# Categorical Data in a Designed Experiment Part 2: Sizing with a Binary Response

*Authored by: Francisco Ortiz, PhD*

*Version 2: 19 July 2018*

*Revised 18 October 2018*


**The goal of the STAT COE is to assist in developing rigorous, defensible test strategies to more effectively quantify and characterize system performance and provide information that reduces risk. This and other COE products are available at [www.AFIT.edu/STAT](www.AFIT.edu/STAT).**

STAT Center of Excellence
2950 Hobson Way – Wright-Patterson AFB, OH 45433

# Table of Contents



*Revision 1, 18 Oct 2018: Formatting and minor typographical/grammatical edits.*

# Introduction

In Part 1 of this best practice series, we learned that the data type used to represent a response can affect the size of the experiment and the quality of its analysis. Categorical data types, such as binary (pass/fail) measures, contain a relatively poor amount of information in comparison to continuous data types. This reduction in information increases the number of samples needed to detect significant changes of a response in the presence of noise. The use of categorical data types for responses should be avoided; however, there will be circumstances in which a pass/fail measure is the only practical way to characterize a system's performance.

Three methods to estimate the sample size needed for a designed experiment using a binary response will be presented in this paper, the arcsine transformation approach, the signal-to-noise method, and the inverse binomial sampling scheme method. The circumstances in which each method can be applied will be described in the paper. The methods will be demonstrated using a Target Location Error (TLE) example, originally presented in part 1 of this series. All methods presented are available with the Binary Response Calculator on the STAT COE website.

Keywords: binary responses, categorical factors, sample size, test and evaluation, design of experiments, confidence, power

# Background

## Target Location Error (TLE) Example Revisited

Let's revisit the missile targeting system example from part 1 of this best practice series. In this example, we are comparing the ability of a missile targeting system to accurately assess the coordinates of a target within a tolerance radius of 10 feet. An example of the error distribution (using notional data) is shown in Figure 1.

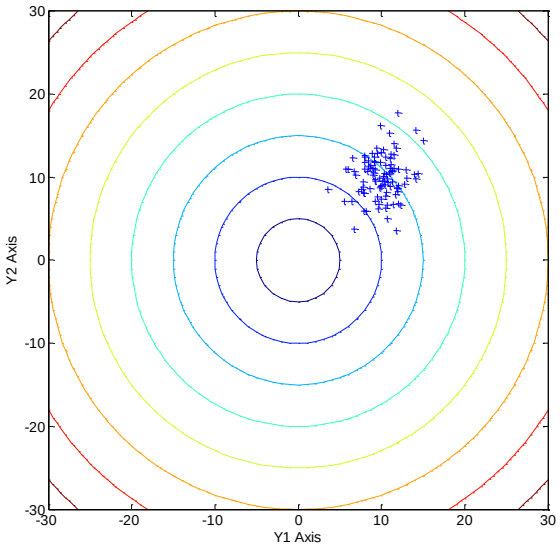



**Figure 1: Distribution of error distance example (notional data)**

Let's assume the only way to measure the systems is to categorize each attempt as either a "Pass" or "Fail" based on whether it falls within a 10 feet radius of the target (see Figure 2). This is a type of nominal response, specifically a binary response.

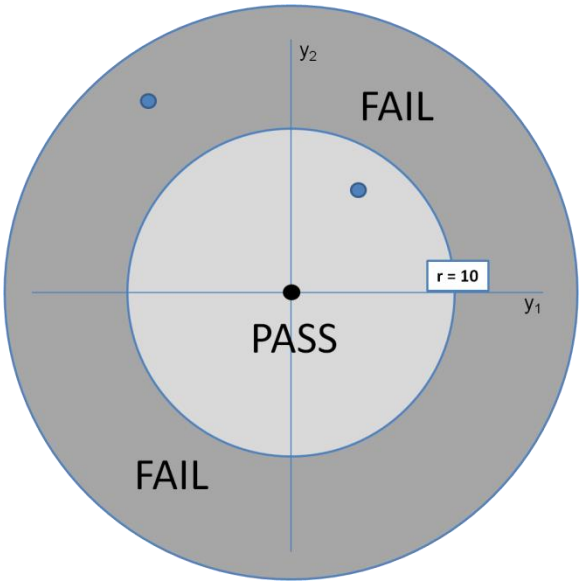



**Figure 2: Measuring using binary (Pass/Fail) response**

The purpose for this test is to evaluate the performance of the system in various simulated engagements and also to determine what factors affect the probability of success. The objective and

threshold probabilities of success $(P_s)$ for the system are 90% and 85% respectively, under all expected conditions.

The test engineers have decided to vary four different factors, Altitude, Range, Aircraft Speed, and AOA. A $2^4$ factorial design has been chosen to set up the simulated engagements, see the coded matrix below (-1 and 1 are the low and high-level values of a factor respectively). This design will allow the practitioner the ability to test what main effects and interactions terms significantly affect the response, the observed proportion of success $(\hat{p})$.

**Table 1: $2^4$ factorial design for TLE example**

| Runs | $X_1$: Altitude | $X_2$: Range | $X_3$: Aircraft Speed | $X_4$: AOA |
|---|---|---|---|---|
| 1 | -1 | -1 | -1 | -1 |
| 2 | 1 | -1 | -1 | -1 |
| 3 | -1 | 1 | -1 | -1 |
| 4 | -1 | -1 | 1 | -1 |
| 5 | -1 | -1 | -1 | 1 |
| 6 | 1 | 1 | -1 | -1 |
| 7 | 1 | -1 | 1 | -1 |
| 8 | 1 | -1 | -1 | 1 |
| 9 | -1 | 1 | 1 | -1 |
| 10 | -1 | 1 | -1 | 1 |
| 11 | -1 | -1 | 1 | 1 |
| 12 | 1 | 1 | 1 | -1 |
| 13 | 1 | 1 | -1 | 1 |
| 14 | 1 | -1 | 1 | 1 |
| 15 | -1 | 1 | 1 | 1 |
| 16 | 1 | 1 | 1 | 1 |

Consider the following graph where the response, probability of success $(P_s)$ , is graphed versus the low and high settings of $X_1$ (Altitude).

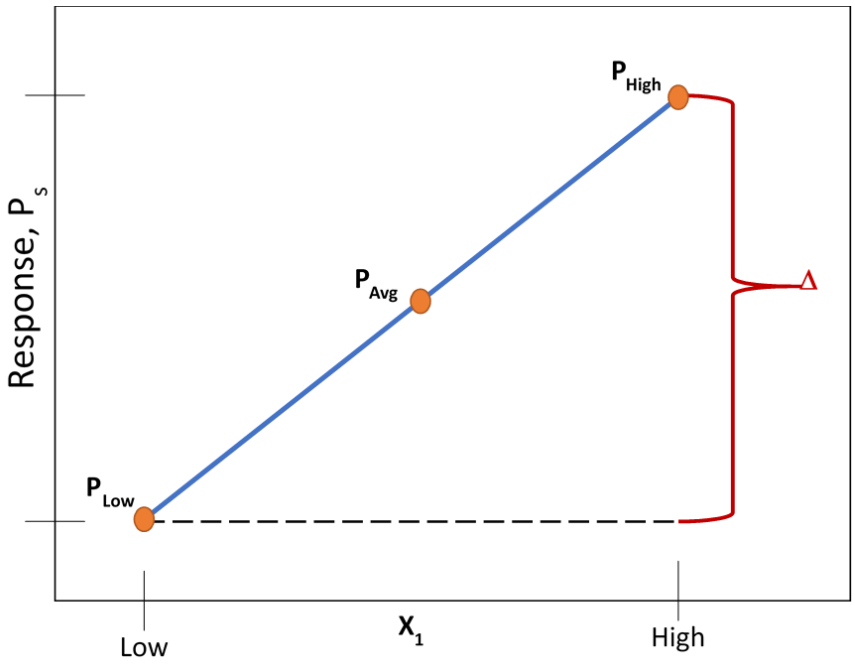



**Figure 3: Power represents the ability to detect a difference Δ between factor levels**

Power is the probability that we can detect a significant difference ($\Delta$) between $P_{Low}$ and $P_{High}$. The values for $P_{Low}$, $P_{Avg}$, $P_{High}$ and $\Delta$ can be derived from the objective and threshold requirement (90% and 85%). So for this TLE example, the objective ($P_S$) of 90% represents the $P_{Avg}$ and the threshold value of 85% represents $P_{Low}$, therefore $P_{High}$ would be 95%. Delta therefore would be ($P_{High}$ - $P_{Low}$), 10%.

