Sample Size and Considerations for Statistical Power

Vance Oas
Nick Garcia
AFOTEC
Release Date: 28 March 17
Power analysis using theoretical SNR is bound by and sensitive to assumptions.

Power should not be the only statistical characteristic considered when creating a test design.

Always consider the system.
Overview

• Definition of Power
• Variables
• Assumptions
• Pitfalls of SNR
• Examples
• Other Design Evaluation Statistics
• Conclusion
Disclaimer

• Although program-like issues may be used to illustrate the topic, this presentation will not cover:
  – Power for specific programs
  – Battlespace conditions for specific programs
  – Measures for specific programs

• This is not a workshop on probability theory

This is a series of parametric, theoretical case studies built to highlight potential issues with power estimation using SNR.
Definition of Power

• In its simplest form, power is nothing more than a probability
  ‐ It is the area under some curve

• Miss Distance: Weapons Testing Example
  ‐ Suppose the test team decides that they need to be confident that they will detect a difference in miss distance
  ‐ Assuming actual difference between Day and Night is at least 2.5 meters
Definition of Power

• Prior to the test, we set our significance level
  – Assume it is 0.1 for this test
  – Indicated by the black “stake”
Definition of Power

• The area to the right of the stake and bounded by the red curve is 
  power

• The area to the left of the stake and bounded by the red curve is 
  called Type II error or \( \beta \) (beta)

• If the sample mean falls anywhere to the left of the stake the test 
  will not conclude that there is an effect (Fail to Reject Ho)
• In this figure we have increased the difference we need to detect to 5 and increased the significance level to 0.2

• Notice that the two curves barely overlap now

• The area to the right of the stake is approximately the entire curve (power is close to 1)
Variables

• Independent Variables ($x_1, x_2, x_3...$)
• Sample Size ($n$)
• Test Design

__________________________________________

Standard Linear Model

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_{12} + \varepsilon$$

Response = Intercept + Effect of $x_1$ * Setting of $x_1$ + Effect of $x_2$ * Setting of $x_2$ + Effect of $x_{1&2}$ * Setting of $x_{1&2}$ + Error
Assumptions

• **Alpha**
  - $\alpha = 0.2$ for Operational Test

• **Signal-to-Noise Ratio (SNR, $\delta/\sigma$)**
  - Presents the effect ($\delta$) as a multiple of unknown standard deviation ($\sigma$)
  - AFOTEC Heuristic: 80% power @ 1.5 SNR

• **Residuals come from an independent, identically distributed (i.i.d.), and normally distributed set**

• **Nuisance variables are controlled/minimized**
  - Blocking
  - Length of test period
Pitfalls of SNR

• Changing Variance (noise)
• Impact to anticipated coefficients (signals)
• Number of independent variables (factors)
• Issues with underlying response distribution

Broken assumptions associated with SNR can create issues with properly characterizing the system performance
Examples (Changing Variance)

Example 1: SNR = 1.5/Small Random Noise

\[1.5 \times X_1 + 1.5 \times X_2 + 1.5 \times X_3 +\]
\[1.5 \times X_4 + 1.5 \times X_5 + 1.5 \times X_6 +\]
\[1.5 \times X_1 \times X_2 + 1.5 \times X_1 \times X_3 +\]
\[1.5 \times X_1 \times X_4 + 1.5 \times X_1 \times X_5 +\]
\[1.5 \times X_1 \times X_6 + 1.5 \times X_2 \times X_3 +\]
\[1.5 \times X_2 \times X_4 + 1.5 \times X_2 \times X_5 +\]
\[1.5 \times X_2 \times X_6 + 1.5 \times X_3 \times X_4 +\]
\[1.5 \times X_3 \times X_5 + 1.5 \times X_3 \times X_6 +\]
\[1.5 \times X_4 \times X_5 + 1.5 \times X_4 \times X_6 +\]
\[1.5 \times X_5 \times X_6 +\]
Random Normal(0, 1)

