Sample Size and Considerations for Statistical Power

Vance Oas

Nick Garcia

AFOTEC

Release Date: 28 March 17

BLUF

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.

<u>Always consider the system.</u>

Overview

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

Disclaimer

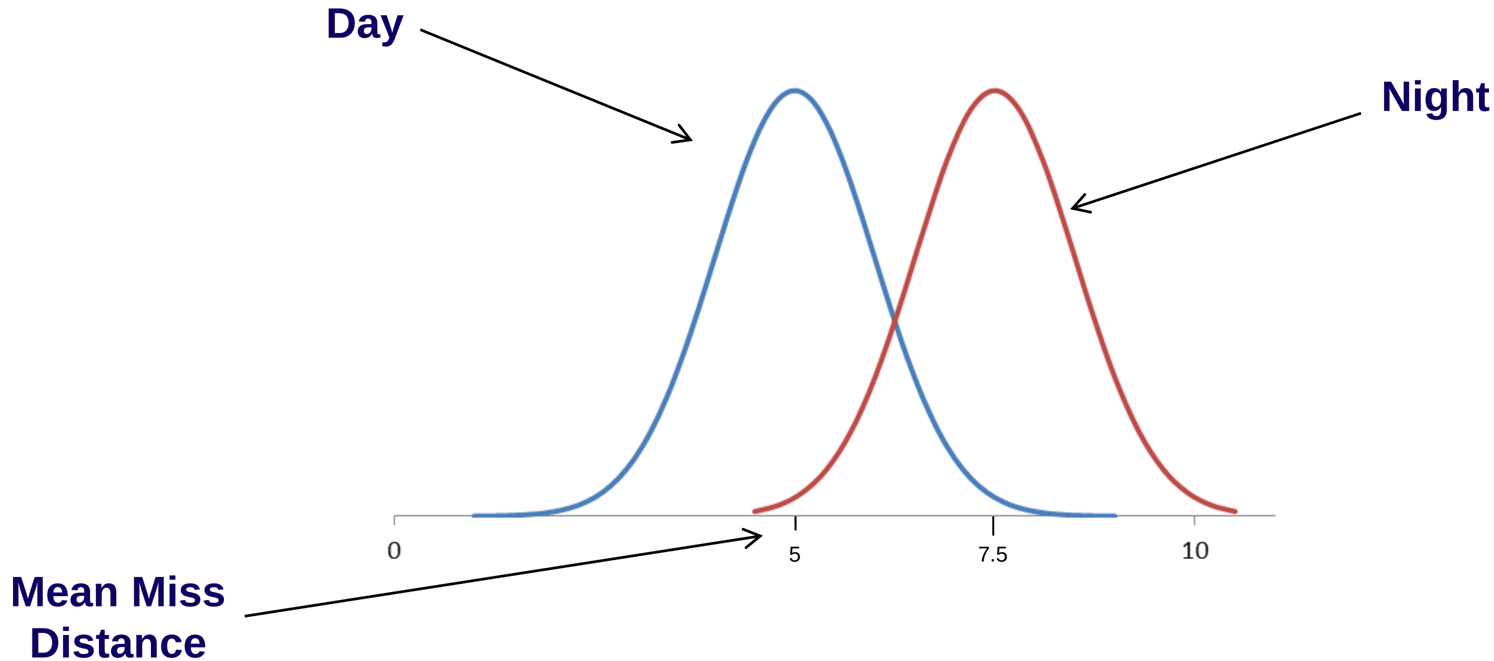
- Although program-like issues may be used to illustrate the topic, this presentation will not cover:
  - Power for specific programs
  - Battlespace conditions for specific programs
  - Measures for specific programs

- 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.

