of not being able to detect the mean's changes. *1 – Beta* is the power of the test. Delta Precision is the accuracy or precision with which the standard deviation may be estimated. For instance, a 0.10% Delta with 5% Alpha for 2 tails means that the estimated mean is plus or minus 0.10%, at a 90% (*1 – 2 x Alpha*) confidence level. Finally, the Sigma Sample is the sample standard deviation of the data set.

SAMPLE SIZE REQUIREMENTS (MEAN)

Input Assumptions

Alpha Error	10.00%
Beta Error	5.00%
Delta Precision	3.00%
Sigma Sample	0.15
One or Two-Tail Test	2

Required Sample Size   273

Figure 1: Sample size determination model for the mean

The remaining models are very similar in that they determine the appropriate sample size to test the standard deviation or sigma levels given some Alpha and Beta levels (*Six Sigma – Sample Size Sigma*), testing proportions (*Six Sigma – Sample Size Proportion*), or defect per unit (*Six Sigma – Sample Size DPU*). In addition, the *Six Sigma – Sample Size Correlation* model is used to determine the minimum required sample size to perform a hypothesis test on correlations. Finally, the *Six Sigma – Delta Precision* model works backwards in that given some Alpha and Beta errors and a given sample size, the Delta precision level is computed instead.

Design of Experiments (DOE) and Combinatorics

Another related issue to sampling and testing or experimentation is that of generating the relevant combinations and permutations of experimental samples to test. Let us say you have 5 different projects and need to decide all the combinations of projects that can occur. For instance, the value 00000 means no projects are chosen; or 10000, where project 1 is chosen and no others; or 010000 where project 2 is chosen and no others; and so forth. This result can also include multiple projects such as 10101 where projects 1, 3, and 5 are chosen. Clearly, if the portfolio has more projects, the combinations can become rather intractable. This is where combinatorial math comes in. We will use Excel to determine the combinatorials.

Procedure
1. Start Excel's Analysis Toolpak (**Tools** | **Add-Ins** | **Analysis Toolpak**).
2. Go to the *Combinatorics* model and view the examples.

The example shown illustrates how to convert a 5-project or 5-asset portfolio into all the possible combinations possible by using the *DEC2BIN* function in Excel. The *DEC2BIN* function stands for decimal to binary function. Specifically, it takes the value of a decimal or number and converts it into the relevant binary structure of 0s and 1s.

For instance, if there are 5 projects in a portfolio, then there are 32 combinations or 2^5 combinations. As binary coding starts from 0, we can number these outcomes from 0, 1, 2, all the way to 31 (providing 32 possible outcomes). We then convert these to a 5-digit binary sequence using the function *DEC2BIN(0,5)* which yields 00000. As another example, *DEC2BIN(10,5)* yields 01010, and so forth. By creating this combinatorial matrix, we see all the possible combinations for 5 projects. We can then pick out the ones required. For example, say we need to select 2 out of 5 projects, and need all the possible combinations. We know that there are 10 combinations, and by using the example model, we can determine exactly what they are. Figure 2 shows an example.

Using this approach, we can add in Risk Simulator to randomly select the projects and cycle through all possible combinations of projects that can be executed in time. As an example, say you have 5 projects in a portfolio and you would like to randomly choose 3 projects to execute in a sequence of three years. In each year you cannot execute more than 3 projects, and no project can repeat itself. In addition, you wish to run all possible combinations of possibilities. Such an example is also included in the model and is self-explanatory.

Design of Experiments and Combinatorics

Value	Combinatorial Matrix for 5 Projects Combinations	Decomposed 1	2	3	4	5	Total	Feasible 5 Choose 4	Feasible 5 Choose 3	Feasible 5 Choose 2
0	00000	0	0	0	0	0	0	0	0	0
1	00001	0	0	0	0	1	1	0	0	0
2	00010	0	0	0	1	0	1	0	0	0
3	00011	0	0	0	1	1	2	0	0	1
4	00100	0	0	1	0	0	1	0	0	0
5	00101	0	0	1	0	1	2	0	0	1
6	00110	0	0	1	1	0	2	0	0	1
7	00111	0	0	1	1	1	3	0	1	0
8	01000	0	1	0	0	0	1	0	0	0
9	01001	0	1	0	0	1	2	0	0	1
10	01010	0	1	0	1	0	2	0	0	1
11	01011	0	1	0	1	1	3	0	1	0
12	01100	0	1	1	0	0	2	0	0	1
13	01101	0	1	1	0	1	3	0	1	0
14	01110	0	1	1	1	0	3	0	1	0
15	01111	0	1	1	1	1	4	1	0	0
16	10000	1	0	0	0	0	1	0	0	0
17	10001	1	0	0	0	1	2	0	0	1
18	10010	1	0	0	1	0	2	0	0	1
19	10011	1	0	0	1	1	3	0	1	0
20	10100	1	0	1	0	0	2	0	0	1
21	10101	1	0	1	0	1	3	0	1	0
22	10110	1	0	1	1	0	3	0	1	0
23	10111	1	0	1	1	1	4	1	0	0
24	11000	1	1	0	0	0	2	0	0	1
25	11001	1	1	0	0	1	3	0	1	0
26	11010	1	1	0	1	0	3	0	1	0
27	11011	1	1	0	1	1	4	1	0	0
28	11100	1	1	1	0	0	3	0	1	0
29	11101	1	1	1	0	1	4	1	0	0
30	11110	1	1	1	1	0	4	1	0	0
31	11111	1	1	1	1	1	5	0	0	0
							SUM	5	10	10

Figure 2: Combinatorics results