Example 2: SNR = 1.5/Large Random Noise

\[1.5 \times X_1 + 1.5 \times X_2 + 1.5 \times X_3 +\]
\[1.5 \times X_4 + 1.5 \times X_5 + 1.5 \times X_6 +\]
\[1.5 \times X_1 \times X_2 + 1.5 \times X_1 \times X_3 +\]
\[1.5 \times X_1 \times X_4 + 1.5 \times X_1 \times X_5 +\]
\[1.5 \times X_1 \times X_6 + 1.5 \times X_2 \times X_3 +\]
\[1.5 \times X_2 \times X_4 + 1.5 \times X_2 \times X_5 +\]
\[1.5 \times X_2 \times X_6 + 1.5 \times X_3 \times X_4 +\]
\[1.5 \times X_3 \times X_5 + 1.5 \times X_3 \times X_6 +\]
\[1.5 \times X_4 \times X_5 + 1.5 \times X_4 \times X_6 +\]
\[1.5 \times X_5 \times X_6 +\]
Random Normal(0, 10)

D-Optimal Design $2^6$, Resolution V, 28 runs
Examples (Changing Variance)

Example 1: SNR = 1.5/Small Random Noise

<table>
<thead>
<tr>
<th>Parameter Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Term</td>
</tr>
<tr>
<td>Intercept</td>
</tr>
<tr>
<td>X1</td>
</tr>
<tr>
<td>X2</td>
</tr>
<tr>
<td>X3</td>
</tr>
<tr>
<td>X4</td>
</tr>
<tr>
<td>X5</td>
</tr>
<tr>
<td>X6</td>
</tr>
<tr>
<td>X1*X2</td>
</tr>
<tr>
<td>X1*X3</td>
</tr>
<tr>
<td>X1*X4</td>
</tr>
<tr>
<td>X1*X5</td>
</tr>
<tr>
<td>X1*X6</td>
</tr>
<tr>
<td>X2*X3</td>
</tr>
<tr>
<td>X2*X4</td>
</tr>
<tr>
<td>X2*X5</td>
</tr>
<tr>
<td>X2*X6</td>
</tr>
<tr>
<td>X3*X4</td>
</tr>
<tr>
<td>X3*X5</td>
</tr>
<tr>
<td>X3*X6</td>
</tr>
<tr>
<td>X4*X5</td>
</tr>
<tr>
<td>X4*X6</td>
</tr>
<tr>
<td>X5*X6</td>
</tr>
</tbody>
</table>

All showed statistical significance near a δ of 1.5

Example 2: SNR = 1.5/Large Random Noise

<table>
<thead>
<tr>
<th>Parameter Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Term</td>
</tr>
<tr>
<td>Intercept</td>
</tr>
<tr>
<td>X1</td>
</tr>
<tr>
<td>X2</td>
</tr>
<tr>
<td>X3</td>
</tr>
<tr>
<td>X4</td>
</tr>
<tr>
<td>X1*X2</td>
</tr>
<tr>
<td>X1*X3</td>
</tr>
<tr>
<td>X1*X4</td>
</tr>
<tr>
<td>X2*X4</td>
</tr>
</tbody>
</table>

Only 8/21 effects (after removing the majority of the effects) showed statistically significance or borderline

Y = Main Effects * 1.5 + Interactions * 1.5 + Random Normal(0, 10)

Examples
(Changing Anticipated Coef.)

Example 3: SNR = Mixed/Small Random Noise

\[
\begin{align*}
5 \cdot X_1 + 6 \cdot X_2 + 2 \cdot X_3 + 3 \cdot X_4 + 1 \cdot X_5 + \\
1 \cdot X_6 + 1 \cdot X_1 \cdot X_2 + 1 \cdot X_1 \cdot X_3 + 1 \cdot X_1 \\
\cdot X_4 + 1 \cdot X_1 \cdot X_5 + 1 \cdot X_1 \cdot X_6 + 1 \cdot X_2 \\
\cdot X_3 + 1 \cdot X_2 \cdot X_4 + 1 \cdot X_2 \cdot X_5 + 1 \cdot X_2 \\
\cdot X_6 + 1 \cdot X_3 \cdot X_4 + 1 \cdot X_3 \cdot X_5 + 1 \cdot X_3 \\
\cdot X_6 + 1 \cdot X_4 \cdot X_5 + 1 \cdot X_4 \cdot X_6 + 1 \cdot X_5 \\
\cdot X_6 + \text{Random Normal}(0, 1)
\end{align*}
\]