The question that must now be answered is how many replicates of each design point must be run in order to achieve appropriate levels of power and confidence. Let's assume for this example that we will go with the DoD standard of 80% confidence and 80% power for a test ($\alpha = \beta = 0.2$). In this paper, we will introduce a calculator/app that will aid practitioners in answering this question. But, before doing so, let's first discuss the distribution the data comes from, the binomial distribution.

## The Binomial Distribution

The binomial distribution is a discrete probability distribution of the number of successes in a series of $n$ independent Bernoulli trials (pass/fail experiments), each trial yields success with probability $p$. The probability mass function is defined as:

$$P(Y = m) = \binom{n}{m} p^m (1-p)^{n-m} \quad (1)$$

for $m = 1,2, \ldots n$. The cumulative distribution function can be expressed as:

$$P(Y \le m) = \sum_{j=0}^{m} P\left(Y = j\right) \tag{2}$$

If a large enough sample size, $n$, is used, the binomial distribution begins to look like the normal distribution and its parameters can be approximated with the following formulas.

Mean:

$$\mu = np \tag{3}$$

Standard Deviation:

$$\sigma = \sqrt{np(1-p)} \tag{4}$$

A rule of thumb commonly used to ensure that the distribution can be approximated by the normal distribution is the "rule of five":

$$\begin{gathered} np \ge 5 \\ \text{and} \\ n(1-p) \ge 5 \end{gathered} \tag{5}$$

The farther $p$ is from 0.5 the larger $n$ needs to be in order for this approximation to work. So, for various $p$'s, the number of reps $(n)$needed are as follows:

**Table 2: Number of reps $(n)$needed versus $p$ based on the "rule of five"**

| $p$ | $n$ |
|---|---|
| 0.1 | 50 |
| 0.2 | 25 |
| 0.3 | 17 |
| 0.4 | 13 |
| 0.5 | 10 |
| 0.6 | 13 |
| 0.7 | 17 |
| 0.8 | 25 |
| 0.9 | 50 |

The following graphs provide a visual representation of how the binomial distribution behaves with varying sample sizes, $n$, while keeping $p$ at 0.1. You see that around $n = 50$ the shape of the histogram begins to look like the normal distribution curve.

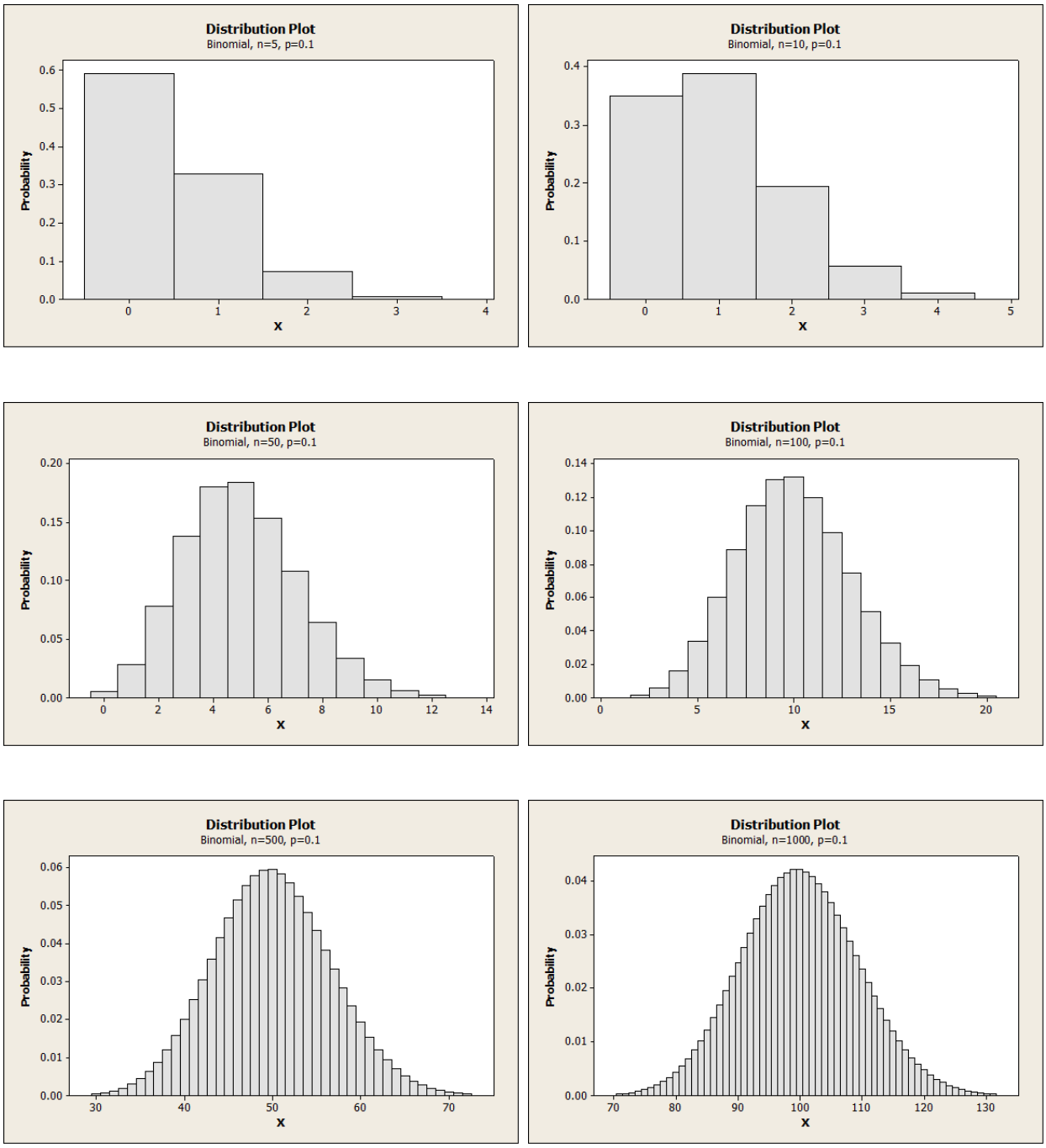
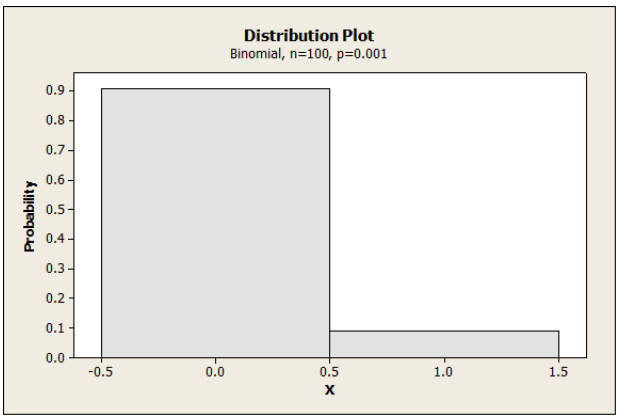
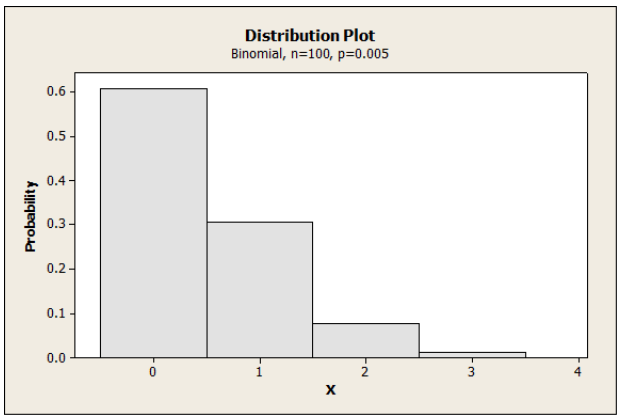
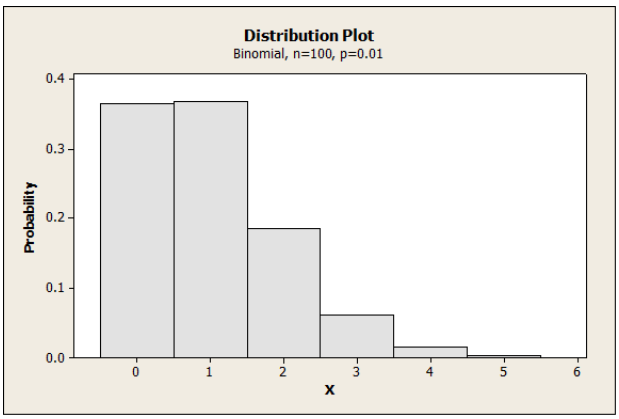