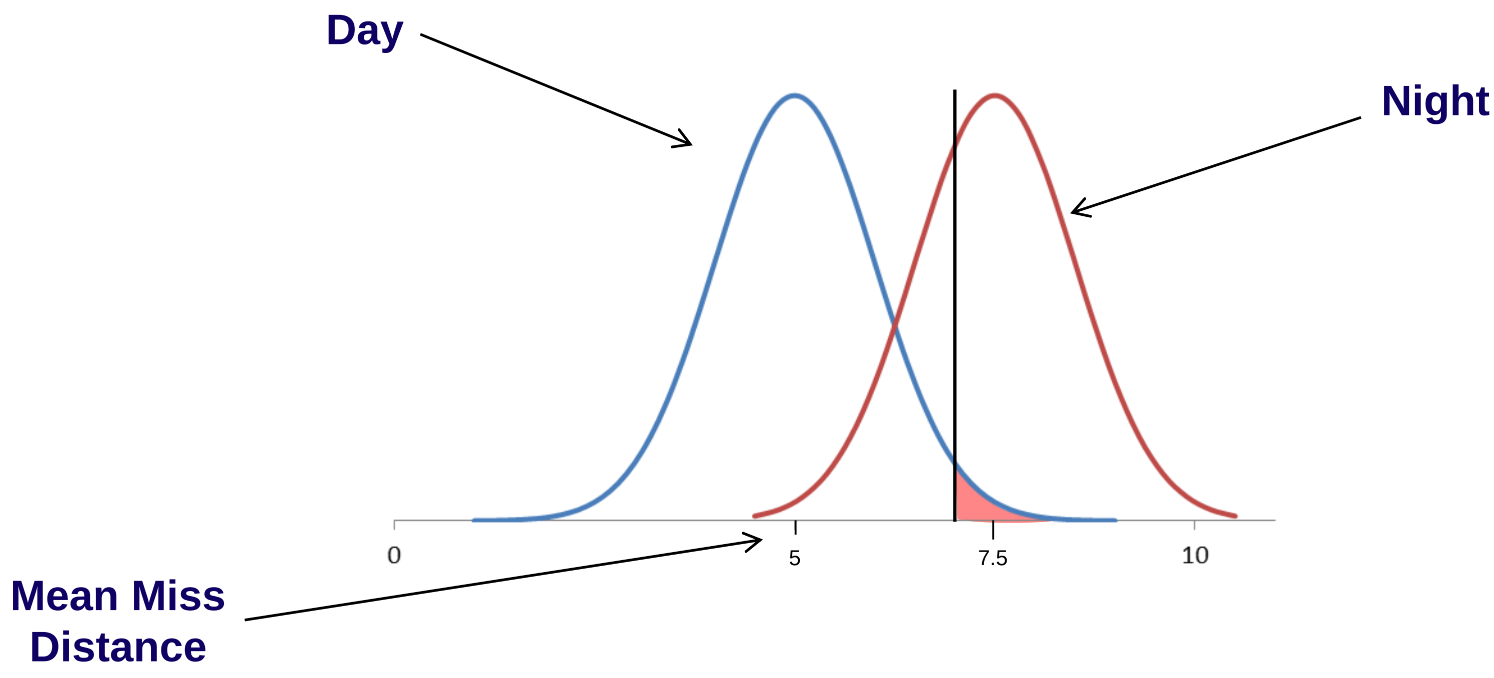
Definition of Power

- 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



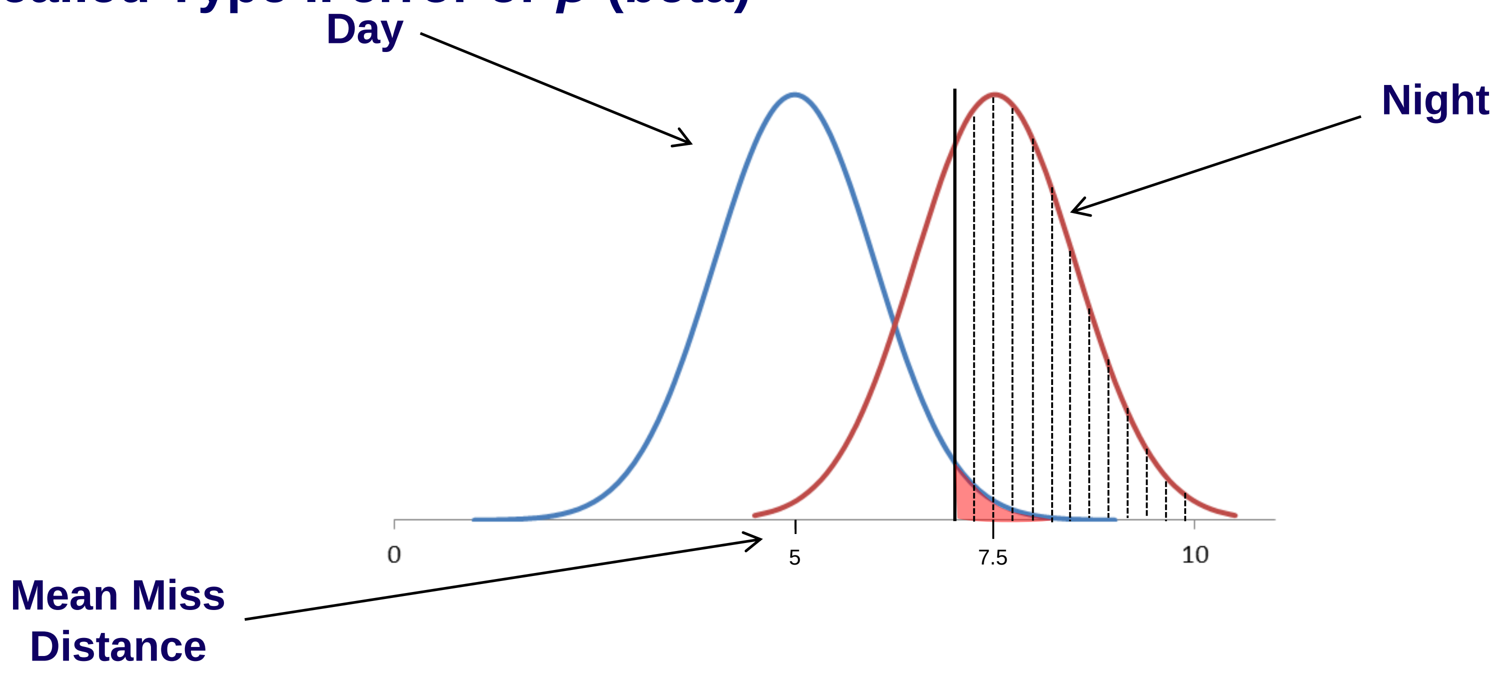
Definition of Power

- **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

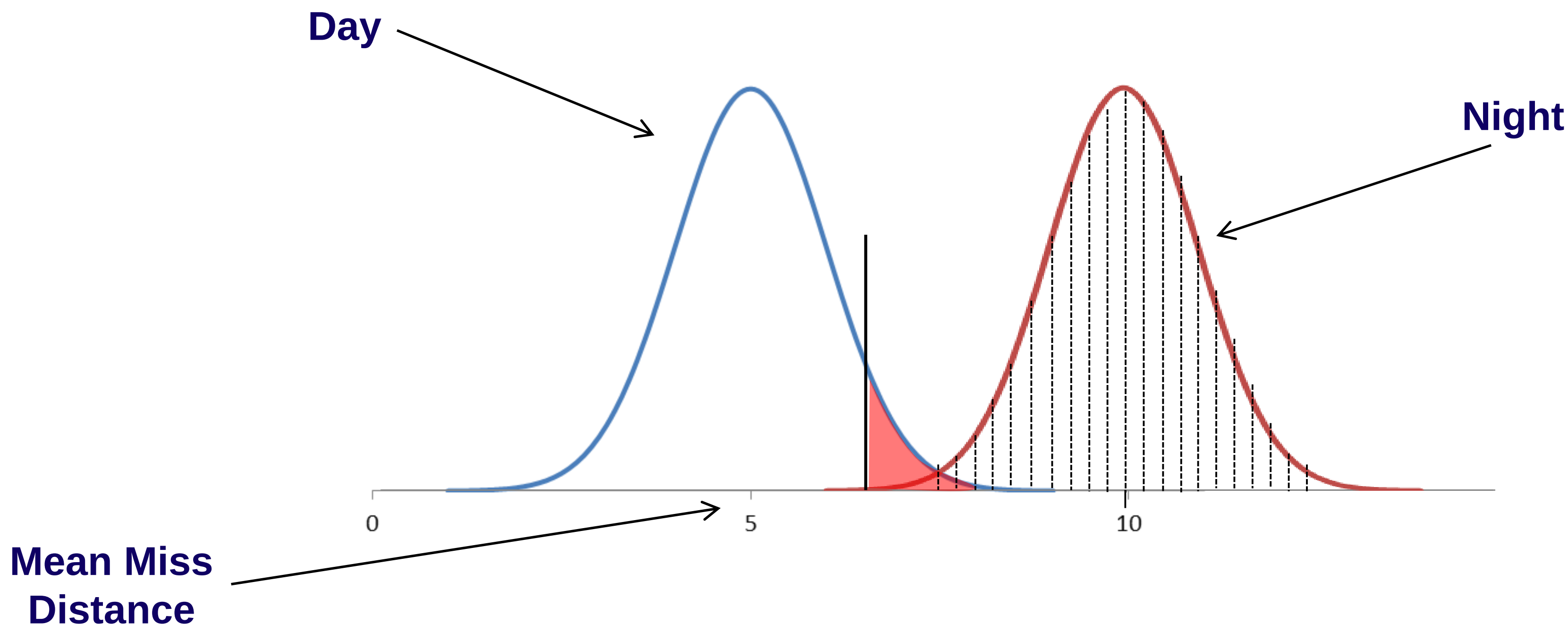
- The area to the right of the stake and bounded by the red curve is power
- The area to the left of the stake and bounded by the red curve is called Type II error or β (beta)