Example 4: SNR = Mixed/Large Random Noise

\[
\begin{align*}
5 \cdot X_1 + 6 \cdot X_2 + 2 \cdot X_3 + 3 \cdot X_4 + 1 \cdot X_5 + \\
1 \cdot X_6 + 1 \cdot X_1 \cdot X_2 + 1 \cdot X_1 \cdot X_3 + 1 \cdot X_1 \\
\cdot X_4 + 1 \cdot X_1 \cdot X_5 + 1 \cdot X_1 \cdot X_6 + 1 \cdot X_2 \\
\cdot X_3 + 1 \cdot X_2 \cdot X_4 + 1 \cdot X_2 \cdot X_5 + 1 \cdot X_2 \\
\cdot X_6 + 1 \cdot X_3 \cdot X_4 + 1 \cdot X_3 \cdot X_5 + 1 \cdot X_3 \\
\cdot X_6 + 1 \cdot X_4 \cdot X_5 + 1 \cdot X_4 \cdot X_6 + 1 \cdot X_5 \\
\cdot X_6 + \text{Random Normal}(0, 10)
\end{align*}
\]

D-Optimal Design $2^6$, Resolution V, 28 runs
Examples (Changing Anticipated Coef.)

Example 3: SNR = Mixed/Small Random Noise

Example 4: SNR = Mixed/Large Random Noise

\[
Y = 5X_1 + 6X_2 + 2X_3 + 3X_4 + 1X_5 + 1X_6 + \text{Interactions} \times 1 + \text{Random Normal}(0, 10)
\]

Only 7/21 effects (after removing the majority of the effects) showed statistically significance or borderline.

Larger effects are easier to estimate.

All showed statistical significance near a \( \delta \) of 1.5.
Examples (Number of Variables)

Example 5: SNR = Mixed/Small Random Noise

\[ 5 \cdot X_1 + 6 \cdot X_2 + 2 \cdot X_3 + 3 \cdot X_4 + 1 \cdot X_5 + \\
1 \cdot X_6 + 5 \cdot X_7 + 1 \cdot X_8 + 1 \cdot X_9 + \\
1 \cdot \text{Interactions} + \text{Random Normal}(0, 1) \]

Example 6: SNR = Mixed/Small(er) Random Noise

\[ 5 \cdot X_1 + 6 \cdot X_2 + 2 \cdot X_3 + 3 \cdot X_4 + 1 \cdot X_5 + \\
1 \cdot X_6 + 5 \cdot X_7 + 1 \cdot X_8 + 1 \cdot X_9 + \\
1 \cdot \text{Interactions} + \text{Random Normal}(0, 0.5) \]

D-Optimal Design $2^9$, Resolution V, 48 runs
Examples (Number of Variables)

Example 5: SNR = Mixed/Small Random Noise

Example 6: SNR = Mixed/Smaller Random Noise

Estimated larger effects very well (X1, X2, X4, and X7), and no returned p-value higher than 0.1573, and only 4 higher than 0.10

Estimated every effect with a p-value of less than 0.05

Y = 5*X1 + 6*X2 + 2*X3 + 3*X4 + 1*X5 + 1*X6 + 5*X7 + 1*X8 + 1*X9 + Interactions * 1 + Random Normal(0, 1)

Y = 5*X1 + 6*X2 + 2*X3 + 3*X4 + 1*X5 + 1*X6 + 5*X7 + 1*X8 + 1*X9 + Interactions * 1 + Random Normal(0, 0.5)
Examples (Number of Variables)

Example 7: SNR = 1/Small Noise, Fewer active effects

\[ 1 \times X_1 + 1 \times X_2 + 1 \times X_3 + 1 \times X_4 + 1 \times X_7 + 1 \times X_1 \times X_2 + 1 \times X_1 \times X_3 + 1 \times X_2 \times X_3 \]

Random Normal(0, 1)

Example 8: SNR = 1/Large Noise, Fewer Active Effects

\[ 1 \times X_1 + 1 \times X_2 + 1 \times X_3 + 1 \times X_4 + 1 \times X_7 + 1 \times X_1 \times X_2 + 1 \times X_1 \times X_3 + 1 \times X_2 \times X_3 \]

Random Normal(0, 2)

D-Optimal Design $2^9$, Resolution V, 48 runs
Examples (Number of Variables)

Example 7: SNR = 1, Fewer active effects

Y = X1 + X2 + X3 + X4 + X7 + X1*X2 + X1*X3 + X2*X3 + Random Normal(0, 1)

Excellent p-values for our selected effects. 31/45 variables remain in this model, other 14 removed and increased confidence in remaining effects.

