INSTITUTE FOR DEFENSE ANALYSES

JSM 2023: Comparing Normal and Binary D-Optimal Design of Experiments by Statistical Power

Rebecca M. Medlin, Project Leader

Addison D. Adams

August 2023

IDA Document 3000032

INSTITUTE FOR DEFENSE ANALYSES
730 East Glebe Road
Alexandria, Virginia 22305

IDA

The Institute for Defense Analyses is a nonprofit corporation that operates three Federally Funded Research and Development Centers. Its mission is to answer the most challenging U.S. security and science policy questions with objective analysis, leveraging extraordinary scientific, technical, and analytic expertise.

About This Publication
This work was conducted by the Institute for Defense Analyses (IDA) under contract HQ0034-19-D-0001, Task C9082, "Statistics and Data Science Working Group." The views, opinions, and findings should not be construed as representing the official position of either the Department of Defense or the sponsoring organization. CRP Statistics WorkGroup

Acknowledgments
The IDA Technical Review Committee was chaired by Dr. V. Bram Lillard and consisted of Dr. Kelly M. Avery, Dr. Keyla Pagan-Rivera, and Dr. Jason P. Sheldon from the Operational Evaluation Division.

For more information:
Dr. Rebecca M. Medlin, Project Leader
rmedlin@ida.org • 703-845-6731

Dr. V. Bram Lillard, Director, Operational Evaluation Division
vlillard@ida.org • (703) 845-2230

Rigorous Analysis | Trusted Expertise | Service to the Nation

INSTITUTE FOR DEFENSE ANALYSES

IDA Document 3000032

JSM 2023: Comparing Normal and Binary D-Optimal Design of Experiments by Statistical Power

Rebecca M. Medlin, Project Leader

Addison D. Adams

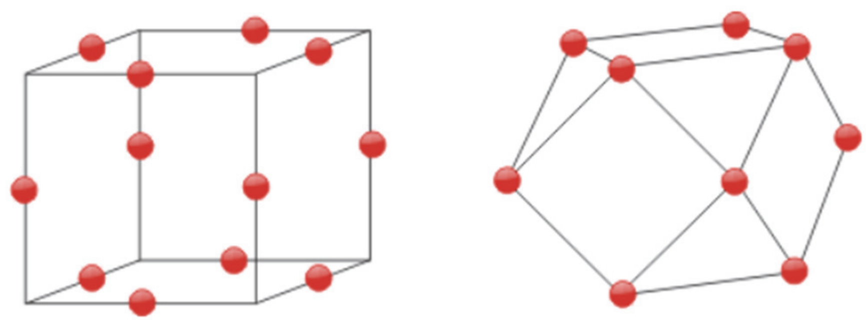
Executive Summary

In many applications binary response variables are unavoidable. Many have considered D-optimal design of experiments for generalized linear models. However, little consideration has been given to assessing how these new designs perform in terms of statistical power for a given hypothesis test. Monte Carlo simulations and exact power calculations suggest that D-optimal designs generally yield higher power than binary D-optimal designs, despite using logistic regression in the analysis after data have been collected. Results from using statistical power to compare designs contradict traditional design of experiments comparisons, which employ D-efficiency ratios and fractional design space plots. Power calculations suggest that practitioners that are primarily interested in the resulting statistical power of a design should use normal D-optimal designs over binary D-optimal designs when logistic regression is to be used in the data analysis after data collection.

JSM 2023: Comparing Normal and Binary D-Optimal Designs by Statistical Power

Addison D. Adams

August 2023



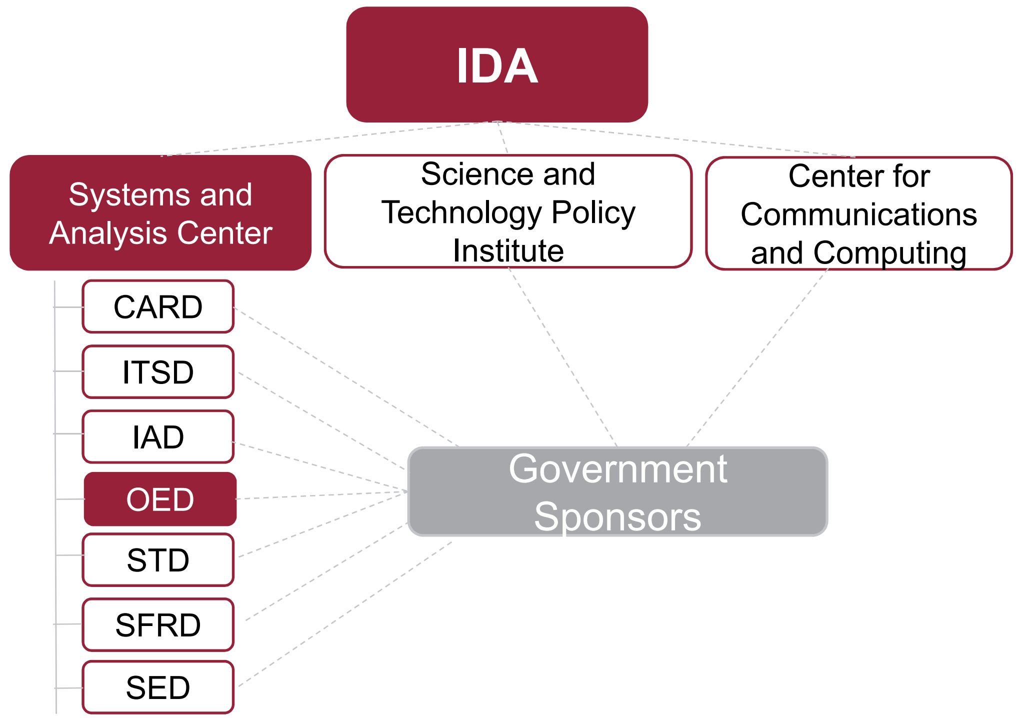
Institute for Defense Analyses
730 East Glebe Road ● Alexandria, Virginia 22305

Institute for Defense Analyses (IDA)

We answer challenging questions using scientific, technical, and analytic expertise.

