INSTITUTE FOR DEFENSE ANALYSES

Design of Experiments for Generalized Linear Models

Rebecca Medlin, Project Leader

Curtis Miller
Addison Adams

August 2022

Approved for Public Release.
Distribution Unlimited.

IDA Document D-33208 NS

Log: H 2022-000348

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

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, Cross-Divisional 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.

Acknowledgments
The IDA Technical Review Committee was chaired by Dr. V. Bram Lillard and consisted of Dr. Jason P. Sheldon, Dr. Keyla Pagan-Rivera, Dr. Mark R. Herrera, and Dr. John T. Haman from the Operational Evaluation Division.

For more information:
Dr. Rebecca 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 D-33208 NS

Design of Experiments for Generalized Linear Models

Rebecca Medlin, Project Leader

Curtis Miller
Addison Adams

This page intentionally left blank.

Executive Summary

Optimal design of experiments (DOE) promises to allocate a fixed number of test runs so that some statistical property of the data is optimized. In the case of D-optimal designs, the optimal design should, in some sense, minimize the standard errors of the parameter estimates of the chosen statistical model, which should lead to better inference of those parameters' value and, thus, the contribution of the associated factors to the response variable. Analysts at IDA and in the operation test community generate D-optimal DOEs that are optimal when using a response variable drawn from a continuum (more specifically, a Normal distribution). These designs may not be optimal under other conditions. Specifically, this presentation investigates how one should generate DOEs when the response variable consists of a binary response, either success or failure, and hence implies using logistic regression for analysis, a particular case of a larger class of statistical models known as generalized linear models (GLMs).

IDA studied optimal DOE for GLMs and developed an R **shiny** app that generates D-optimal DOEs for logistic regression as part of a Summer Associate project. IDA conducted simulation studies comparing the statistical power of hypothesis tests for model utility when one uses the D-optimal DOE for the logistic regression model versus the standard practice of using the D-optimal DOE for a Normal response, which is not optimal for logistic regression models.

Despite the fact that the DOE accounting for logistic regression is theoretically optimal—and thus superior—to the design for the wrong model, the D-optimal design had markedly worse power in the sample sizes that planners would expect to use in operational test events relative to the improperly specified DOEs currently used. Some conjectures to explain this phenomenon are presented but, given the surprising nature of this result, more research is needed to explain it.

Presently, we do not recommend any changes to how people construct DOEs. However, we may adopt different recommendations based on follow-on research.

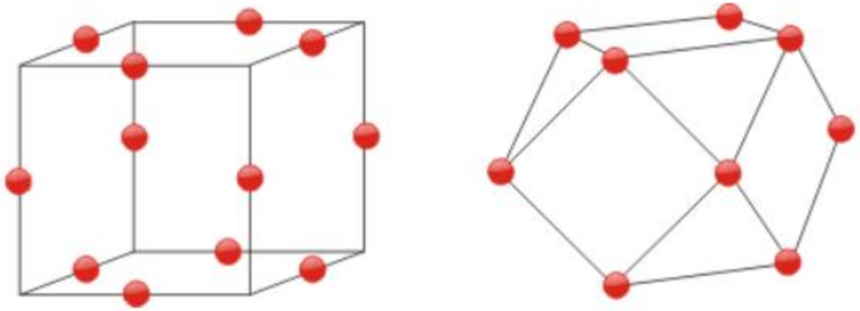
This page intentionally left blank.