- 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)

Definition of Power

- 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

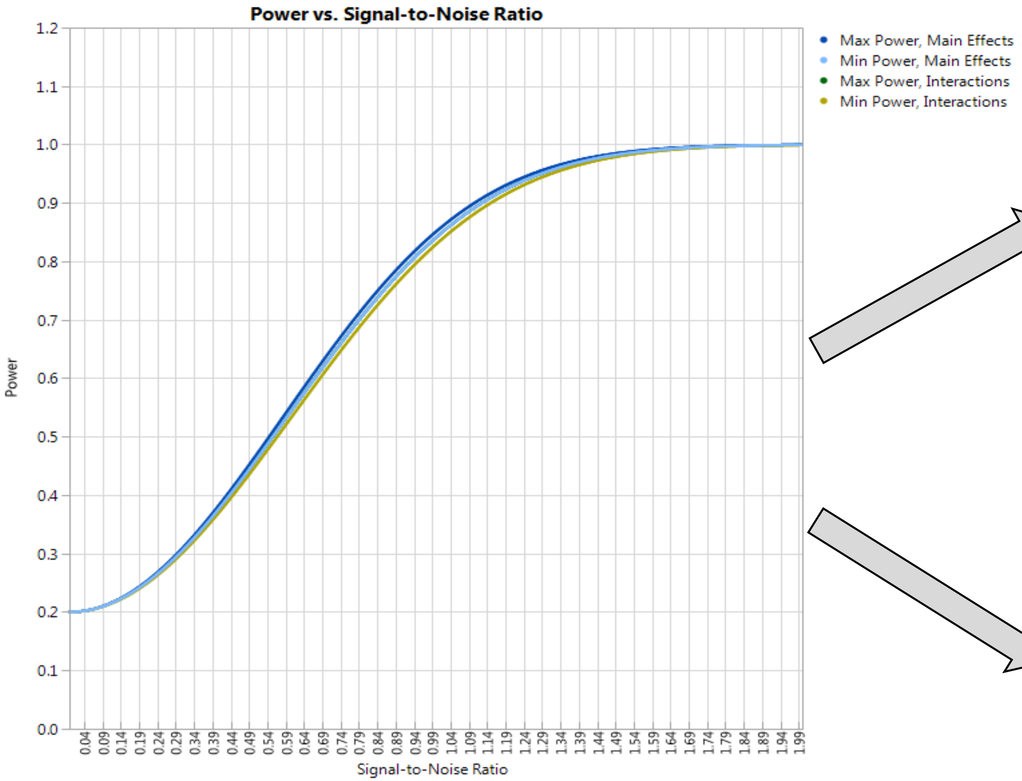
- **Alpha**
  - **α = 0.2 for Operational Test**
- **Signal-to-Noise Ratio (SNR, δ/σ)**
  - **Presents the effect (δ) as a multiple of unknown standard deviation (σ)**
  - **AFOTEC Heuristic: 80% power @ 1.5 SNR**
- **Residuals come from an independent, identically distributed (i.i.d.), and normally 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)



D-Optimal Design 2^6,
Resolution V,
28 runs

Example 1: SNR = 1.5/Small Random Noise

1.5 * :X1 + 1.5 * :X2 + 1.5 * :X3 +
1.5 * :X4 + 1.5 * :X5 + 1.5 * :X6 +
1.5 * :X1 * :X2 + 1.5 * :X1 * :X3 +
1.5 * :X1 * :X4 + 1.5 * :X1 * :X5 +
1.5 * :X1 * :X6 + 1.5 * :X2 * :X3 +
1.5 * :X2 * :X4 + 1.5 * :X2 * :X5 +
1.5 * :X2 * :X6 + 1.5 * :X3 * :X4 +
1.5 * :X3 * :X5 + 1.5 * :X3 * :X6 +
1.5 * :X4 * :X5 + 1.5 * :X4 * :X6 +
1.5 * :X5 * :X6 +
Random Normal(0, 1)

Example 2: SNR = 1.5/Large Random Noise

1.5 * :X1 + 1.5 * :X2 + 1.5 * :X3 +
1.5 * :X4 + 1.5 * :X5 + 1.5 * :X6 +
1.5 * :X1 * :X2 + 1.5 * :X1 * :X3 +
1.5 * :X1 * :X4 + 1.5 * :X1 * :X5 +
1.5 * :X1 * :X6 + 1.5 * :X2 * :X3 +
1.5 * :X2 * :X4 + 1.5 * :X2 * :X5 +
1.5 * :X2 * :X6 + 1.5 * :X3 * :X4 +
1.5 * :X3 * :X5 + 1.5 * :X3 * :X6 +
1.5 * :X4 * :X5 + 1.5 * :X4 * :X6 +
1.5 * :X5 * :X6 +
Random Normal(0, 10)

Examples (Changing Variance)