IDA is a private, nonprofit corporation headquartered in Alexandria, Virginia

IDA operates three Federally Funded Research and Development Centers

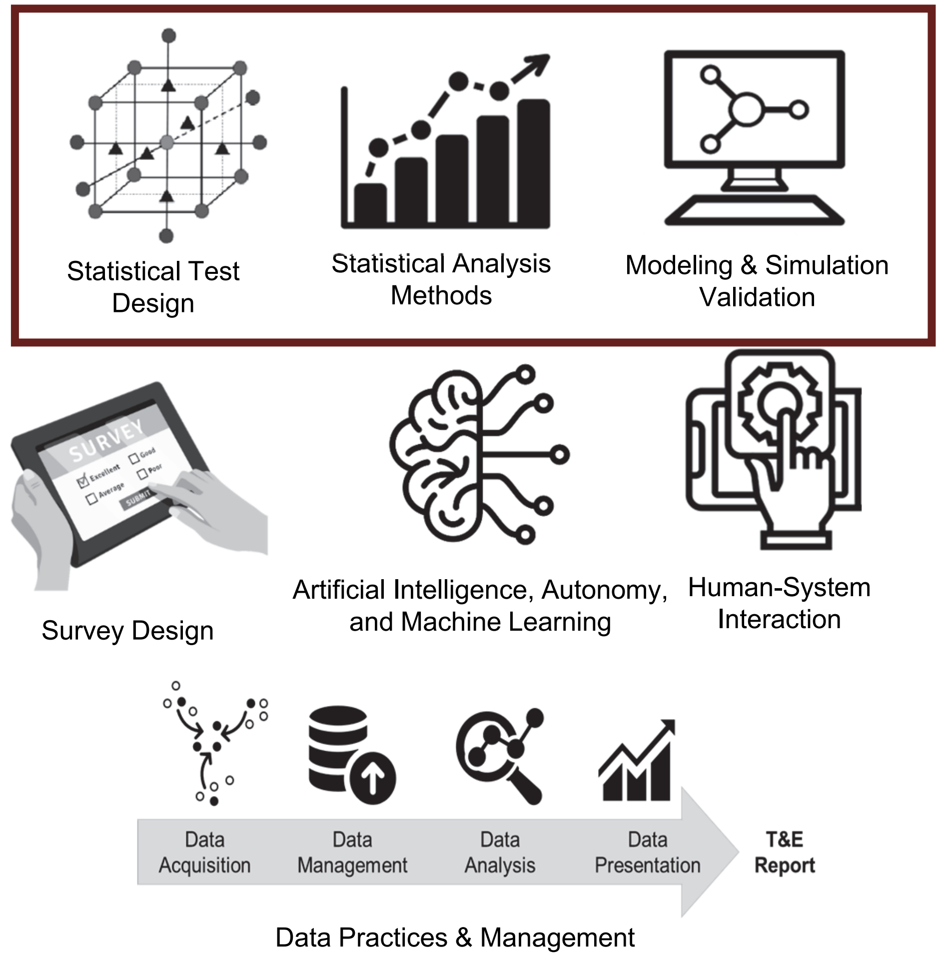


CARD – Cost Analysis and Research Division; ITSD – Information Technology and Systems Division; IAD – Intelligence Analyses Division; OED – Operational Evaluation Division; STD – Science and Technology Division; SFRD – Strategy, Forces and Resources Division; SED – System Evaluation Division

The Test Science Team provides expertise to all warfare areas in OED

We develop, apply, and disseminate statistical, psychological, and data science methodologies

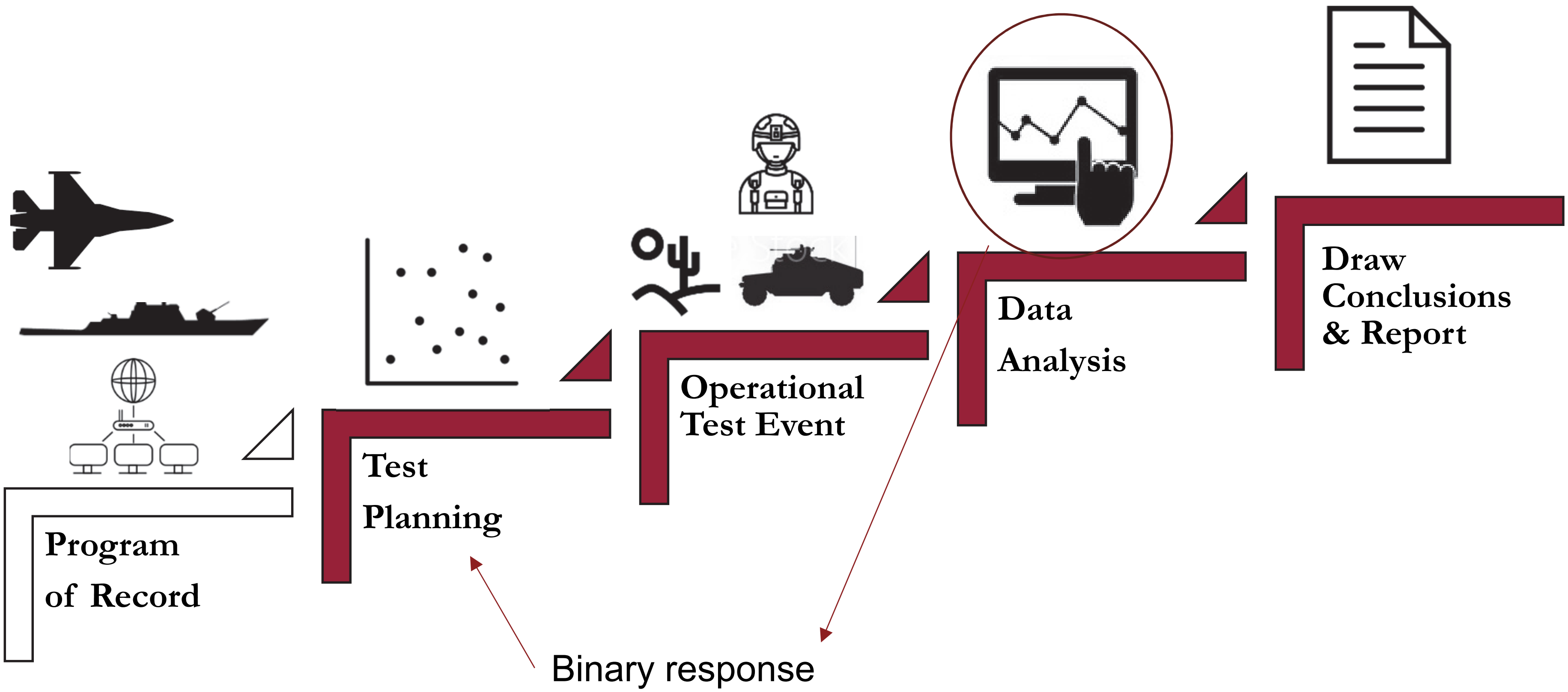
Core Areas of Expertise



TestScience
Data . Driven . Defense

OED: Operational Evaluation Division; T&E: Test and Evaluation

Central Question: How should an operational test event be planned when the response of interest is a success or failure?



- Compare two designs
  - Normal D-optimal design
  - Binary D-optimal design: special case of DOE for GLM

DOE: Design of Experiments; GLM: Generalized Linear Models

Key findings

- We advocate to compare design by statistical power.
- Standard design comparisons favor a binary design.
- Generally, a normal D-optimal design results in higher statistical power than a binary D-optimal design.

DOT&E: Director, Operational Test & Evaluation; SECDEF: Secretary of Defense

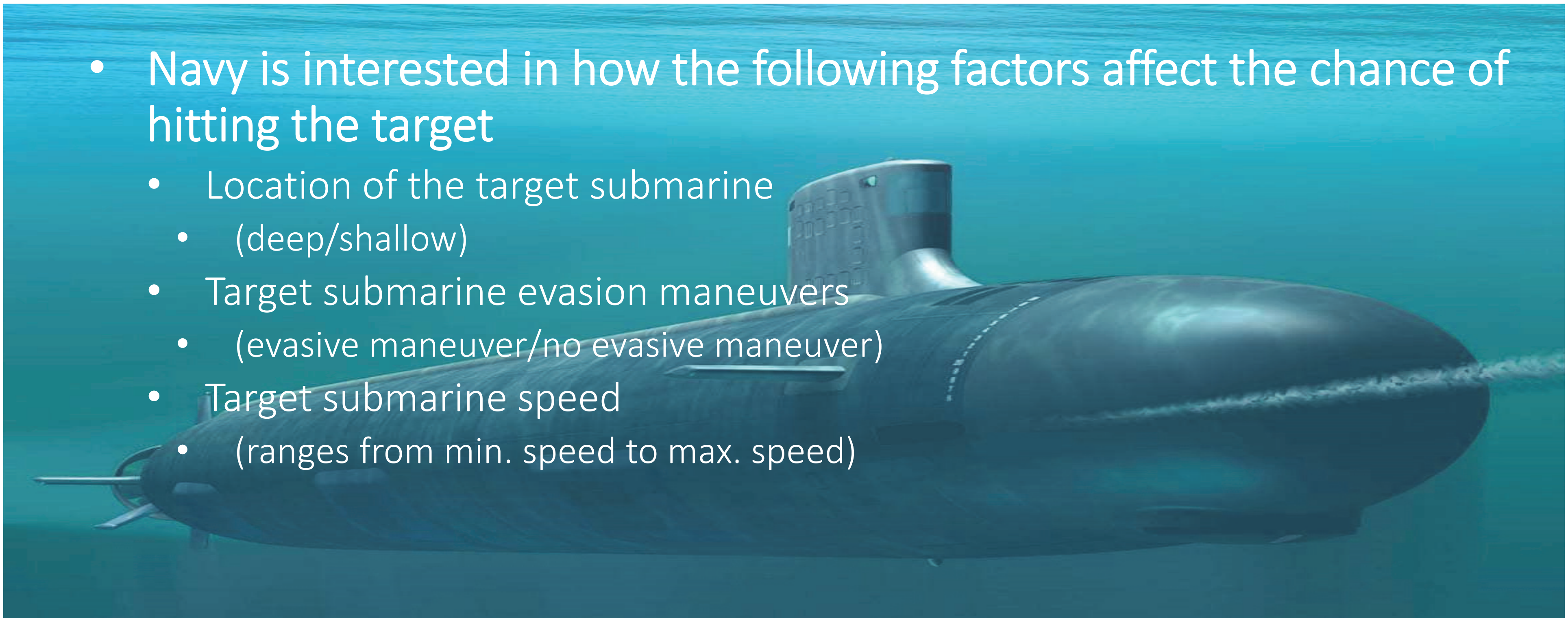
Experimental design for torpedo hit probabilities