Design of Experiments for Generalized Linear Models

Addison Adams

Mentors: Dr. Curtis Miller and Dr. Rebecca Medlin

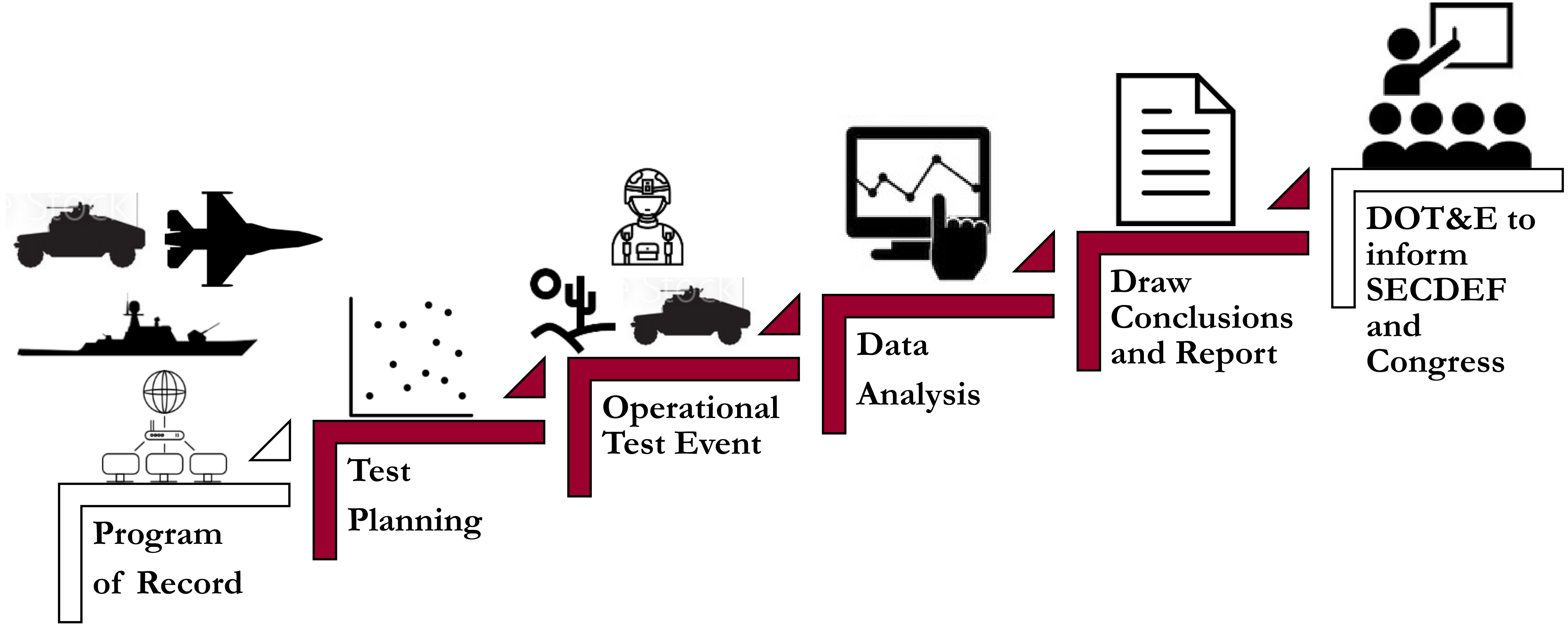
Summer 2022



Institute for Defense Analyses

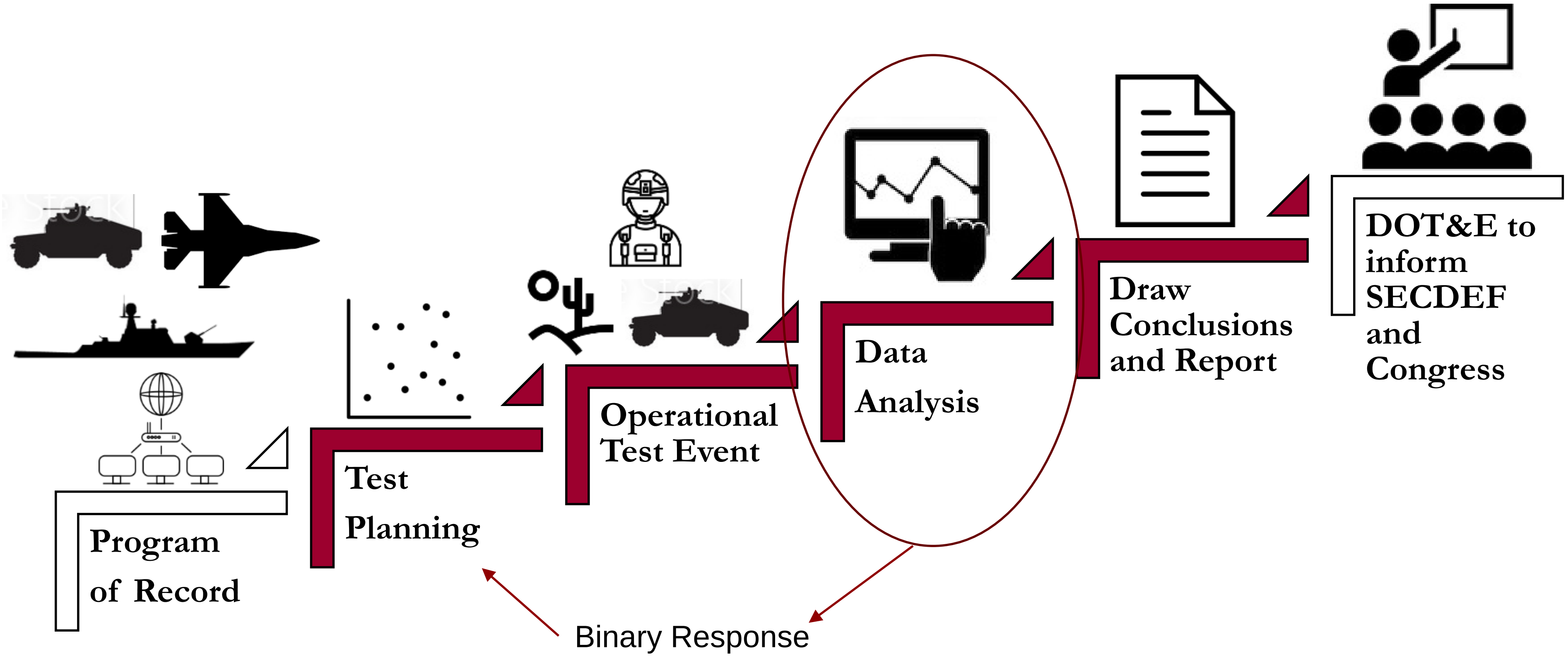