Example 1: SNR = 1.5/Small Random Noise

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-0.000912	0.24661	-0.00	0.9972
X1	1.4281723	0.251228	5.68	0.0013*
X2	1.7073158	0.251228	6.80	0.0005*
X3	1.3465355	0.24661	5.46	0.0016*
X4	1.5579553	0.251228	6.20	0.0008*
X5	1.6534084	0.251228	6.58	0.0006*
X6	1.1800554	0.251228	4.70	0.0033*
X1*X2	1.7495098	0.256579	6.82	0.0005*
X1*X3	1.7150713	0.251228	6.83	0.0005*
X1*X4	1.2474969	0.256579	4.86	0.0028*
X1*X5	1.4873403	0.256579	5.80	0.0012*
X1*X6	1.5206268	0.256579	5.93	0.0010*
X2*X3	1.169788	0.251228	4.66	0.0035*
X2*X4	1.4238502	0.256579	5.55	0.0014*
X2*X5	1.7499522	0.256579	6.82	0.0005*
X2*X6	1.4992837	0.256579	5.84	0.0011*
X3*X4	1.6366822	0.251228	6.51	0.0006*
X3*X5	1.5987737	0.251228	6.36	0.0007*
X3*X6	1.4814616	0.251228	5.90	0.0011*
X4*X5	1.6296658	0.256579	6.35	0.0007*
X4*X6	1.2902756	0.256579	5.03	0.0024*
X5*X6	1.5574534	0.256579	6.07	0.0009*

Y = Main Effects * 1.5 +
Interactions * 1.5 +
Random Normal(0, 1)

All showed statistical significance near a δ of 1.5

Example 2: SNR = 1.5/Large Random Noise[2]

Parameter Estimates

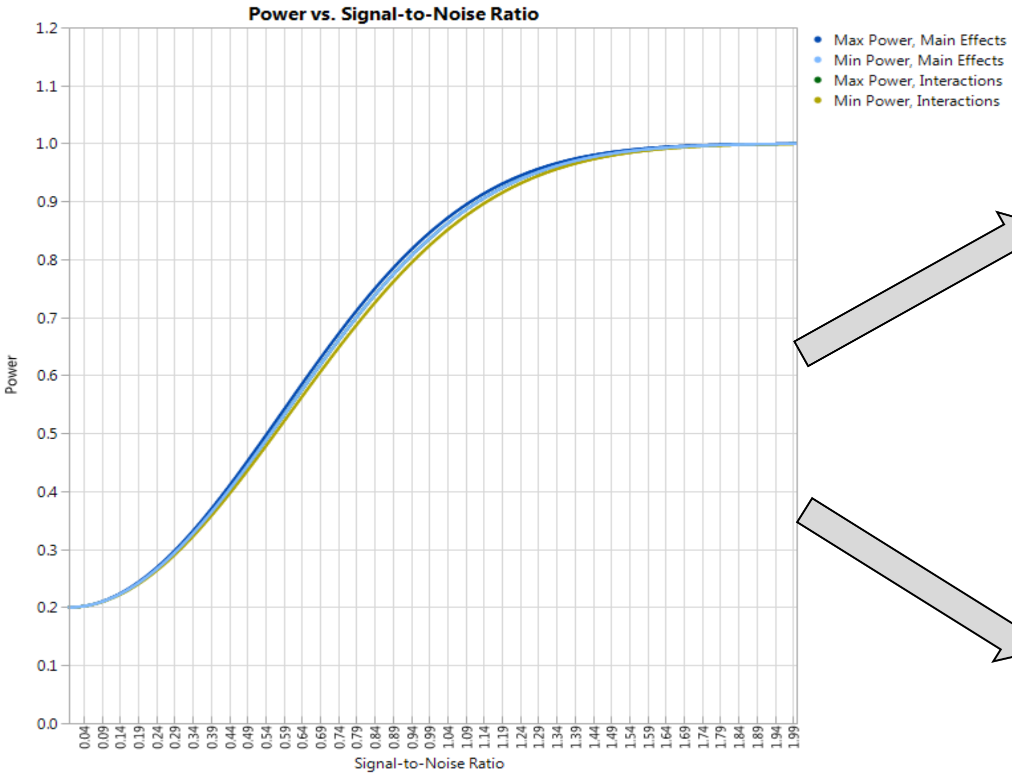
Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-2.012547	1.425876	-1.41	0.1743
X1	2.0683575	1.414193	1.46	0.1599
X2	3.9870283	1.413754	2.82	0.0109*
X3	3.5544782	1.425876	2.49	0.0221*
X5	4.0584983	1.411188	2.88	0.0097*
X1*X6	2.5741324	1.444434	1.78	0.0907
X2*X4	3.745176	1.424494	2.63	0.0165*
X2*X5	4.6327113	1.426766	3.25	0.0042*
X3*X6	4.331452	1.414193	3.06	0.0064*

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

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

[2] Montgomery, Douglas C. (2007), Design and Analysis of Experiments, *John Wiley & Sons, Inc., New York*.

Examples (Changing Anticipated Coef.)



D-Optimal Design 2^6,
Resolution V,
28 runs

Example 3: SNR = Mixed/Small Random Noise

5 * :X1 + 6 *:X2 + 2 *:X3 + 3 * :X4 + 1*:X5 + 1*:X6 + 1*:X1 * :X2 + 1 * :X1 * :X3 + 1 * :X1 * :X4 + 1 * :X1 * :X5 + 1 * :X1 * :X6 + 1 * :X2 * :X3 + 1 * :X2 * :X4 + 1 * :X2 * :X5 + 1 * :X2 * :X6 + 1 * :X3 * :X4 + 1 * :X3 * :X5 + 1 * :X3 * :X6 + 1 * :X4 * :X5 + 1 * :X4 * :X6 + 1 * :X5 * :X6 + Random Normal(0, 1)

Example 4: SNR = Mixed/Large Random Noise

5 * :X1 + 6 *:X2 + 2 *:X3 + 3 * :X4 + 1*:X5 + 1*:X6 + 1*:X1 * :X2 + 1 * :X1 * :X3 + 1 * :X1 * :X4 + 1 * :X1 * :X5 + 1 * :X1 * :X6 + 1 * :X2 * :X3 + 1 * :X2 * :X4 + 1 * :X2 * :X5 + 1 * :X2 * :X6 + 1 * :X3 * :X4 + 1 * :X3 * :X5 + 1 * :X3 * :X6 + 1 * :X4 * :X5 + 1 * :X4 * :X6 + 1 * :X5 * :X6 + Random Normal(0, 10)

Examples (Changing Anticipated Coef.)