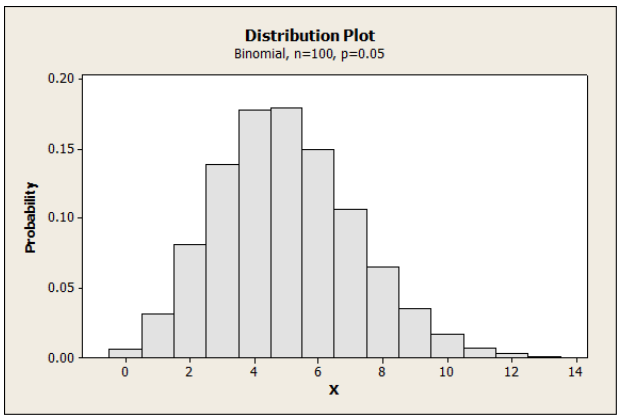

**Figure 4: Distribution plots for $p = 0.1$ for varying sample sizes, $n$**

The following graphs provide a visual representation of how the binomial distribution behaves with varying proportions $p$ and a constant sample size $n = 100$. You can see that the closer you are to the min and max values of 0 and 1 the distribution begins to looks less normal. Therefore, caution should be taken when dealing with $(P_s)$ greater than 95% or less than 1%.



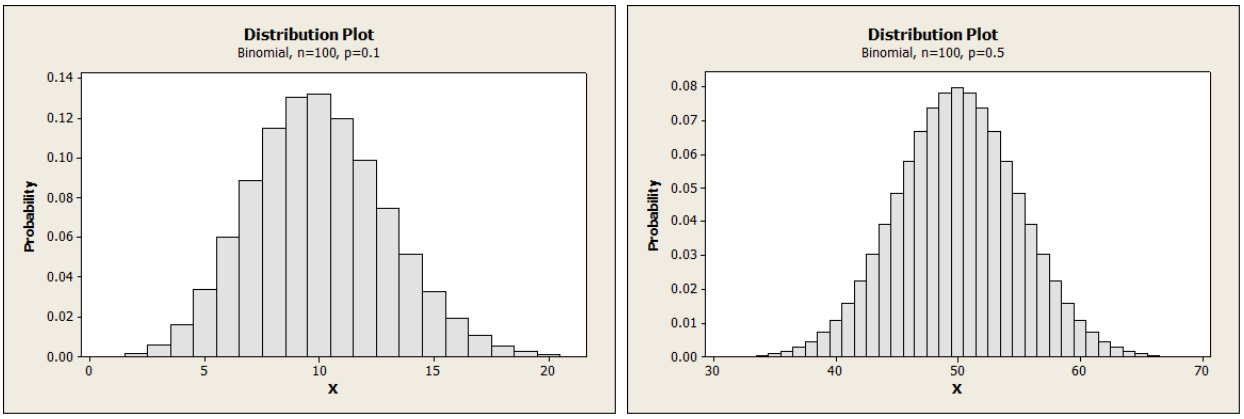



**Figure 5: Distribution plots for $n = 100$ for varying proportions, $p$**

# Method

## Method 1: Arcsine Transformation Approach

Note that the formulation for standard deviation in equation (4) is a function of $p$, which is the very response we are monitoring and wish to change by varying factor levels. Due to this, the assumption of constant variance is violated. An approach to deal with this problem is to perform a variance stabilizing transformation on the observed response $\hat{p}$. The most commonly used transformation when dealing with binomial data is the arcsine square root transformation (see equation 6). This new transformed response would be the response used in the analysis.

$$\hat{p}_1^* = \arcsin\sqrt{\hat{p}} \tag{6}$$

Bisgaard and Fuller (1995) use this transformation to derive the number of replicates needed when using a $2^{k-f}$ factorial design with binary responses. Their formulation for the signal of interest (the change in the response we wish to detect) on the transformed scale is:

$$\delta = \arcsin\left(\sqrt{\bar{p} + \frac{\Delta}{2}}\right) - \arcsin\left(\sqrt{\bar{p} - \frac{\Delta}{2}}\right) \tag{7}$$

where $\bar{p}$ is the expected proportion across the design space, and $\Delta$ is the signal in the original scale/units.

The individual point sample size is then calculated by the following formula:

$$n = \frac{(z_{1-\alpha/2} + z_{1-\beta})^2}{N\delta^2} \tag{8}$$

where $z_{1-\alpha/2}$ and $z_{1-\beta}$ are the critical z values based on the specified power and confidence, $N$ is the total number of design points, and $\delta$ is defined in equation (7).

To demonstrate further, let's use the TLE example introduced earlier with the Binary Response Calculator available on at the STAT COE website. The following is a screen shot of the calculator:

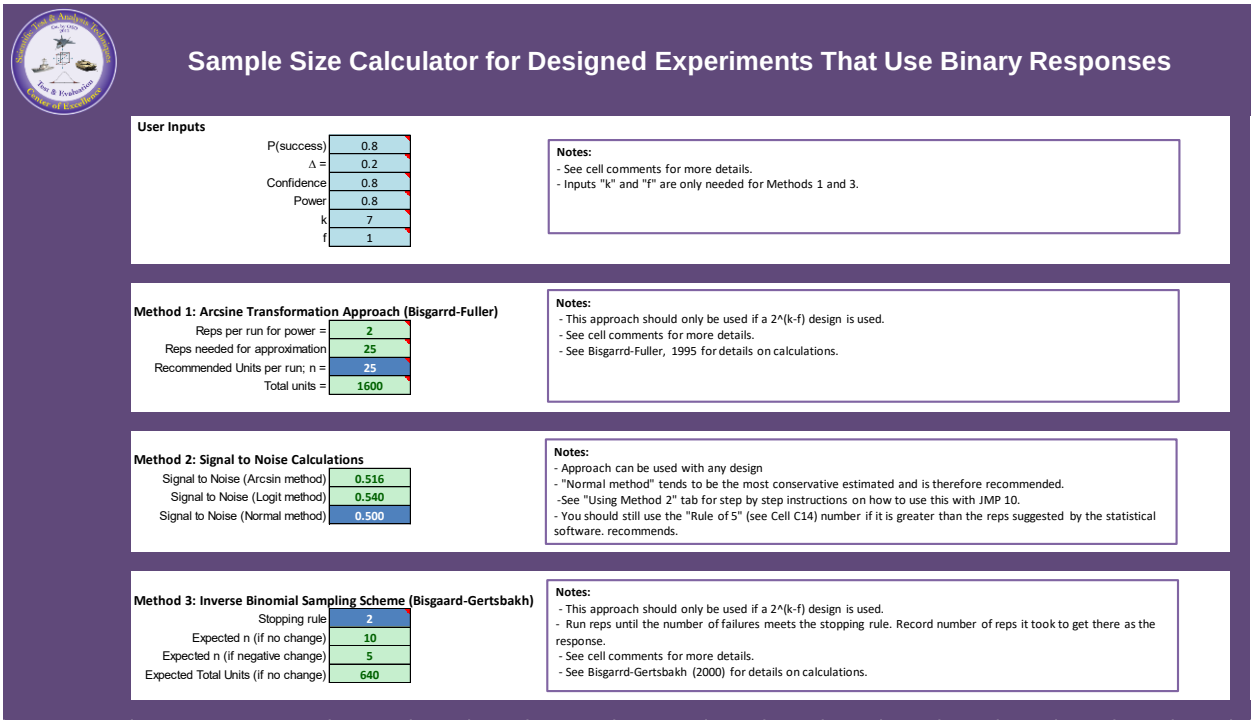



**Figure 6: Sample size calculator for binary responses screenshot**

The calculator consists of 4 sections.