730 East Glebe Road ● Alexandria, Virginia 22305

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



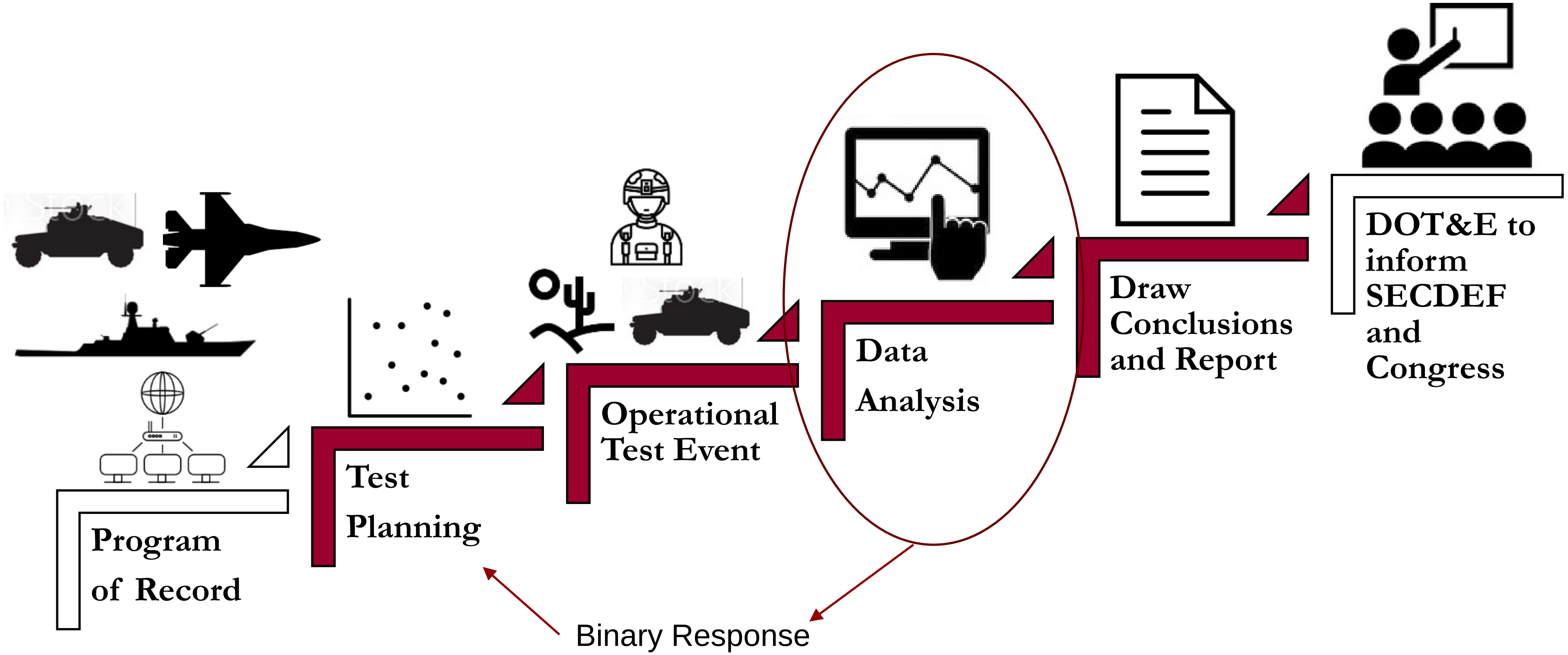
Acronyms: DOT&E – Director, Operational Test and Evaluation; SECDEF – Secretary of Defense