Example 3: SNR = Mixed/Small Random Noise

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	0.3789268	0.216105	1.75	0.1301
X1	4.6654884	0.220151	21.19	<.0001*
X2	5.5645574	0.220151	25.28	<.0001*
X3	1.817613	0.216105	8.41	0.0002*
X4	2.699293	0.220151	12.26	<.0001*
X5	1.5287799	0.220151	6.94	0.0004*
X6	1.0059115	0.220151	4.57	0.0038*
X1*X2	1.1431812	0.22484	5.08	0.0023*
X1*X3	1.3970103	0.220151	6.35	0.0007*
X1*X4	1.1499705	0.22484	5.11	0.0022*
X1*X5	1.0811429	0.22484	4.81	0.0030*
X1*X6	0.9361872	0.22484	4.16	0.0059*
X2*X3	1.2357017	0.220151	5.61	0.0014*
X2*X4	1.0308738	0.22484	4.58	0.0038*
X2*X5	0.5397632	0.22484	2.40	0.0532
X2*X6	0.8964636	0.22484	3.99	0.0072*
X3*X4	0.7441731	0.220151	3.38	0.0149*
X3*X5	1.0335791	0.220151	4.69	0.0033*
X3*X6	0.9646004	0.220151	4.38	0.0047*
X4*X5	1.0215228	0.22484	4.54	0.0039*
X4*X6	1.0400882	0.22484	4.63	0.0036*
X5*X6	0.8796928	0.22484	3.91	0.0079*

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

All showed statistical significance near a δ of 1.5

Example 4: SNR = Mixed/Large Random Noise

Parameter Estimates

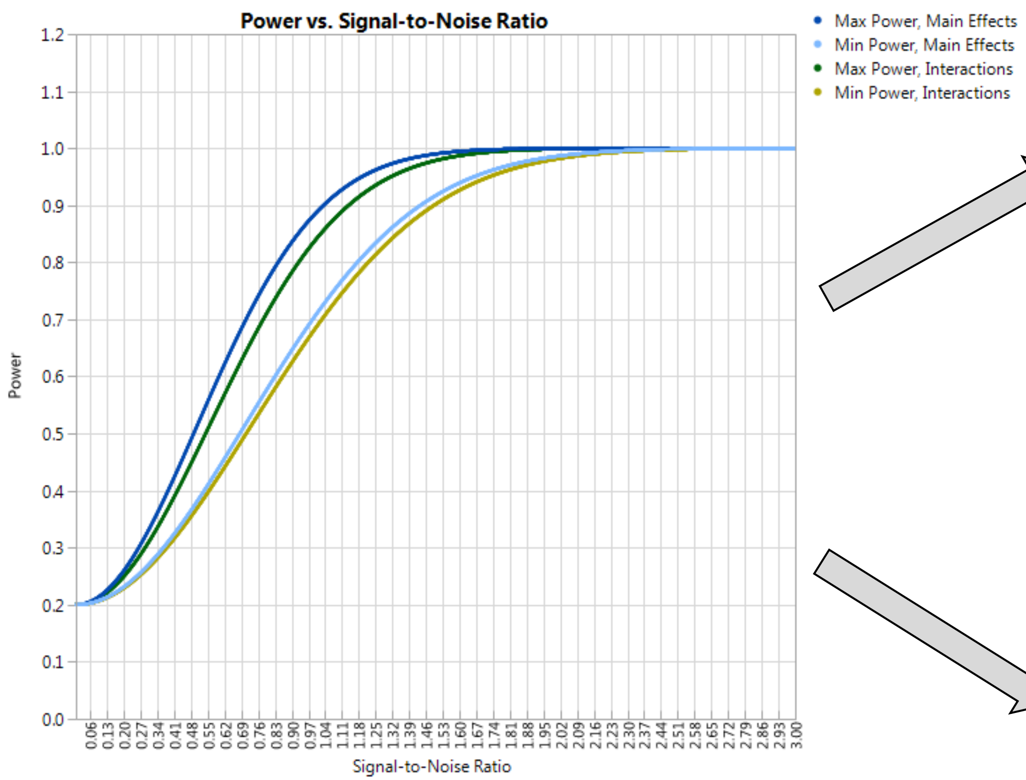
Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	0.9606652	1.20811	0.80	0.4358
X1	5.9764693	1.207979	4.95	<.0001*
X2	6.0195669	1.208242	4.98	<.0001*
X4	2.635605	1.207979	2.18	0.0412*
X6	-2.185199	1.208242	-1.81	0.0856
X1*X4	2.8194321	1.20811	2.33	0.0302*
X1*X5	2.6514972	1.20811	2.19	0.0402*
X4*X5	3.5645839	1.20811	2.95	0.0079*

Y = 5*X1 + 6*X2 + 2*X3 + 3*X4 + 1*X5 + 1*X6
+ Interactions * 1 + 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

Examples (Number of Variables)



D-Optimal Design 2^9,
Resolution V,
48 runs

Example 5: SNR = Mixed/Small Random Noise

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

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

5 * :X1 + 6 *:X2 + 2 *:X3 + 3 * :X4 + 1*:X5 + 1*:X6 + 5 * :X7 + 1*:X8 + 1*:X9 + 1*:Interactions + Random Normal(0, 0.5)

Examples (Number of Variables)