- **D-optimal:** A design which minimizes the generalized variance of the parameter estimates

- **Hypothetical:** An operational test event is to be planned to explore condition effects on torpedo hit probability against an adversary submarine
- **Response:** Did the torpedo hit or miss the target boat?

Experimental design for torpedo hit probabilities

- Hypothetical Factors:
  - 2 categorical
  - 1 continuous



- Navy is interested in how the following factors affect the chance of hitting the target
  - Location of the target submarine
    - (deep/shallow)
  - Target submarine evasion maneuvers
    - (evasive maneuver/no evasive maneuver)
  - Target submarine speed
    - (ranges from min. speed to max. speed)

Experimental design for torpedo hit probabilities

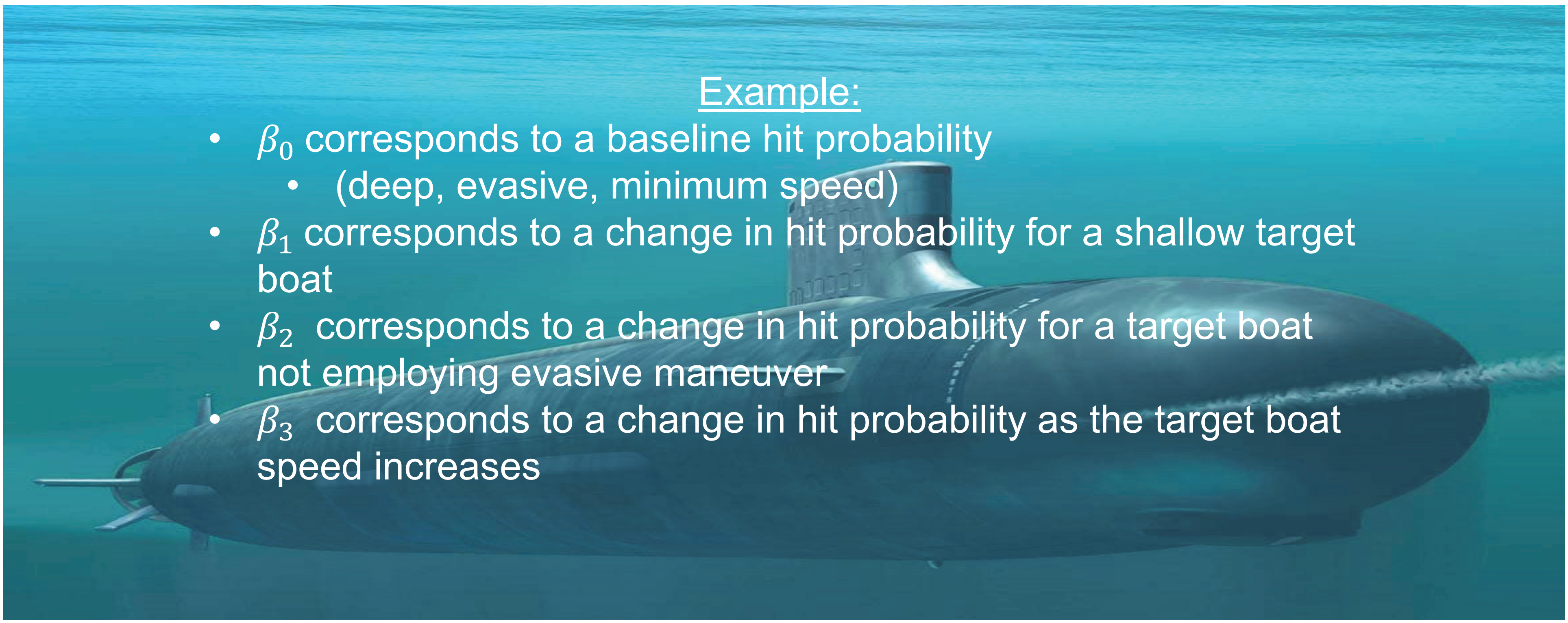
Analysis Model:

$$y_i \sim Bernoulli(\pi_i)$$

$$\text{where } \pi_i = \frac{\exp(x_i^T \boldsymbol{\beta})}{1 + \exp\left(x_i^T \boldsymbol{\beta}\right)}$$

Hypothesis Test:

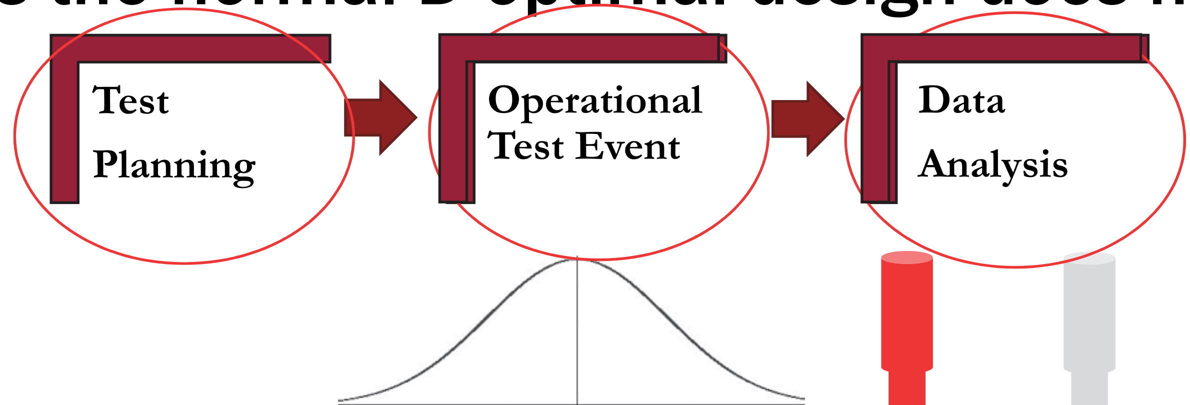
$H_0: \beta_1 = \beta_2 = \beta_3 = 0$ vs H_A: *At least one* $\beta_i \neq 0$ *for* $i = 1,2,3$



Example:

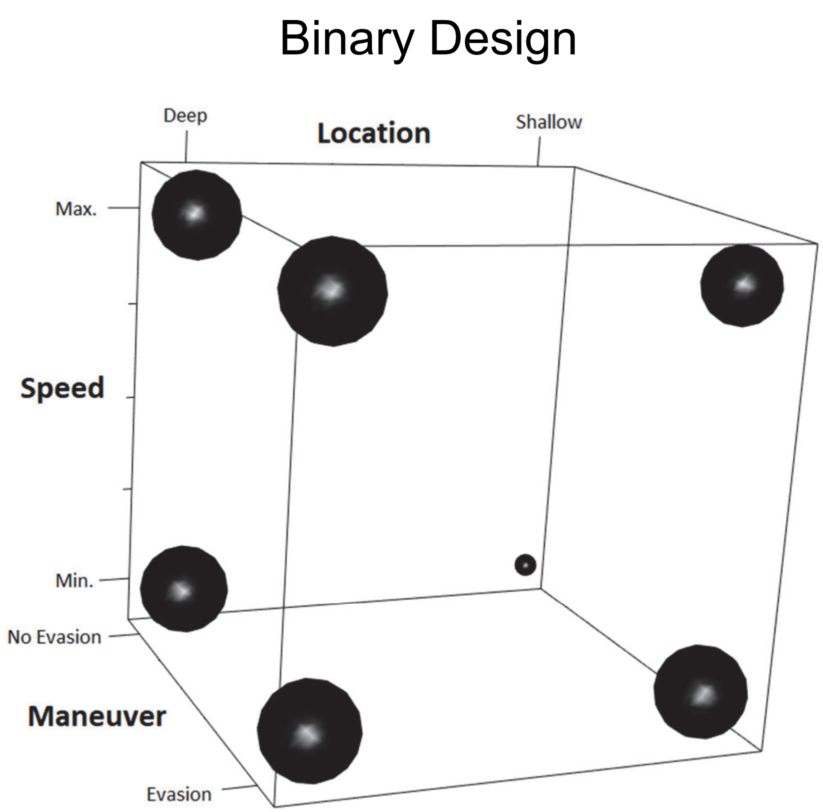
- β_0 corresponds to a baseline hit probability
  - (deep, evasive, minimum speed)
- β_1 corresponds to a change in hit probability for a shallow target boat
- β_2 corresponds to a change in hit probability for a target boat not employing evasive maneuver
- β_3 corresponds to a change in hit probability as the target boat speed increases

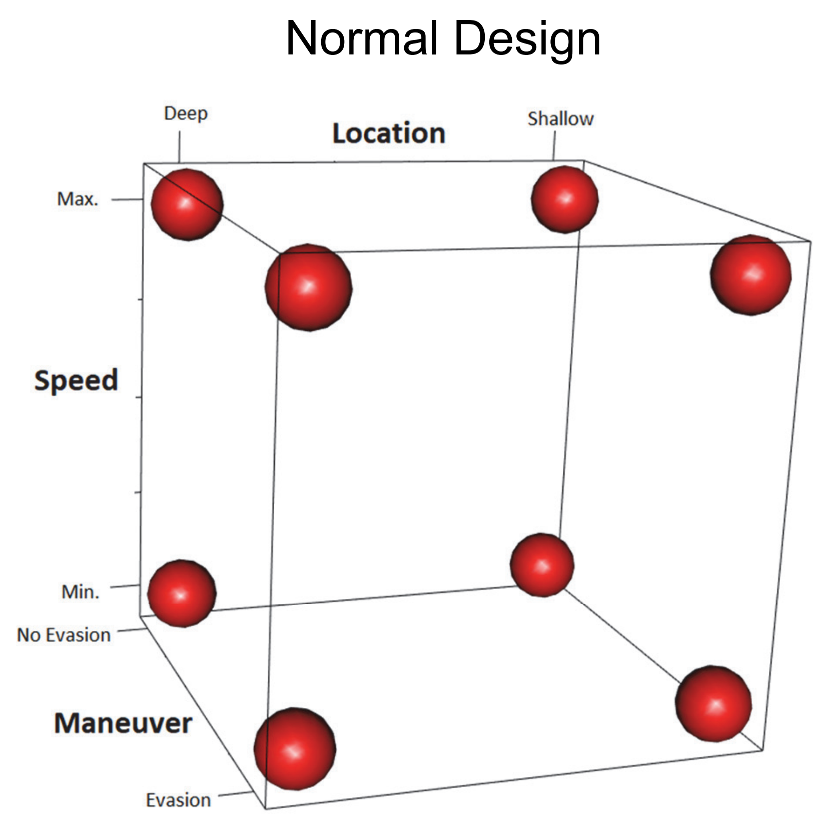
Binary D-optimal design anticipates binary data, whereas the normal D-optimal design does not



	Normal Design	Binary Design
Response Model	$y_i = x_i^T \boldsymbol{\beta} + \epsilon_i$ where $\epsilon_i \sim N(0, \sigma^2)$	$y_i \sim Bernoulli(\pi_i)$ where $\pi_i = \frac{\exp(x_i^T \boldsymbol{\beta})}{1+\exp(x_i^T \boldsymbol{\beta})}$
D-Criterion	Maximize the determinant of the information matrix	Maximize the determinant of the information matrix
Information Matrix	$X^T X$	$X^T V_B X$ where V_B depends on $\boldsymbol{\beta}$
Data Collection	Collects binary response variable	Collects binary response variable
Analysis Model	Logistic Regression	Logistic Regression

The binary D-optimal design is unbalanced

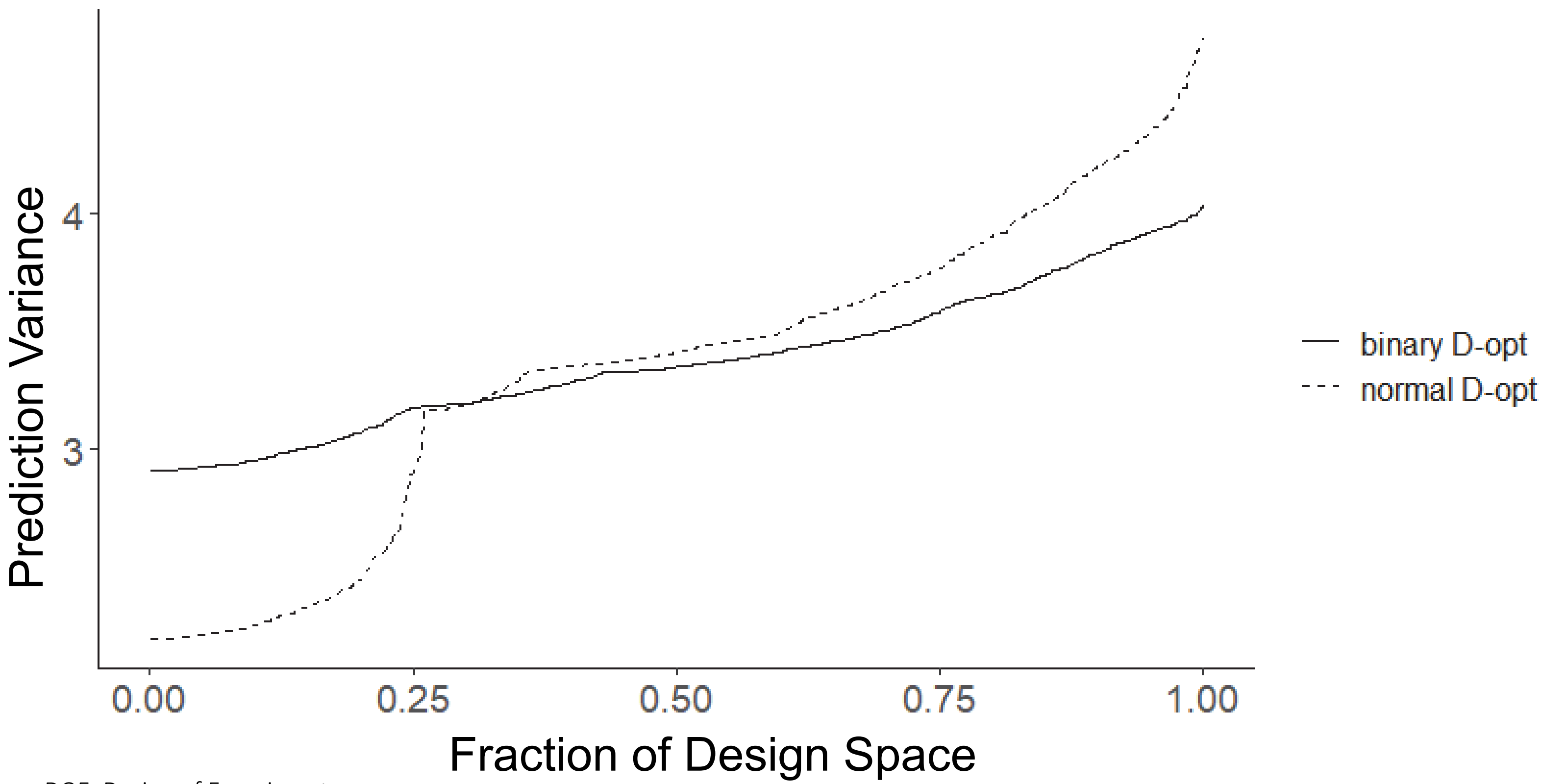




$$\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3) = (-0.5, 1.61, 1.1, 0.205)$$