Example 8: SNR = 0.5, Fewer active effects

Y = X1 + X2 + X3 + X4 + X7 + X1*X2 + X1*X3 + X2*X3 + Random Normal(0, 2)

Excellent p-values for our selected effects. 28/45 variables remain in this model, other 17 removed and increased confidence in remaining effects.
Issues with underlying response distribution

- Physical (e.g., a breaker on a circuit) or artificial (e.g., threshold on a measure) limits or constraints on response variables have an effect on our ability to differentiate effects.
Other Design Evaluation Statistics

- Correlation/Aliasing/Confounding

- Fraction of Design Space/Variance Inflation Factor
Other Design Evaluation Statistics

• Efficiency
  ▪ D—efficiency
    • Minimizes maximum variance of parameter estimates\(^1\)
  ▪ G—efficiency
    • Minimizes the maximum prediction variance for predicted responses\(^1\)
  ▪ A—efficiency
    • Measure for independence, minimizes average variance of parameter estimates\(^1\)

• Balance of Quality

---
\(^1\) Harman, Michael. (2014), Test Design Comparison and Selection Method Best Practice. STAT T&E COE.
Conclusion

• Power only give us an idea of how well the test will be able to characterize the system, and is very sensitive to its constituent variables

• Consider other Design Evaluation Metrics/Statistics

• Generate, generate, generate (and compare)!

Thinking through the system will always provide better power estimates, but may point a team towards additional metrics
Questions
• For a continuous response, to calculate power, 
  – First, calculate the NCP for an effect: 
    \[ \text{NCP}_i = \lambda_i = (L_i b)' \left( L_i (X'X)^{-1} L_i' \right)^{-1} L_i b \]
    ▪ X is the coding table 
    ▪ L_i is the submatrix of rows from the Identity matrix corresponding to columns of the X matrix 
      • This serves to parse only the portions of the b and X’X matrices relevant to that effect 
    ▪ b is the column matrix of anticipated coefficients 
    ▪ Find corresponding \( F_{\text{crit}} = F_{1-\alpha, df_1, df_2} \)
Calculation of Power

• For a continuous response, to calculate power,
  – Next, feed this NCP into the Non-Central F CDF

\[
CDF = \sum_{i=0}^{\infty} \left( \frac{\left(\frac{\lambda}{2}\right)^i}{i!} \times e^{-\frac{\lambda}{2}} \right)
\]

\[
\times \sum_{j=\frac{df_1}{2}+i}^{\infty} \left( \frac{\Gamma \left( \frac{df_1}{2} + \frac{df_2}{2} + i \right)}{i! \times \Gamma \left( \frac{df_1}{2} + \frac{df_2}{2} + i - j \right)} \right) \times \left( \frac{df_1 F_{crit}}{df_2 + df_1 F_{crit}} \right)^j
\]
Other Design Evaluation Statistics

- **Correlation/Aliasing/Confounding**
  - $A = (X'X)^{-1}X'Z$
  - $R = \frac{1}{n-1} \left( D^{-\frac{1}{2}} \left( X'X - \frac{1}{n} (X'1)(1'X) \right) D^{-\frac{1}{2}} \right)$

- **Fraction of Design Space/Variance Inflation Factor**
  - $VIF_i = n \times (X'X)^{-1}_{ii}$
• Efficiency

  ▪ **D—efficiency** = 100 × \( \frac{|X'X|^{\frac{1}{p}}}{N} \)
    - Uses the maximum determinant (X’X) available and minimizes maximum variance of parameter estimates

  ▪ **G—efficiency** = 100 × \( \frac{\sqrt{\frac{p}{N}}}{\sigma_M} \)
    - minimizes the maximum prediction variance for predicted responses

  ▪ **A—efficiency** = 100 × \( \frac{p}{(N \times (X'X)^{-1})} \)
    - measure for independence, minimizes average variance of parameter estimates

* \( p \): number of columns in X matrix, \( N \) = number of runs