***The User Input section,*** this is where the basic information about the test needs to be specified.

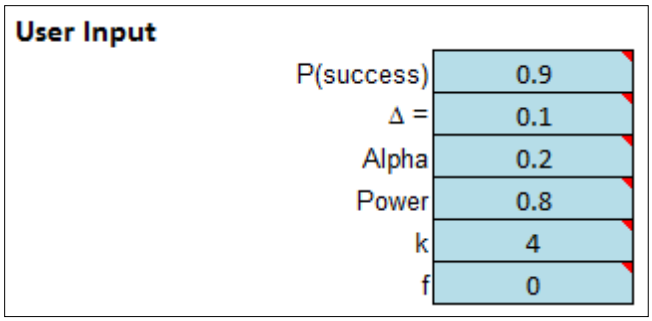



**Figure 7: User input section**

The following information must be specified in this section:

- P(success): The expected probability of success across the design space
- Δ: The signal of interest (the change in the response we wish to detect)
- Alpha: Allowable Type I error, Confidence is 1-Alpha
- Power: The probability of detecting Δ, 1-Power is the Type II error
- k: The number of factors in the design (4 in this case)
- f: The level we wish to fractionate the factorial (0 in this case since this is a full factorial)

Remember, for our example, the objective and threshold $P_s$ for the system are 90% and 85% respectively. Therefore, $P(success)$ is set to 0.9 and Δ is equal to 0.1. We're going with the DoD standard of 80% confidence and 80% power for a test ($\alpha = \beta = 0.2$). We are using a $2^4$ full factorial design so $k = 4$ and $f = 0$.

***The Method 1 section*** displays the results from applying the approach defined by Bisgaard and Fuller (1995).

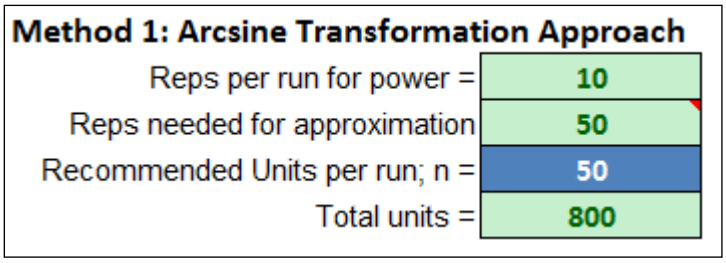



**Figure 8: Method 1- Arcsine Transformation Approach**

The following describes the output:

- Reps per run for power: Based on equation (8)
- Reps needed for approximation: Based on the "rule of 5"

- Recommended Units per run: Takes the maximum value between the reps needed for power and the reps needed for the approximation
- Total units: The total number of runs times the recommended number of reps

For the TLE example, the calculator is recommending 50 reps for each of the $2^4 = 16$ design points, resulting in a total of 800 runs.

## Method 2: Signal to Noise Calculations

Method 1, the Arcsine Transformation Approach, only works if a $2^{k-f}$ designs is used. If another type of design is used, a better approach would be to use the signal-to-noise ratio (SNR) method. The signal to noise ratio is simply the ratio between the measured change in the response we wish to detect ($\delta$, the signal of interest) and the estimated standard deviation of the system (noise). See the formula below:

$$SNR = \frac{\delta}{\sigma} \tag{9}$$

Three methods of calculating the SNR are presented in the calculator.

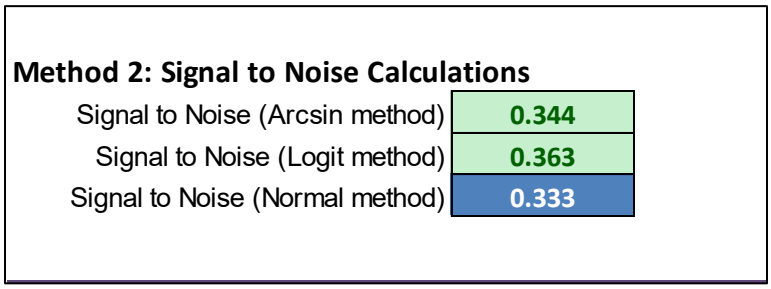



**Figure 9: Method 2- Signal to Noise Calculations**

### Arcsine Formulation

This method uses the same formulations for delta and sigma that were derived in the Bisgaard and Fuller (1995) paper. Delta (in the transformed scale) is the same as in equation (7) repeated here for convenience:

$$\delta_1 = \arcsin\left(\sqrt{\bar{p} + \frac{\Delta}{2}}\right) - \arcsin\left(\sqrt{\bar{p} - \frac{\Delta}{2}}\right)$$

The standard deviation for the arcsine transformation is as follows:

$$\sigma_1 = \frac{1}{\sqrt{4n}} = \frac{1}{2} \tag{10}$$

where $n = 1$ here since we wish to determine what the SNR is before replication.

## Logit Formulation

This approach uses the Logit transformation, which is the traditional solution used when applying logistic regression to fit a model where the dependent variable is a proportion. The transformation takes the log of the odds:

$$\hat{p}_2^* = \ln\left(\frac{\hat{p}}{1-\hat{p}}\right) \tag{11}$$

where $\hat{p}$ is the probability of an event occurring, $1-\hat{p}$ is the probability of an event not occurring, and $\frac{\hat{p}}{1-\hat{p}}$ is the odds of the event. Delta in the transformed scale is defined below:

$$\delta_2 = \left|\ln\left(\frac{p_1}{1-p_1}\right) - \ln\left(\frac{p_2}{1-p_2}\right)\right| \tag{12}$$

where $p_1 = \bar{p} + \frac{\Delta}{2}$ and $p_2 = \bar{p} - \frac{\Delta}{2}$. The standard deviation is defined as follows:

$$\sigma_2 = \sqrt{n\bar{p}(1-\bar{p})} = \sqrt{\bar{p}(1-\bar{p})} \tag{13}$$

where $n = 1$ here, since we wish to determine what the SNR is before replication.

## Normal Approximation Formulation

The final SNR formulation is based on the Normal Approximation of the binomial. This is the simplest of the formulations presented in this paper. Delta is defined as:

$$\delta_3 = \left|p_1 - p_2\right| \tag{14}$$

where $p_1 = \bar{p} + \frac{\Delta}{2}$ and $p_2 = \bar{p} - \frac{\Delta}{2}$. The standard deviation is defined the same as the logit formulation in equation 13:

$$\sigma_3 = \sqrt{n\bar{p}(1-\bar{p})} = \sqrt{\bar{p}(1-\bar{p})}$$

where $n = 1$ here, since we wish to determine what the SNR is before replication.

Figure 9 shows the results from the calculator using the TLE Example. Note that all three methods provide similar results. Table 3 shows the SNR results of the three methods when varying $p$. The Normal Approximation method consistently produces the most conservative estimate of the SNR.