Standard DOE comparisons favor the binary D-optimal design

- Standard DOE comparisons favor the binary design including:
  - D-efficiency: binary D-optimal is 1.064x more efficient than normal D-optimal
- Fraction of Design Space (FDS) plot



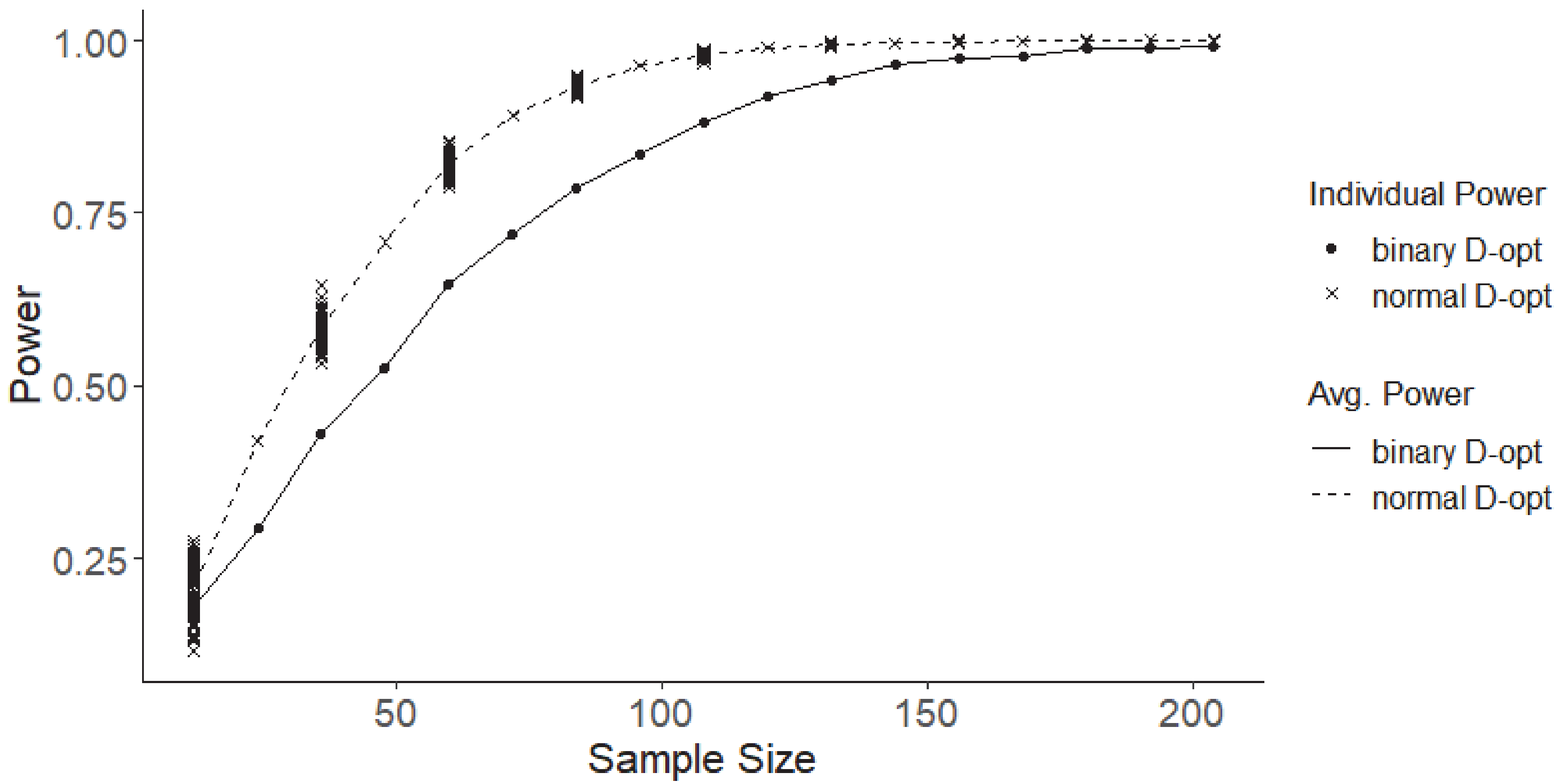
DOE: Design of Experiments

Approximate designs to exact designs

Location	Maneuver	Speed	Normal Design Weight	Binary Design Weight
deep	no evasion	minimum	12.50%	16.42%
deep	no evasion	maximum	12.50%	16.23%
deep	evasion	minimum	12.50%	16.72%
deep	evasion	maximum	12.50%	16.52%
shallow	no evasion	minimum	12.50%	4.40%
shallow	no evasion	maximum	12.50%	0.00%
shallow	evasion	minimum	12.50%	16.12%
shallow	evasion	maximum	12.50%	13.54%

- Approximate designs to exact designs
  - (Pukelsheim, 1993) method
- Statistic used:
  - Likelihood ratio

Binary design underperforms in power analysis



* Using likelihood ratio statistic ($\alpha = 0.05$)