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



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

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



Key Findings:

- Results show the new design is better theoretically, but practically the current design outperforms the new design
- Relatively low performance of both designs, more research may be required
- More types of responses can be considered based on the framework built

Acronym: DOT&E – Director, Operational Test and Evaluation

Design of experiment for torpedo hit probabilities

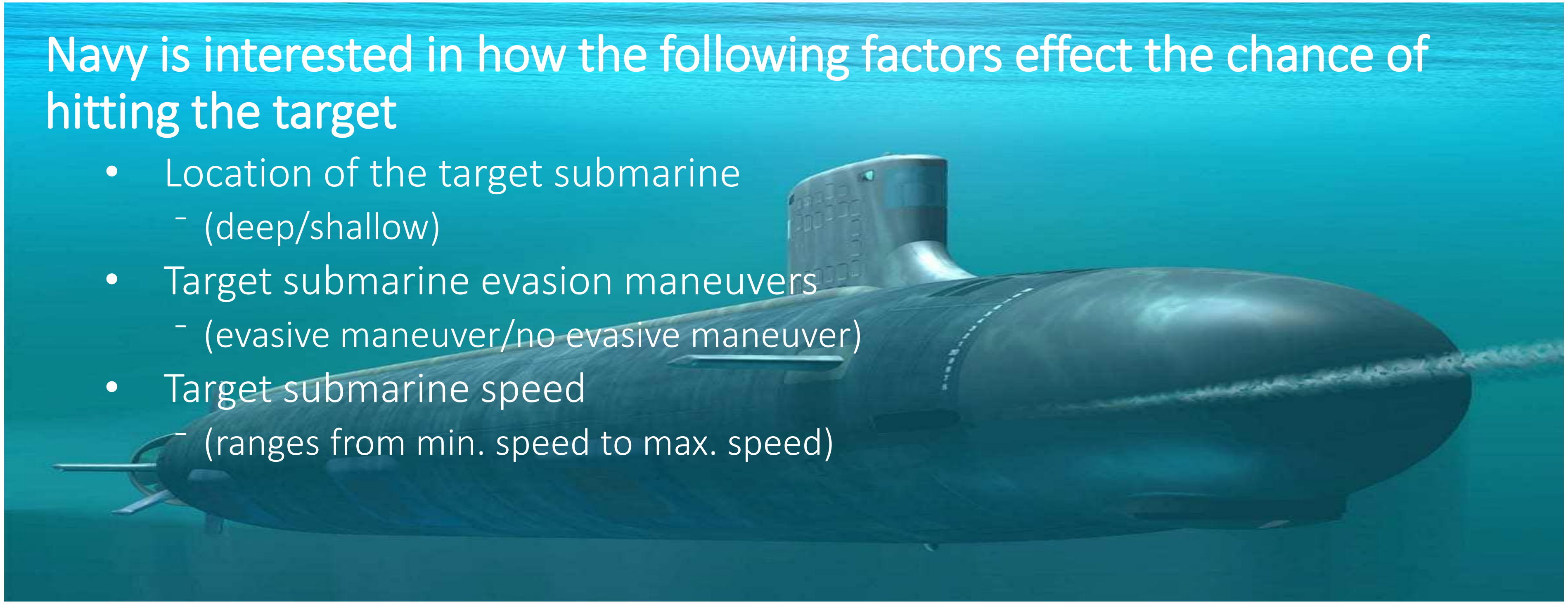
A **design of experiment** (DOE) is the planning of experiment with the statistical analysis in mind