**Table 3: Comparison of SNR calculation methods**

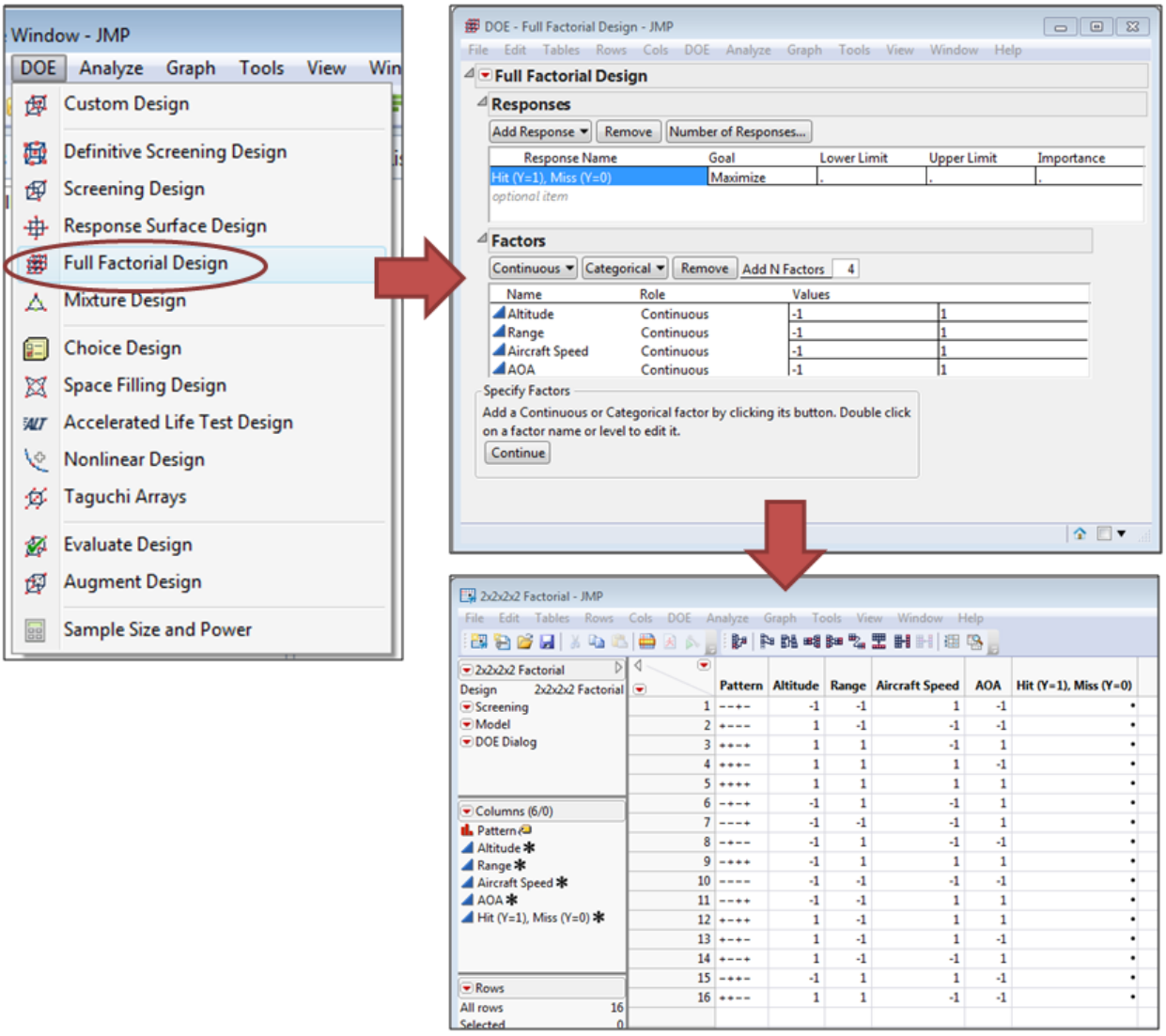
| p | Δ | SNR (arcsin) | SNR (logit) | SNR (normal) |
|---|---|---|---|---|
| 0.9 | 0.100 | 0.3444 | 0.3630 | 0.3333 |
| 0.85 | 0.100 | 0.2838 | 0.2896 | 0.2801 |
| 0.8 | 0.100 | 0.2518 | 0.2544 | 0.2500 |
| 0.75 | 0.100 | 0.2320 | 0.2334 | 0.2309 |
| 0.7 | 0.100 | 0.2189 | 0.2198 | 0.2182 |
| 0.65 | 0.100 | 0.2102 | 0.2107 | 0.2097 |
| 0.6 | 0.100 | 0.2045 | 0.2050 | 0.2041 |
| 0.55 | 0.100 | 0.2014 | 0.2017 | 0.2010 |
| 0.5 | 0.100 | 0.2003 | 0.2007 | 0.2000 |
| 0.45 | 0.100 | 0.2014 | 0.2017 | 0.2010 |
| 0.4 | 0.100 | 0.2045 | 0.2050 | 0.2041 |
| 0.35 | 0.100 | 0.2102 | 0.2107 | 0.2097 |
| 0.3 | 0.100 | 0.2189 | 0.2198 | 0.2182 |
| 0.25 | 0.100 | 0.2320 | 0.2334 | 0.2309 |
| 0.2 | 0.100 | 0.2518 | 0.2544 | 0.2500 |
| 0.15 | 0.100 | 0.2838 | 0.2896 | 0.2801 |
| 0.1 | 0.100 | 0.3444 | 0.3630 | 0.3333 |

## Using the signal-to-noise ratio (JMP 10 Demo)

Most design of experiments (DOE) software will allow you to input the SNR in order to calculate the power of the test. In this section, we will demonstrate how to use the SNR with JMP 10.

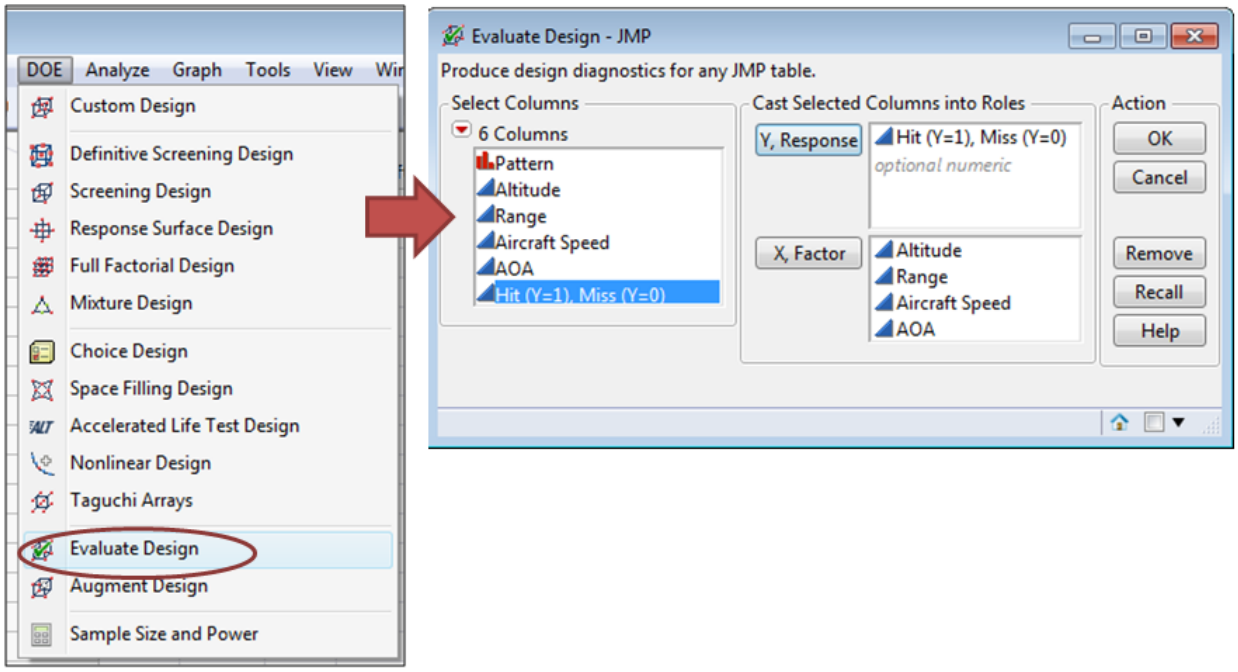
**Step 1: Create your design in JMP**

- Create a $2^4$ factorial design to match our TLE example. The columns X1, X2, X3, X4 columns represent our factors Altitude, Range, Aircraft Speed, AOA respectively. The Y column is our pass/fail response.