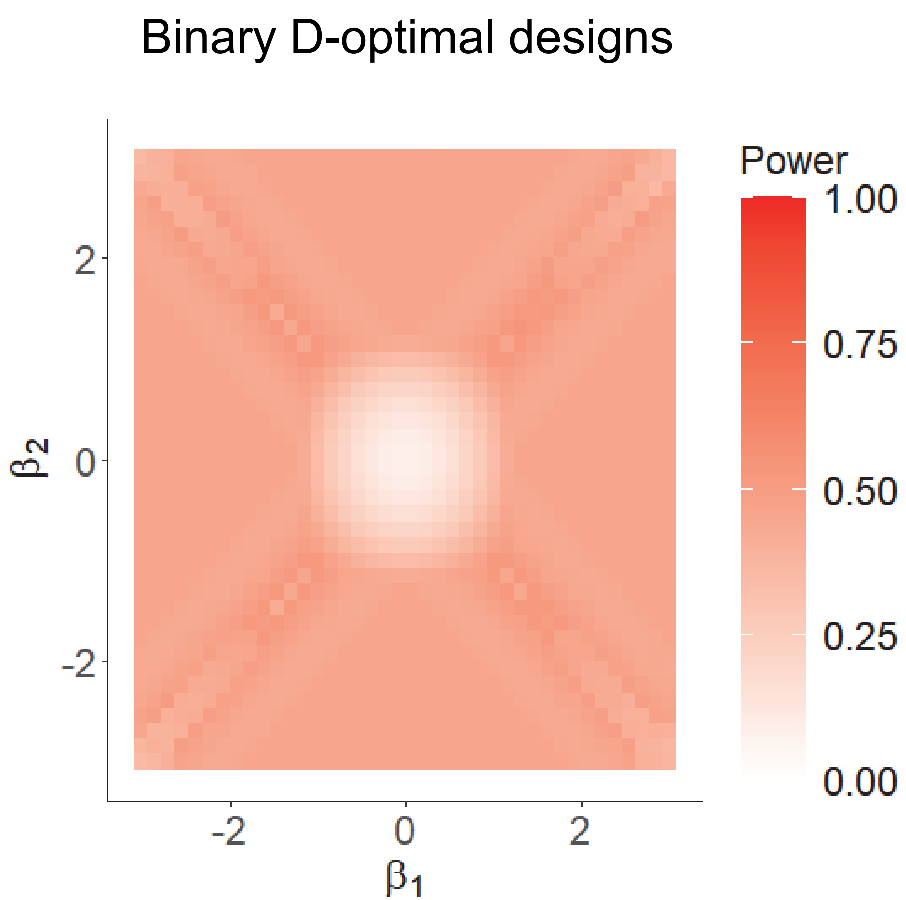
Study power over parameter space

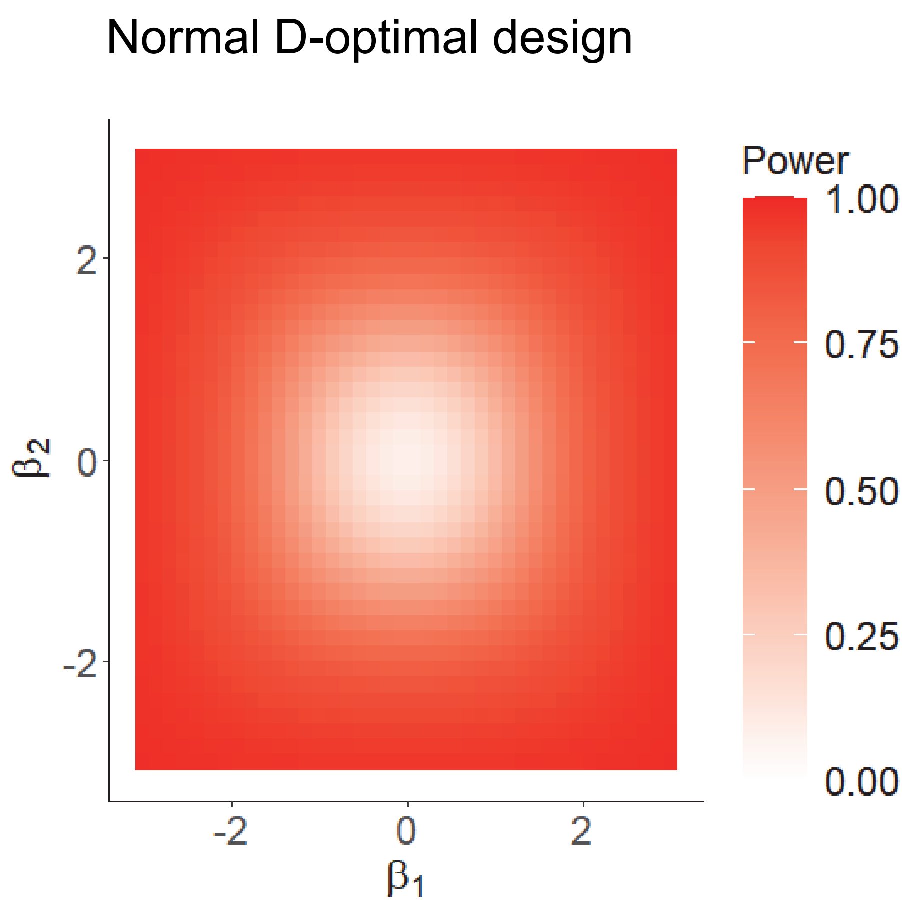
Two continuous factors

- $\eta_i = x_i^T \boldsymbol{\beta}$
- $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)^T$
- $x_i^T = (1, x_1, x_2)$
  - Where x_1 and x_2 are continuous variables in $[-1,1]$

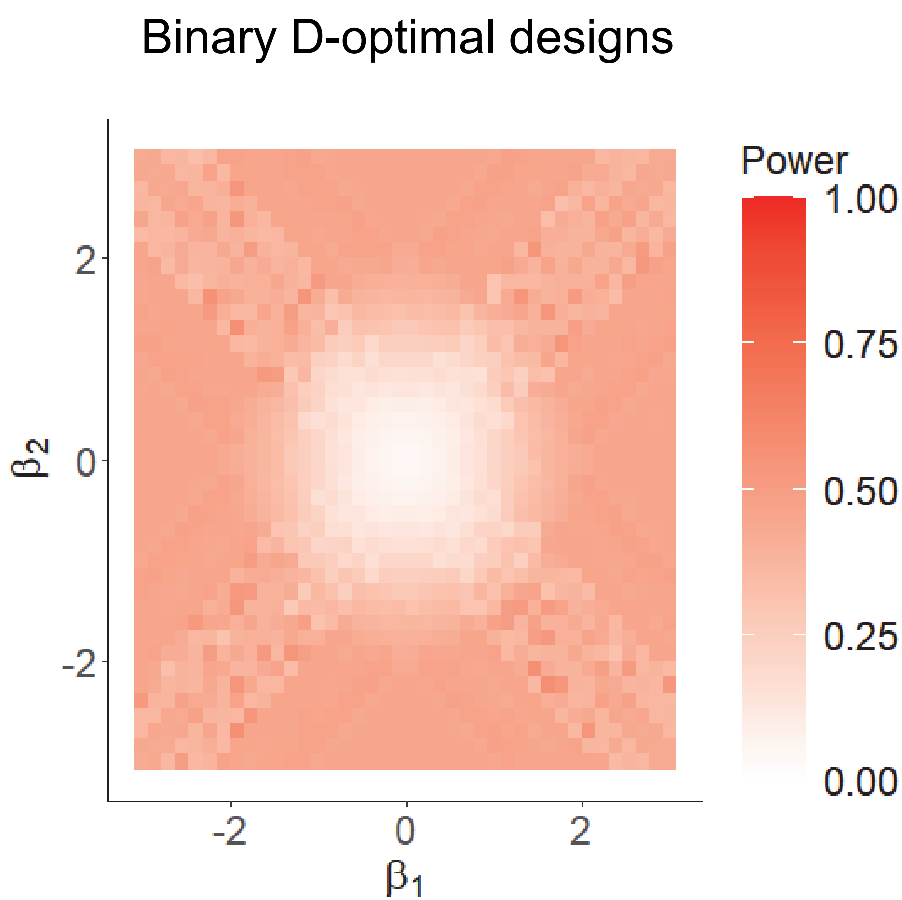
- Fix sample size $n = 12$
- $\boldsymbol{\mathcal{B}} = \{\boldsymbol{\beta}: \beta_0 \in \{0,1,2\}, \beta_1, \beta_2 \in [-3,3]\}$
- For each $\boldsymbol{\beta} \in \boldsymbol{\mathcal{B}}$ over a fine grid,
  - A local binary D-optimal design is found at $\boldsymbol{\beta}$
  - We calculate power for the local binary D-optimal and for the normal design at $\boldsymbol{\beta}$