- A grape analogy
- DOE: Planning the methods of how to grow grapes

Design of experiment for torpedo hit probabilities

Factors: Conditions under which an experiment is conducted

- May be continuous or categorical



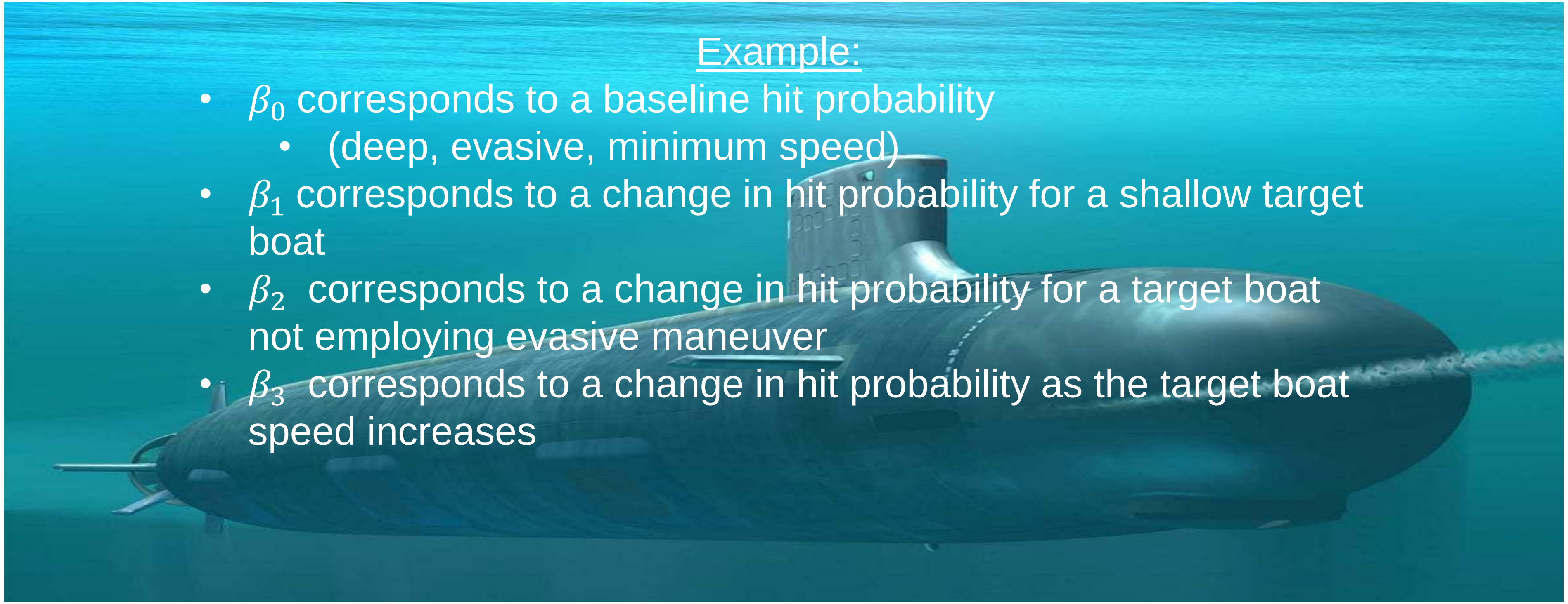
Navy is interested in how the following factors effect the chance of hitting the target