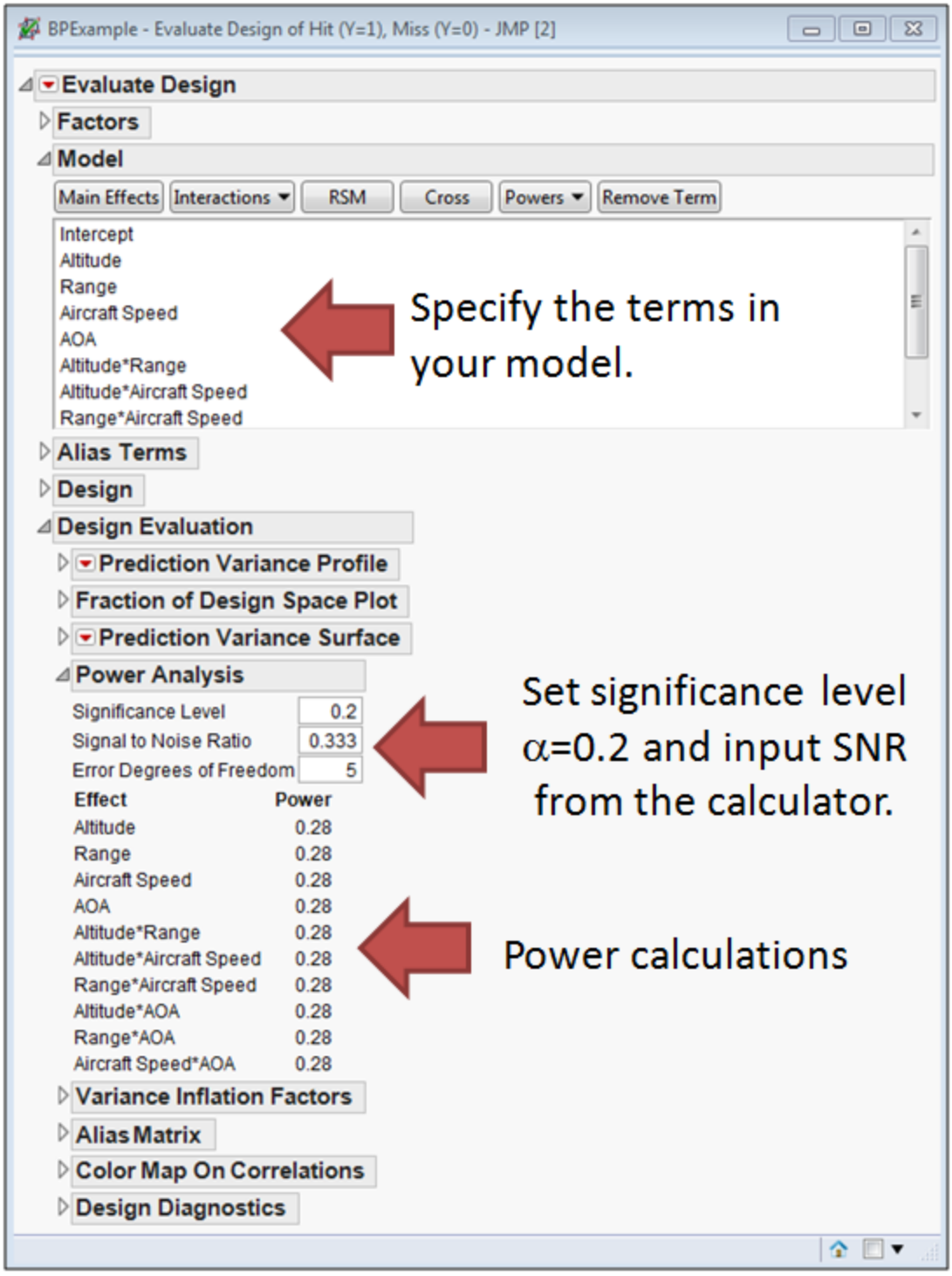



**Step 2: Evaluate design**

- Once the design is created, select DOE > Evaluate Design.



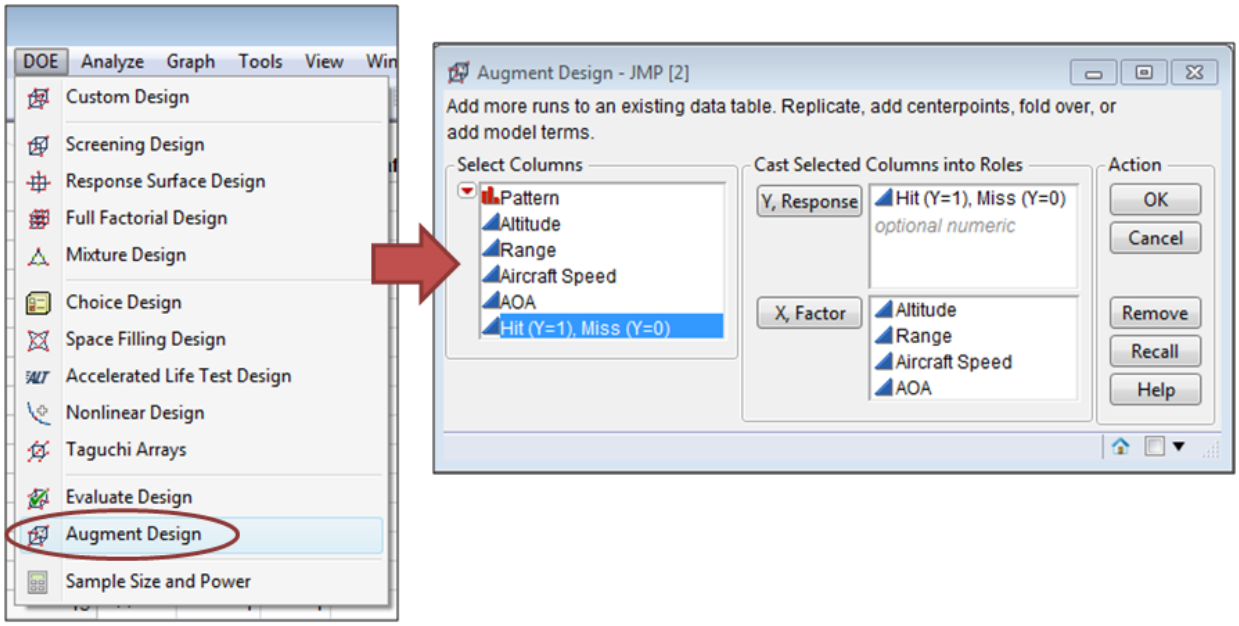

- Specify the response and factor columns and then click OK. The Evaluate Design dialog box will appear.
- In the Evaluate Design dialog box:
    - Specify the terms in your model. For this example, we are interested in main effects and two factor interactions.
    - Set significance level at $\alpha = 0.2$.
    - Input SNR from the calculator.



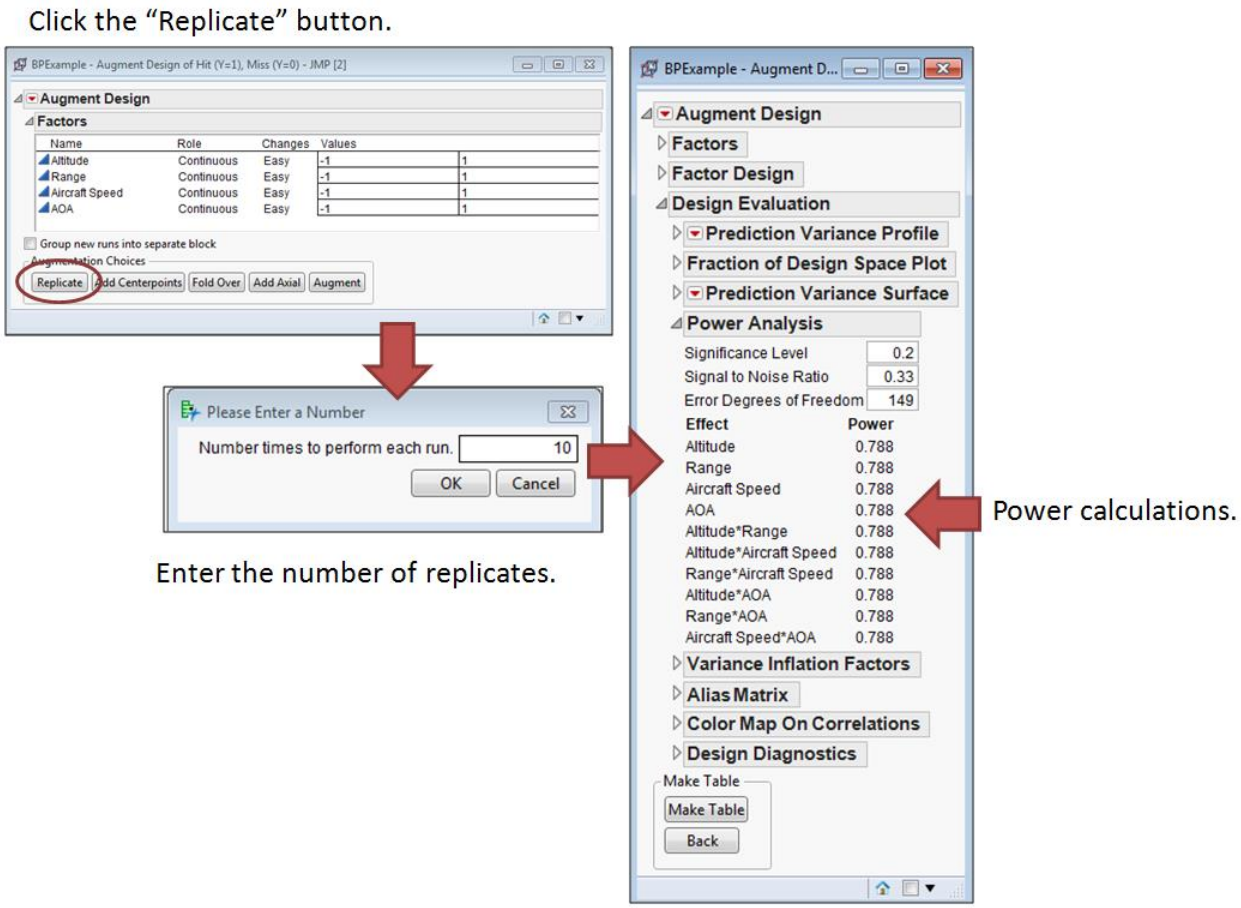

- Note: Power is 28% for all terms if we only run each setting once.