Example 5: SNR = Mixed/Small Random Noise

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	0.097319	0.218682	0.45	0.6998
X1	4.8962435	0.253134	19.34	0.0027*
X2	5.9411824	0.251165	23.65	0.0018*
X3	1.9309866	0.27034	7.14	0.0190*
X4	3.0421026	0.291519	10.44	0.0091*
X5	0.897182	0.281217	3.19	0.0858
X6	0.9799089	0.265362	3.69	0.0661
X7	5.2015701	0.30444	17.09	0.0034*
X8	0.9081694	0.264593	3.43	0.0754
X9	0.9277789	0.255349	3.63	0.0681
X1*X2	0.9358782	0.249983	3.74	0.0645
X1*X3	1.016537	0.273787	3.71	0.0655
X1*X4	0.9181712	0.257387	3.57	0.0704
X1*X5	1.0970939	0.265284	4.14	0.0538
X1*X6	0.8843252	0.251435	3.52	0.0722
X1*X7	1.1432184	0.249165	4.59	0.0444*
X1*X8	0.9067534	0.244477	3.71	0.0656
X1*X9	1.1950772	0.259278	4.61	0.0440*
X2*X3	1.2168388	0.252652	4.82	0.0405*
X2*X4	1.0927994	0.272563	4.01	0.0569
X2*X5	1.0387566	0.270286	3.84	0.0615
X2*X6	0.8327738	0.284242	2.93	0.0994
X2*X7	0.7807718	0.276971	2.82	0.1062
X2*X8	0.7804882	0.270245	2.89	0.1019
X2*X9	1.0780366	0.267109	4.04	0.0563
X3*X4	1.0420093	0.271558	3.84	0.0617
X3*X5	1.2872822	0.266693	4.83	0.0403*
X3*X6	1.2710508	0.277411	4.58	0.0445*
X3*X7	1.2538295	0.256781	4.88	0.0395*
X3*X8	1.1289033	0.255121	4.42	0.0475*
X3*X9	1.2543037	0.260294	4.82	0.0405*
X4*X5	0.9704309	0.316281	3.07	0.0918
X4*X6	1.102279	0.302661	3.64	0.0678
X4*X7	0.6591063	0.297805	2.21	0.1573
X4*X8	1.0755143	0.274835	3.91	0.0595
X4*X9	0.6647368	0.272675	2.44	0.1350
X5*X6	1.2154463	0.277191	4.38	0.0483*
X5*X7	1.1773664	0.288252	4.08	0.0550
X5*X8	0.8636515	0.259808	3.32	0.0798
X5*X9	0.8913691	0.243732	3.66	0.0673
X6*X7	1.0616675	0.282864	3.75	0.0642
X6*X8	1.1923147	0.26635	4.48	0.0465*
X6*X9	1.3669269	0.271129	5.04	0.0372*
X7*X8	1.0989721	0.2669	4.12	0.0542
X7*X9	1.0759148	0.279492	3.85	0.0613
X8*X9	0.9722652	0.240876	4.04	0.0562

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

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)

Example 6: SNR = Mixed/Smaller Random Noise

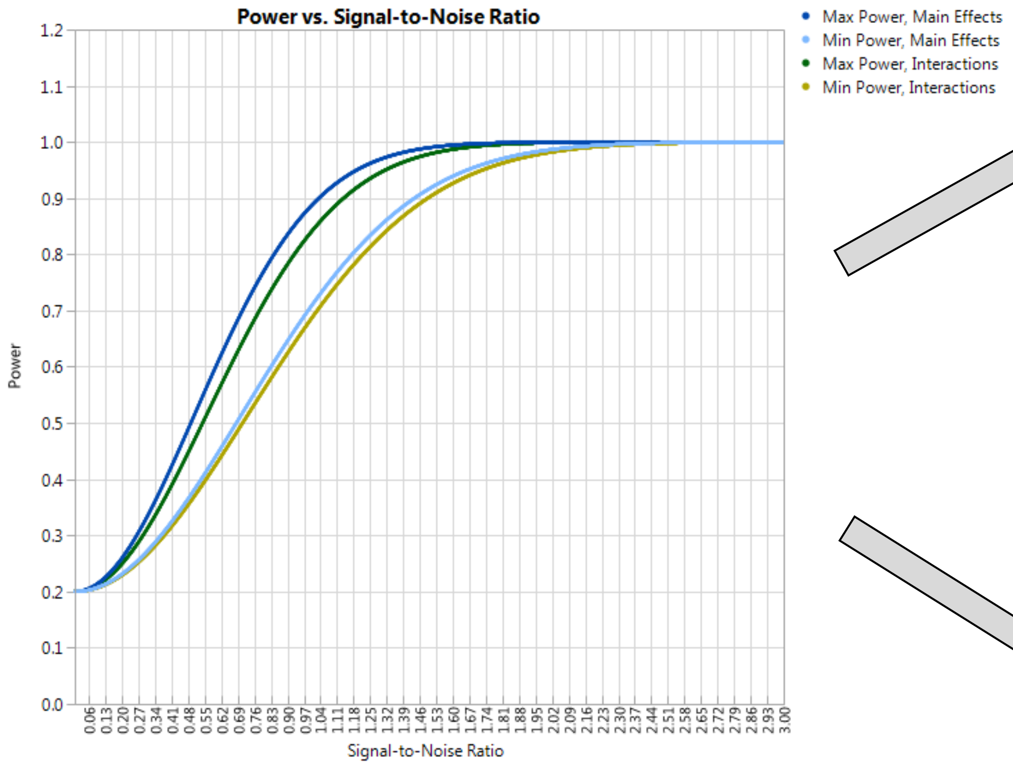
Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	0.0067494	0.081306	0.08	0.9414
X1	4.8180578	0.094116	51.19	0.0004*
X2	6.0556487	0.093383	64.85	0.0002*
X3	1.9598361	0.100513	19.50	0.0026*
X4	3.0019059	0.108387	27.70	0.0013*
X5	0.9744163	0.104557	9.32	0.0113*
X6	0.9382834	0.098662	9.51	0.0109*
X7	4.9532937	0.113191	43.76	0.0005*
X8	0.9119056	0.098376	9.27	0.0114*
X9	1.0152702	0.094939	10.69	0.0086*
X1*X2	0.9972663	0.092944	10.73	0.0086*
X1*X3	1.0017163	0.101794	9.84	0.0102*
X1*X4	1.033529	0.095697	10.80	0.0085*
X1*X5	0.8704556	0.098633	8.83	0.0126*
X1*X6	1.0950582	0.093484	11.71	0.0072*
X1*X7	0.9761633	0.09264	10.54	0.0089*
X1*X8	1.0171688	0.090897	11.19	0.0079*
X1*X9	1.1867969	0.0964	12.31	0.0065*
X2*X3	0.9740676	0.093936	10.37	0.0092*
X2*X4	0.9430883	0.101339	9.31	0.0114*
X2*X5	1.0774138	0.100492	10.72	0.0086*
X2*X6	0.9197031	0.105681	8.70	0.0129*
X2*X7	1.1460805	0.102978	11.13	0.0080*
X2*X8	1.0025513	0.100478	9.98	0.0099*
X2*X9	1.00424	0.099312	10.11	0.0096*
X3*X4	1.0435044	0.100966	10.34	0.0092*
X3*X5	0.8264563	0.099157	8.33	0.0141*
X3*X6	1.1868903	0.103142	11.51	0.0075*
X3*X7	1.0753211	0.095471	11.26	0.0078*
X3*X8	1.0251548	0.094854	10.81	0.0085*
X3*X9	1.2012949	0.096778	12.41	0.0064*
X4*X5	1.0869078	0.117594	9.24	0.0115*
X4*X6	1.0526171	0.11253	9.35	0.0112*
X4*X7	0.8892908	0.110724	8.03	0.0152*
X4*X8	0.9880957	0.102184	9.67	0.0105*
X4*X9	0.8602498	0.101381	8.49	0.0136*
X5*X6	1.028198	0.10306	9.98	0.0099*
X5*X7	1.0227952	0.107172	9.54	0.0108*
X5*X8	0.8867394	0.096597	9.18	0.0117*
X5*X9	0.94109	0.09062	10.39	0.0091*
X6*X7	0.8558807	0.105169	8.14	0.0148*
X6*X8	1.1260782	0.099029	11.37	0.0076*
X6*X9	0.9806254	0.100806	9.73	0.0104*
X7*X8	0.8394987	0.099234	8.46	0.0137*
X7*X9	1.0770816	0.103915	10.36	0.0092*
X8*X9	1.0047705	0.089558	11.22	0.0079*