- 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)

Design of experiment for torpedo hit probabilities

Parameters: The unknown but true effects of the factors on the response

- We will denote the 4 parameters in our submarine example by β_0, β_1, β_2, and β_3



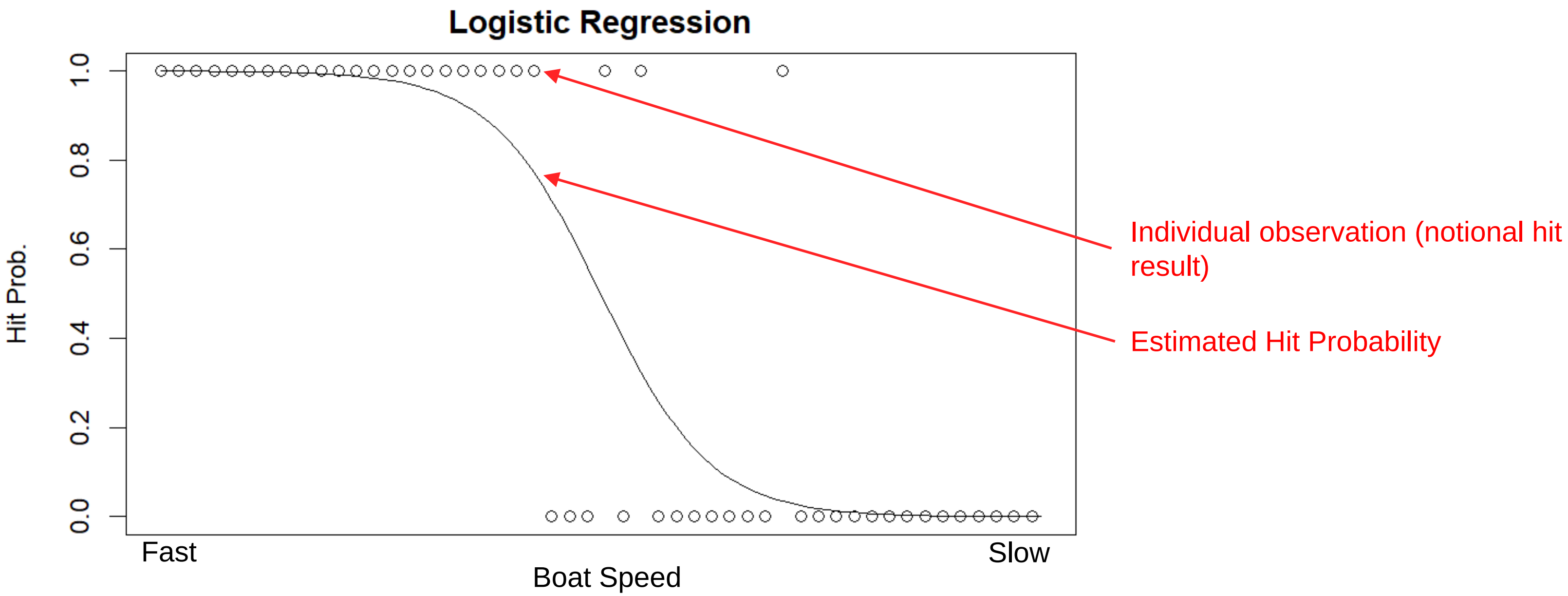
<u>Example:</u>

- 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

Design Methods

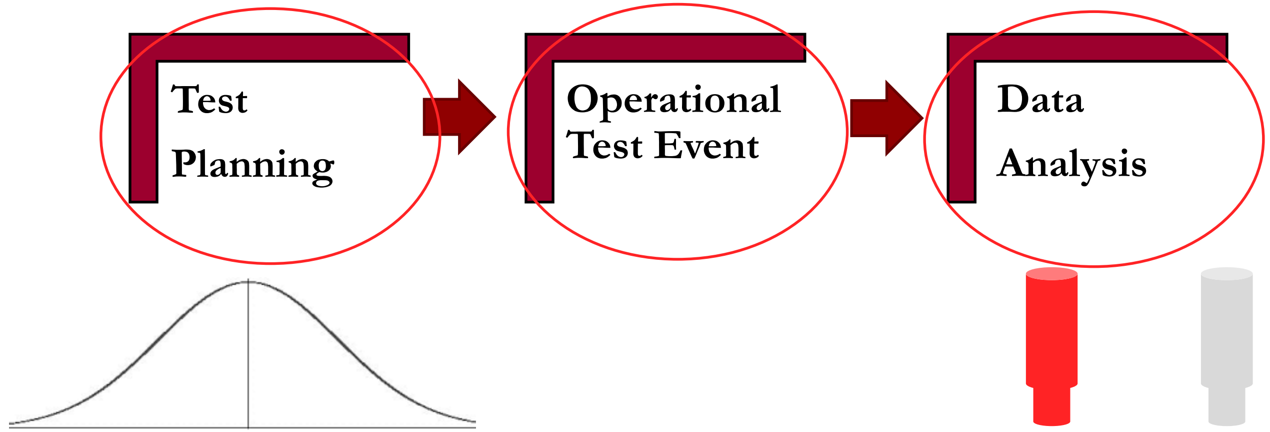
Logistic regression model is used in data analysis for both designs

- Logistic regression:
  - Common model used for evaluating binary responses
  - Special case of generalized linear models



Acronym: DOE – Design of Experiment