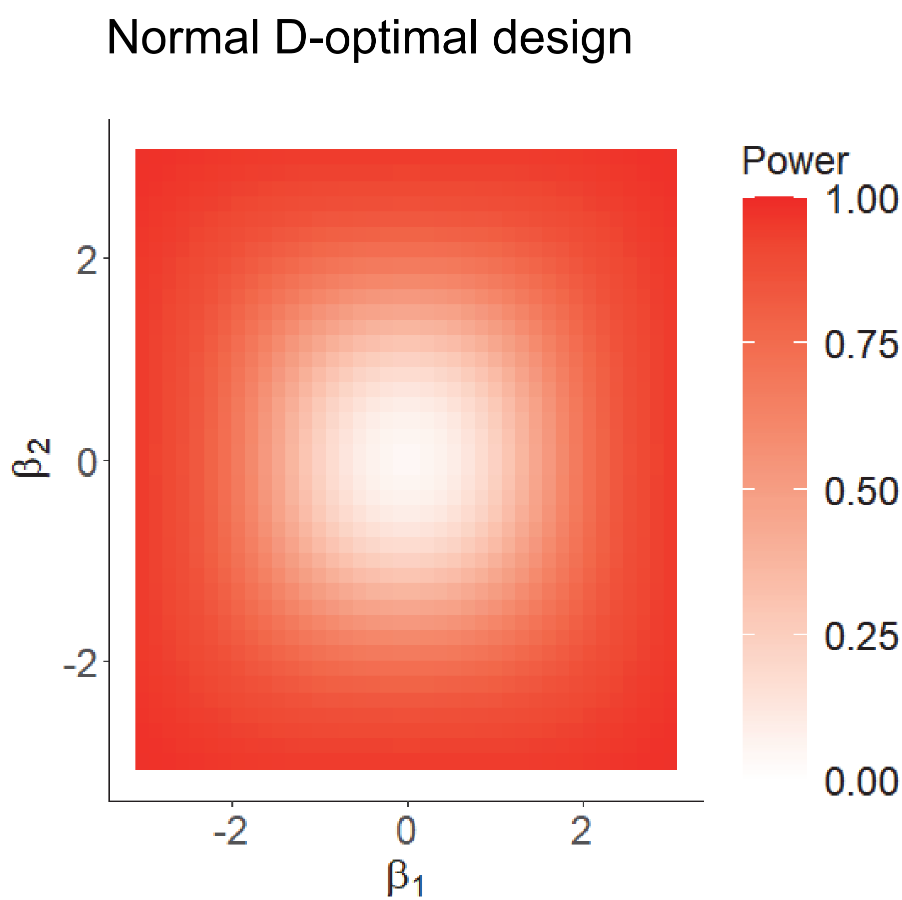
For $\beta_0 = 0$ the normal D-optimal design results in higher power everywhere



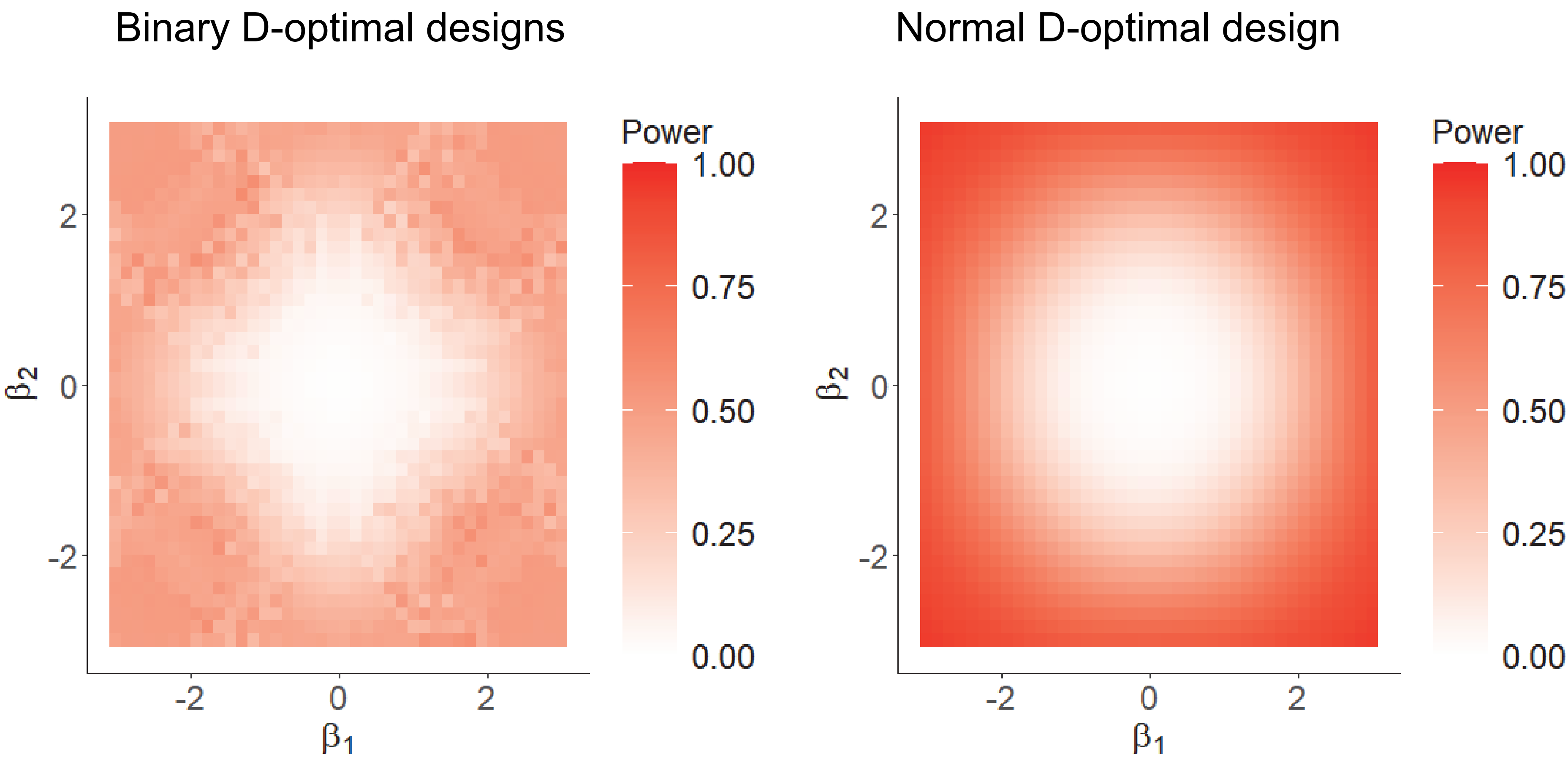


For $\beta_0 = 1$ the normal D-optimal design results in higher power everywhere





For $\beta_0 = 2$ the normal D-optimal design results in higher power nearly everywhere



There is a small, low-power area where local binary D-optimal designs result in slightly higher power than the normal D-optimal design

Where the binary D-optimal design outperforms the normal D-optimal design

Three problems where the binary D-optimal design outperforms the normal D-optimal design:

1. The binary D-optimal design outperforms the normal D-optimal design in low-power regions
2. The binary D-optimal design is constructed using the unknown parameter values
3. As sample size increases, the advantage of the binary D-optimal design dissipates

Fix $\boldsymbol{\beta}$ and increase sample size n

Largest increase for binary D-opt

- $\boldsymbol{\beta}' = (2, -1.31, 1.31)$
- The binary D-optimal design at $\boldsymbol{\beta}'$ results in power ≈ 0.39 at $\boldsymbol{\beta}'$
- The normal D-optimal design results in power ≈ 0.29 at $\boldsymbol{\beta}'$
- Take the approx. designs at $\boldsymbol{\beta}'$ and increase sample size