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, 0.5)

Examples
(Number of Variables)



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

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

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

1 * :X1 + 1 *:X2 + 1 *:X3 + 1 * :X4 + 1 * :X7 + 1*:X1*:X2 + 1*:X1*:X3 + 1*:X2*:X3 + Random Normal(0, 2)

D-Optimal Design 2^9,
Resolution V,
48 runs

Examples (Number of Variables)

Example 7: SNR = 1, Fewer active effects

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-0.0138	0.058086	-0.24	0.8152
X1	1.0555978	0.064126	16.46	<.0001*
X2	1.1726213	0.060882	19.26	<.0001*
X3	0.94467	0.062201	15.19	<.0001*
X4	0.7937893	0.072106	11.01	<.0001*
X7	0.8646491	0.072857	11.87	<.0001*
X1*X2	0.6092814	0.063184	9.64	<.0001*
X1*X3	1.048496	0.064127	16.35	<.0001*
X1*X5	0.1689882	0.06247	2.71	0.0156*
X1*X6	-0.13767	0.062341	-2.21	0.0422*
X1*X7	-0.282124	0.063277	-4.46	0.0004*
X1*X8	0.1435384	0.062342	2.30	0.0351*
X1*X9	0.4798177	0.060814	7.89	<.0001*
X2*X3	1.1555886	0.062483	18.49	<.0001*
X2*X4	-0.22851	0.067688	-3.38	0.0039*
X2*X5	0.2724824	0.061585	4.42	0.0004*
X2*X8	0.1171309	0.065557	1.79	0.0929
X2*X9	-0.194707	0.064915	-3.00	0.0085*
X3*X5	-0.243616	0.061981	-3.93	0.0012*
X3*X6	0.3008849	0.064689	4.65	0.0003*
X3*X8	0.2608889	0.064918	4.02	0.0010*
X3*X9	-0.193055	0.063673	-3.03	0.0079*
X4*X5	0.2971485	0.076776	3.87	0.0014*
X4*X6	0.1824722	0.064503	2.83	0.0121*
X5*X6	0.1402655	0.065625	2.14	0.0484*
X5*X7	0.1827837	0.068581	2.67	0.0169*
X5*X8	-0.193841	0.063906	-3.03	0.0079*
X6*X7	-0.268923	0.064895	-4.14	0.0008*
X6*X9	0.1361832	0.068625	1.98	0.0646
X7*X8	-0.28726	0.061563	-4.67	0.0003*
X7*X9	-0.198219	0.061833	-3.21	0.0055*
X8*X9	-0.247728	0.060714	-4.08	0.0009*

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

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

Example 8: SNR = 0.5, Fewer active effects

Parameter Estimates

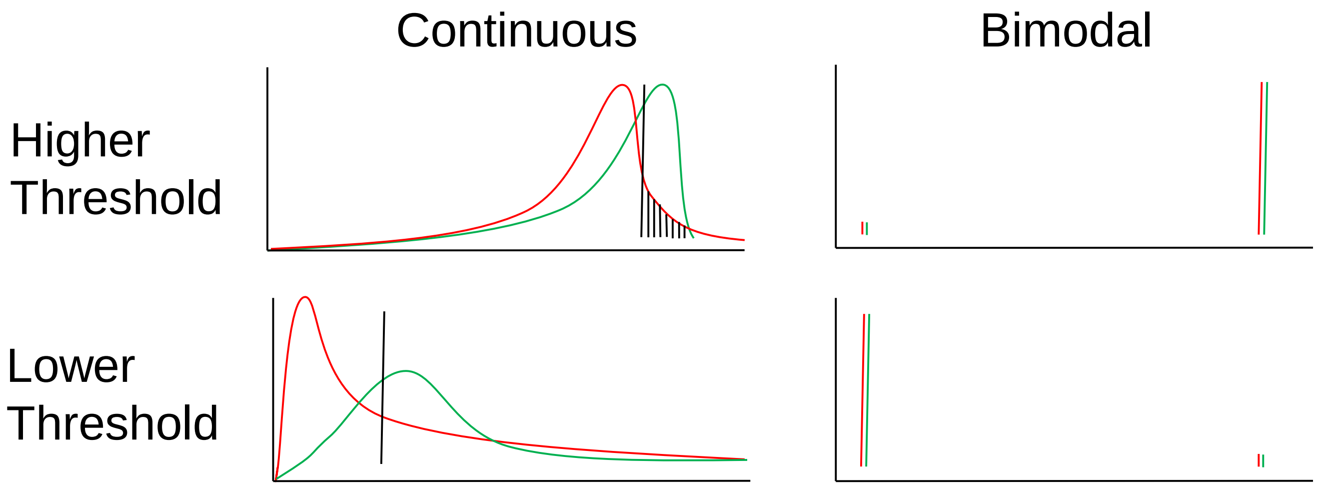
Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	0.0072438	0.137553	0.05	0.9586
X1	0.7556947	0.149921	5.04	<.0001*
X2	0.8974775	0.144324	6.22	<.0001*
X3	1.273112	0.147554	8.63	<.0001*
X4	0.9912152	0.147686	6.71	<.0001*
X5	0.4992768	0.157548	3.17	0.0051*
X7	1.9098324	0.169254	11.28	<.0001*
X1*X2	1.3404323	0.147771	9.07	<.0001*
X1*X3	0.7301233	0.154315	4.73	0.0001*
X1*X8	0.8818146	0.151109	5.84	<.0001*
X2*X3	1.7738565	0.147244	12.05	<.0001*
X2*X4	-0.505288	0.150738	-3.35	0.0033*
X2*X5	0.2882327	0.156464	1.84	0.0811
X2*X7	-0.323246	0.145087	-2.23	0.0382*
X2*X8	0.7541419	0.157193	4.80	0.0001*
X3*X4	0.4911772	0.152114	3.23	0.0044*
X3*X5	-0.294001	0.152524	-1.93	0.0690
X3*X6	0.6665825	0.158849	4.20	0.0005*
X3*X8	-0.313113	0.144715	-2.16	0.0434*
X4*X5	0.5404511	0.180644	2.99	0.0075*
X4*X6	-0.509129	0.152738	-3.33	0.0035*
X4*X7	-0.438728	0.161607	-2.71	0.0137*
X4*X8	1.0562294	0.156893	6.73	<.0001*
X5*X8	-0.320377	0.155775	-2.06	0.0537
X5*X9	-0.488048	0.145195	-3.36	0.0033*
X6*X7	-0.537179	0.158272	-3.39	0.0030*
X6*X8	0.2945943	0.156027	1.89	0.0744
X7*X8	-0.55461	0.14508	-3.82	0.0011*
X7*X9	0.5509539	0.157129	3.51	0.0024*