**Step 3: Augment design**

- Augment the design to add replicates and bring power up to appropriate level. For this example, we want 80%.
- Select DOE > Augment Design.




- Specify the response and factor columns and then click OK. The Augment Design dialog box will appear.




- Make sure all factors to be replicated are listed.
- Click the "replicate" button.
- Enter the number of times to replicate each design point.
- Check the Power Analysis section of the resulting design. Confirm that calculations are above or close to predetermined power objectives. In this case, 80%.
- If not, click the back button and increase the number of replicates until you achieve your power objective.

## Method 3: Inverse Binomial Sampling Scheme

The final method provided by the calculator is the Inverse Binomial Sampling Scheme proposed in Bisgaard and Gertsbakh (2000). This method can be used with a $2^{k-f}$ design where the purpose is to reduce the rate of defectives. Instead of determining a fixed sample size for each design run, this approach suggests sampling until a fixed number of defects $r$, are observed. The derivation for the stopping rule will not be covered in this paper, for more details please refer to Bisgaard and Gertsbakh (2000). The number of defects observed, *r,* is based on the number of factorial trials in a $2^{k-f}$ design,

the change in probability to detect Δ, and the fixed levels of $\alpha$ and $\beta$. The total number of reps until $r$ defects occurs is used as the response. This approach could significantly reduce the number of total runs needed if the system does not meet the $P_s$ requirement.

| Method 3: Inverse Binomial Sampling Scheme | |
|---|---|
| Stopping rule | 3 |
| Expected n (if no change) | 30 |
| Expected n (if negative change) | 15 |

**Figure 10: Inverse Binomial Sampling Scheme output**

The following describes the calculator's output:

- Stopping rule: The number of defects to observe for each design run
- Expected $n$ (if no change): Based on the estimated $P_s$, this is the expected number of reps needed to observe the stopping rule
- Expected $n$ (if negative change): If a negative change of Δ has occurred, this is the expected number of reps needed to observe the stopping rule

Note that an unequal number of reps for each design run is likely; therefore, a general linear model (GLM) or weighted least squares (WLS) approach is recommended for the analysis.

## Conclusion

Three methods to estimate the samples size needed for a designed experiment using binary responses were presented in this paper. The arcsine transformation approach can be used if a $2^{k-f}$ design is employed. The signal-to-noise method can be used for any design but requires iterative exploration of the number of replicates needed using statistical software. A JMP 10 tutorial on how to do this was provided. The Inverse Binomial Sampling Scheme method can also be used if a $2^{k-f}$ design is employed. This method could be a potential resource saving approach for system with a high expected probability of success ($P_s$), and if the goal is to simply demonstrate that the system meets that objective $P_s$. All methods presented are available for use on the Binary Response Calculator available on at the STAT COE website.

There are opportunities for future work on this subject. A Monte Carlo approach should be considered in order to produce more accurate power calculations that are robust to the experimental design used. Also, future research should explore the use of OC curves and sequential probability ratio testing in order to truncate and quickly stop testing if it is abundantly clear that the system is passing or failing the requirements.

## References

Bisgaard, Søren, and Howard T. Fuller. "Sample Size Estimates for 2k-p Designs with Binary Responses." *Journal of Quality Technology,* vol. 27, no. 4, 1995, pp. 344-354., doi: 10.1080/00224065.1995.1179616.

Bisgaard, Søren, and Ilya Gertsbakh. "2k- Experiments with Binary Responses: Inverse Binomial Sampling." *Journal of Quality Technology,* vol. 32, no. 2, 2000, pp. 148-156., doi: 10.1080/00224065.2000.11979986.

Gotwalt, C., "JMP Script for Computing Binary Power using the Logit Transformation," JMP, 2012.

Lenth, Russ. *Java Applets for Power and Sample Size.* Retrieved 2014 from http://www.stat.uiowa.edu/~rlenth/Power.

Whitcomb, Pat, and Mark Anderson. "Excel Sample Size Calculator for Binary Responses." Stat-Ease, 2000.

## Addendum (Updated July 19, 2018):

The STAT COE has created a new calculator to help estimate signal-to-noise for a binary response. The new tool allows a practitioner to evaluate the effects of increasing the number of replicates for each test design point to the signal-to-noise ratio (SNR). The following is a brief tutorial on how to use the new tool.

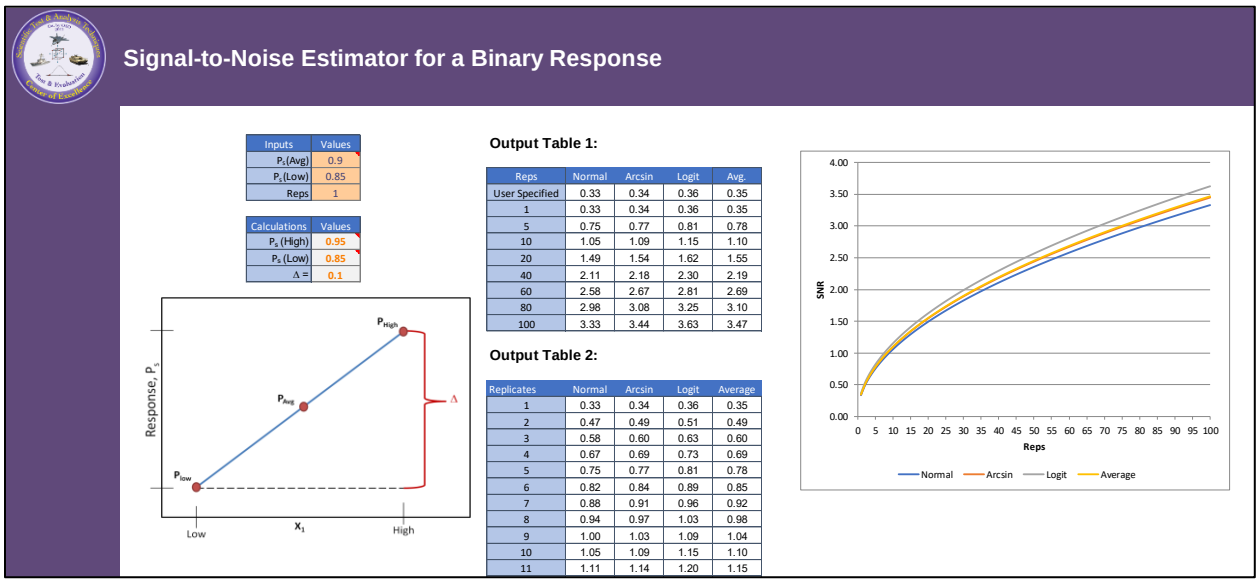



**Figure 11: SNR estimator for a binary response**

Recall our TLE Example, where the objective and threshold probability of success $(P_s)$ for the system is 90% and 85% respectively, under all expected conditions. The following graph shows the response, probability of success $(P_s)$ , versus the low and high settings of a factor $X_1$.