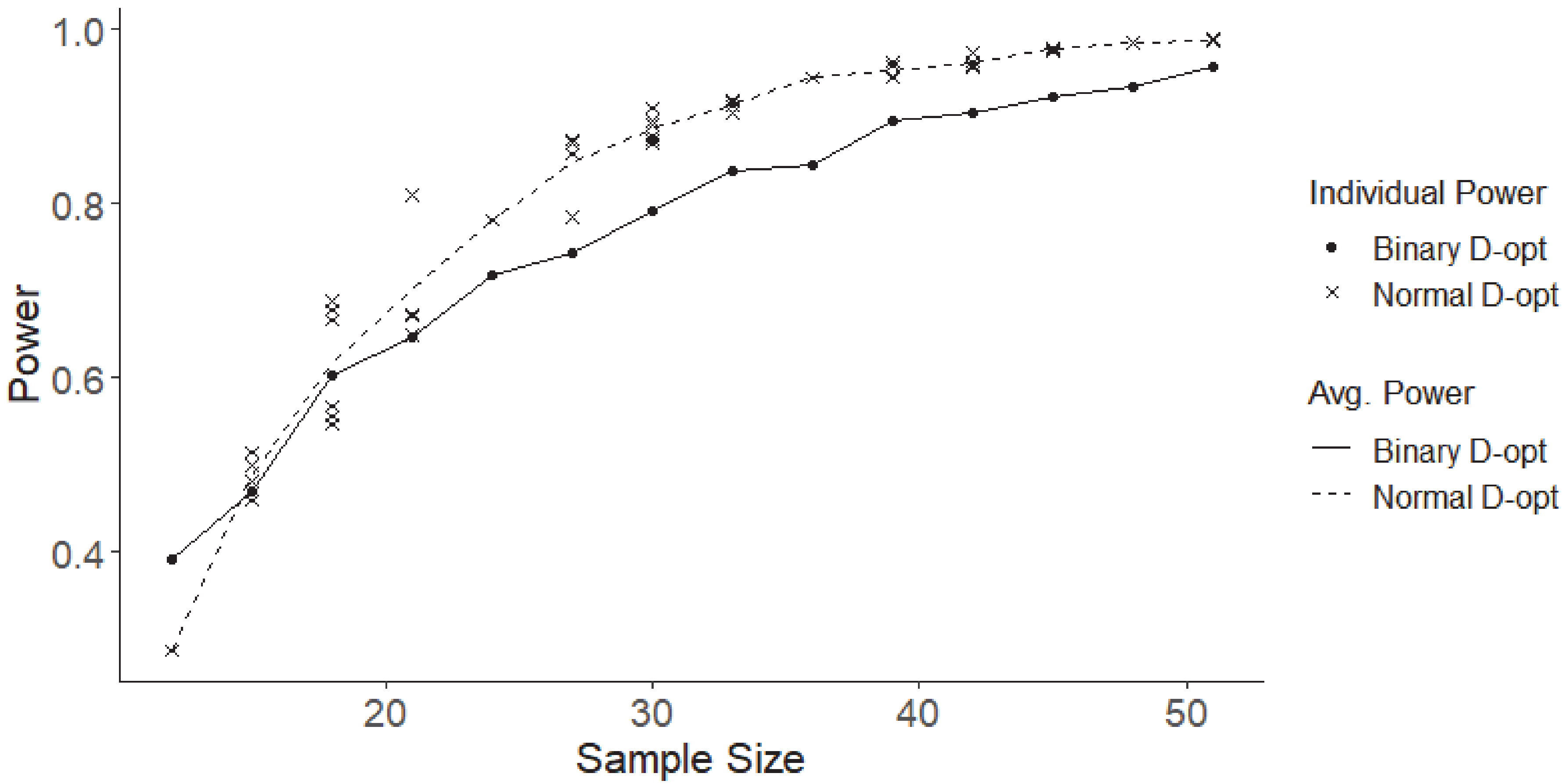
Binary D-optimal design at $\boldsymbol{\beta}'$

Factor 1	Factor 2	Design Weights
1	0.75	1/3
0.64	-1	1/6
1	-0.64	1/6
-0.75	-1	1/3

Normal D-optimal design

Factor 1	Factor 2	Design Weights
1	1	1/4
1	-1	1/4
-1	1	1/4
-1	-1	1/4

The binary D-optimal power advantage disappears with an increasing sample size



Primary Findings and Discussion

- Normal D-optimal designs generally result in higher power than the binary D-optimal designs
- Standard DOE comparisons favor the binary D-optimal design
- Similar findings with interactions and quadratic terms

* Performance is measured by power calculations.

DOE: Design of Experiments

References

- Box, G. E.; Hunter, W. H.; & Hunter, S. (1978). *Statistics for experimenters* (Vol. 664). New York: John Wiley and sons.
- Burke, S. E.; Montgomery, D. C.; Anderson-Cook, C. M.; Silvestrini, R. T.; & Borror, C. M. (2021). Optimal designs for dual response systems for the normal and binomial case. *Quality and Reliability Engineering International*, *37*(7), 3034-3054.
- Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika*, *80*(1), 27-38.
- Hothorn, T., Zeileis, A.; Farebrother, R. W.; Cummins, C.; Millo, G.; Mitchell, D.; & Zeileis, M. A. (2015). Package 'lmtest'. *Testing linear regression models. https://cran. r-project. org/web/packages/lmtest/lmtest. pdf. Accessed*, *6*.
- Johnson, R. T., & Montgomery, D. C. (2009). Choice of second-order response surface designs for logistic and Poisson regression models. *International Journal of Experimental Design and Process Optimisation*, *1*(1), 2-23.
- McCullagh, P., & Nelder, J. A. (2019). *Generalized linear models*. Routledge.
- Morgan-Wall, T., & Khoury, G. (2021). Optimal Design Generation and Power Evaluation in R: The skpr Package. *Journal of Statistical Software*, *99*, 1-36.
- Ozol-Godfrey, A.; Anderson-Cook, C.; & Robinson, T. J. (2008). Fraction of design space plots for generalized linear models. *Journal of Statistical Planning and Inference*, *138*(1), 203-219.
- Pukelsheim, F. (1993). *Optimal Design of Experiments.* New York: John Wiley & Sons.
- Russell, K. G. (2018). *Design of experiments for generalized linear models*. Chapman and Hall/CRC.
- Zahran, A.; Anderson-Cook, C. M.; & Myers, R. H. (2003). Fraction of design space to assess prediction capability of response surface designs. *Journal of Quality Technology*, *35*(4), 377-386.

How Power Analysis was conducted

- Power analysis is conducted via Monte Carlo simulation using the anticipated parameters
- Null Hypothesis: All non-intercept parameters are zero
  - Alternative Hypothesis: At least one non-intercept parameter is non-zero
- Test Statistic is Likelihood Ratio (R package lmtest)
- Parameter estimation is done using the firth correction
  - R package mbest to modify the glm object

REPORT DOCUMENTATION PAGE	Form Approved OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
XX-07-2023	Final	

4. TITLE AND SUBTITLE
JSM 2023: Comparing Normal and Binary D-Optimal Design of Experiments by Statistical Power

5a. CONTRACT NUMBER
HQ0034-19-D-0001

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Adams, Addison, D

5d. PROJECT NUMBER

5e. TASK NUMBER
C9082

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Institute for Defense Analyses
730 East Glebe Road
Alexandria, Virginia 22305

8. PERFORMING ORGANIZATION REPORT NUMBER
NS 3000032

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Central Research Project

10. SPONSOR/MONITOR'S ACRONYM(S)
CRP

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
This OED Draft has not been approved by the sponsor for distribution and release. Reproduction or use of this material is not authorized without prior permission from the responsible IDA Division Director.

13. SUPPLEMENTARY NOTES
Project Leader: Medlin, Rebecca M.

14. ABSTRACT
In many Department of Defense (DoD) Test and Evaluation (T&E) applications, binary response variables are unavoidable. Many have considered D-optimal design of experiments (DOEs) for generalized linear models (GLMs). However, little consideration has been given to assessing how these new designs perform in terms of statistical power for a given hypothesis test. Monte Carlo simulations and exact power calculations suggest that D-optimal designs generally yield higher power than binary D-optimal designs, despite using logistic regression in the analysis after the data have been collected. Results from using statistical power to compare designs contradict traditional DOE comparisons which employ D-efficiency ratios and fractional design space (FDS) plots. Power calculations suggest that practitioners that are primarily interested in the resulting statistical power of a design should use normal D-optimal designs over binary D-optimal designs when logistic regression is to be used in the data analysis after data collection

15. SUBJECT TERMS
Design of Experiment; Operational Testing; Statistics

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON	19b. TELEPHONE NUMBER *(Include area code)*
Unclassified	Unclassified	Unclassified	SAR	TBD	Rebecca M. Medlin	703-845-6731