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

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

Examples (Underlying Response Dist.)

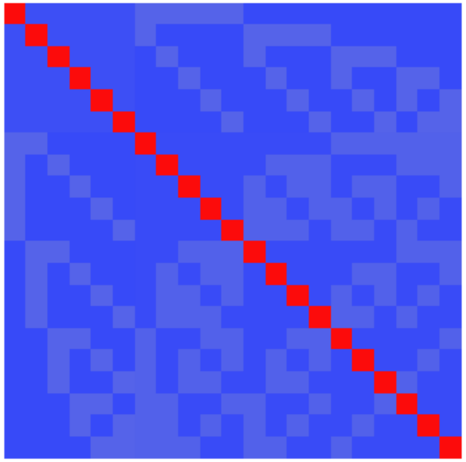
- **Issues with underlying response distribution**



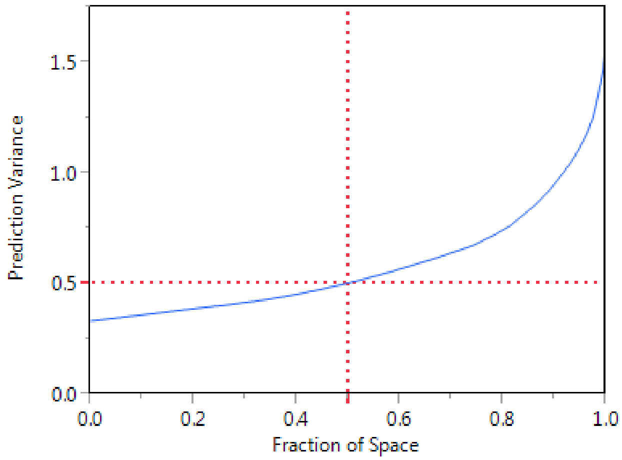
  - **Physical (e.g., a breaker on a circuit) or artificial (e.g., threshold on a measure) limits or constraints on response variables have an effect on our ability to differentiate effects**

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

?

Calculation of Power

- **For a continuous response, to calculate power,**
  - **First, calculate the NCP for an effect:**

$$\text{NCP}_i = \lambda_i = (\mathbf{L_i b})'\left(\mathbf{L_i}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}_i'\right)^{-1}\mathbf{L_i b}$$

    - **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 $F_{crit} = F_{1-\alpha,df_1,df_2}$**

Calculation of Power

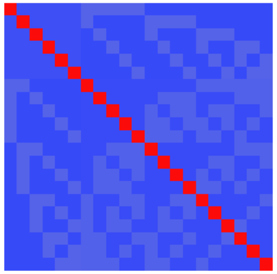
- **For a continuous response, to calculate power,**
  - **Next, feed this NCP into the Non-Central F CDF**

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

$$\left.\times \sum_{j=\frac{df_1}{2}+i}^{\infty}\left(\left(\frac{\Gamma\left(\frac{df_1}{2}+\frac{df_2}{2}+\mathrm{i}\right)}{i!\times\Gamma\left(\frac{df_1}{2}+\frac{df_2}{2}+i-j\right)}\right)\times\left(\frac{df_1F_{crit}}{df_2+df_1F_{crit}}\right)^{j}\right)\right)$$

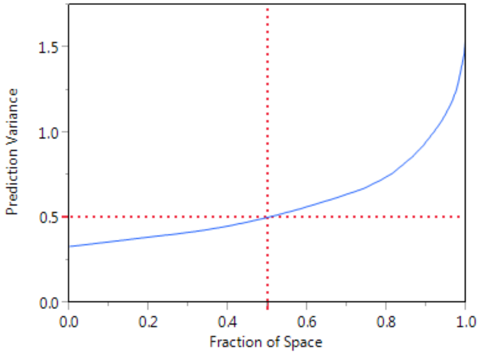
Other Design Evaluation Statistics

- Correlation/Aliasing/Confounding



- $\mathbf{A} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}$
- $\mathbf{R} = \frac{1}{n-1}\left(\mathbf{D}^{-\frac{1}{2}}\left(\mathbf{X}'\mathbf{X} - \frac{1}{n}(\mathbf{X}'\mathbf{1})(\mathbf{1}'\mathbf{X})\right)\mathbf{D}^{-\frac{1}{2}}\right)$

- Fraction of Design Space/Variance Inflation Factor



- $\mathrm{VIF}_i = n \times (\mathbf{X}'\mathbf{X})_{ii}^{-1}$

Other Design Evaluation Statistics

- **Efficiency**
  - $\text{D-efficiency} = 100 \times \left(\frac{|\mathbf{X'X}|^{\frac{1}{p}}}{\mathbf{N}} \right)$
    - **Uses the maximum determinant (X'X) available and minimizes maximum variance of parameter estimates**
  - $\text{G-efficiency} = 100 \times \frac{\sqrt{\frac{p}{N}}}{\sigma_M}$
    - **minimizes the maximum prediction variance for predicted responses**
  - $\text{A-efficiency} = 100 \times \frac{p}{\left(N \times (\mathbf{X'X})^{-1}\right)}$
    - **measure for independence, minimizes average variance of parameter estimates**

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