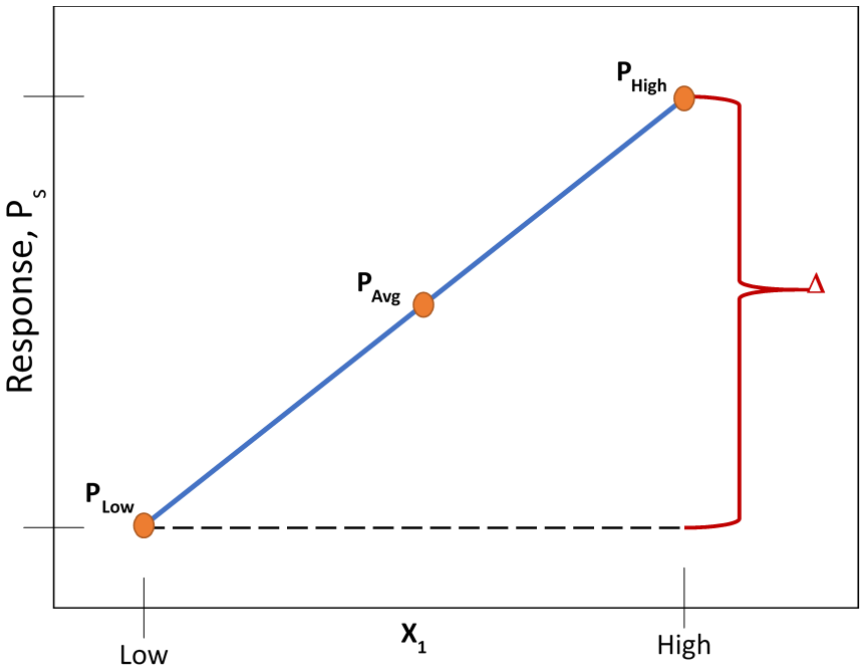



**Figure 12: Power represents the ability to detect a difference Δ between factor levels**

The values for $P_{Low}$, $P_{Avg}$, $P_{High}$, and Δ can be derived from the objective and threshold requirement. $P_{Avg}$ is our objective value 90%, $P_{Low}$ is 85%, $P_{High}$ would be 95%, and Δ therefore is 10%.

In the input section, we would enter the following information:

| Inputs | Values |
|---|---|
| $P_s$(Avg) | 0.9 |
| $P_s$(Low) | 0.85 |
| Reps | 1 |

**Figure 13: SNR calculator inputs**

The calculation section would display the following:

| Calculations | Values |
|---|---|
| $P_s$ (High) | 0.95 |
| $P_s$ (Low) | 0.85 |
| Δ = | 0.1 |

**Figure 14: SNR calculations for $P_{Low}$, $P_{High}$, and Δ**

The Output Table 1 shows what the estimated signal-to-noise would be using the three formulations described in this paper (Normal, Arcsine, and Logit). The first row shows the estimates based on the number of reps entered in the inputs sections. The following rows show how the SNR changes for differing numbers of reps (1, 5, 10, etc.).

## Output Table 1:

| Reps | Normal | Arcsin | Logit | Avg. |
|---|---|---|---|---|
| User Specified | 0.33 | 0.34 | 0.36 | 0.35 |
| 1 | 0.33 | 0.34 | 0.36 | 0.35 |
| 5 | 0.75 | 0.77 | 0.81 | 0.78 |
| 10 | 1.05 | 1.09 | 1.15 | 1.10 |
| 20 | 1.49 | 1.54 | 1.62 | 1.55 |
| 40 | 2.11 | 2.18 | 2.30 | 2.19 |
| 60 | 2.58 | 2.67 | 2.81 | 2.69 |
| 80 | 2.98 | 3.08 | 3.25 | 3.10 |
| 100 | 3.33 | 3.44 | 3.63 | 3.47 |

**Figure 15: Output Table 1, summary of the effects of reps to SNR**

Output Table 2, shows a finer or more detailed analysis of the effect replicates have on the SNR estimates.

## Output Table 2:

| Replicates | Normal | Arcsin | Logit | Average |
|---|---|---|---|---|
| 1 | 0.33 | 0.34 | 0.36 | 0.35 |
| 2 | 0.47 | 0.49 | 0.51 | 0.49 |
| 3 | 0.58 | 0.60 | 0.63 | 0.60 |
| 4 | 0.67 | 0.69 | 0.73 | 0.69 |
| 5 | 0.75 | 0.77 | 0.81 | 0.78 |
| 6 | 0.82 | 0.84 | 0.89 | 0.85 |
| 7 | 0.88 | 0.91 | 0.96 | 0.92 |

**Figure 16: Output Table 2, detailed report of the effects of reps to SNR**

A graphical representation of the results is also provided.

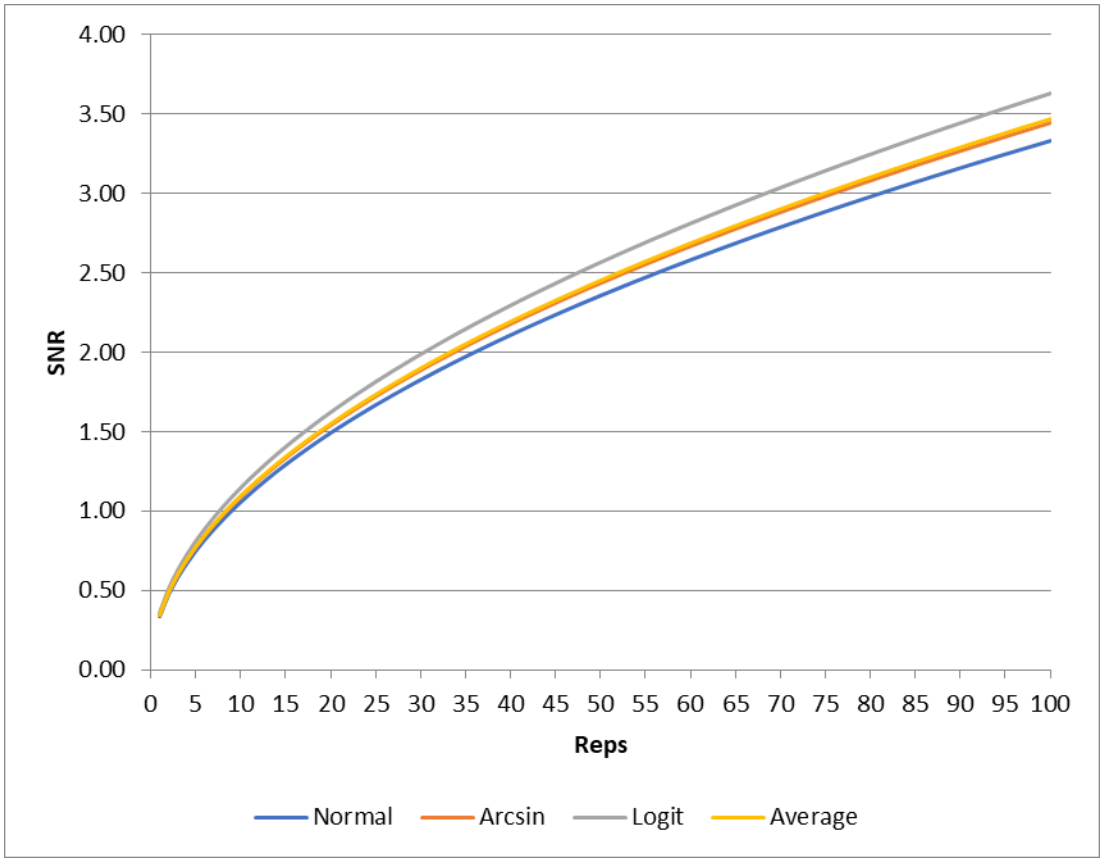



**Figure 17: Graphical representation of SNR estimation results**

In practice, a SNR of about 2 is usually a good target to aim for when it comes to SNR. This means that results that are 2 sigma away from what should be expected will be flag as significant in your analysis. In this case you see that 40 reps produce an average SNR of 2.19. In Output Table 2, we get a more precise value of 34 reps.

The next step is to create a test design and use the estimated SNR to calculate power and confidence. In our example, a $2^4$ design with 40 reps will produce the following power numbers:

**Power Analysis**

Significance Level 0.05

Anticipated RMSE 1

| Term | Anticipated Coefficient | Power |
|---|---|---|
| Intercept | 1.095 | 0.933 |
| X1 | 1.095 | 0.933 |
| X2 | 1.095 | 0.933 |
| X3 | 1.095 | 0.933 |
| X4 | 1.095 | 0.933 |
| X1*X2 | 1.095 | 0.933 |
| X1*X3 | 1.095 | 0.933 |
| X2*X3 | 1.095 | 0.933 |
| X1*X4 | 1.095 | 0.933 |
| X2*X4 | 1.095 | 0.933 |
| X3*X4 | 1.095 | 0.933 |

**Figure 18: Power calculations based on SNR estimate**

All terms are well above 80% with 95% confidence and satisfy the DoD standard of 80% confidence and 80% power.