Logistic design anticipates binary data in data analysis, whereas the normal design does not



Normal design (current)

- Anticipates normal data in data analysis
- Optimizes precision of parameter estimates using normal theory
- Collects binary response in operational test event
- Uses logistic regression in data analysis

Logistic design (new)

- Anticipates binary data in data analysis
- Optimizes precision of parameter estimates using logistic theory
- Collects binary response in operational test event
- Uses logistic regression in data analysis

Acronym: DOE – Design of Experiment

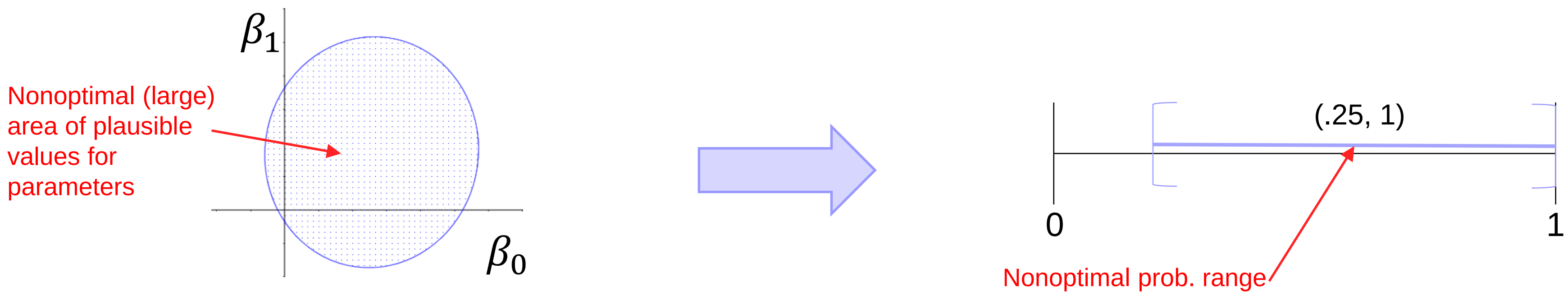
Optimizing the precision of the parameter estimates

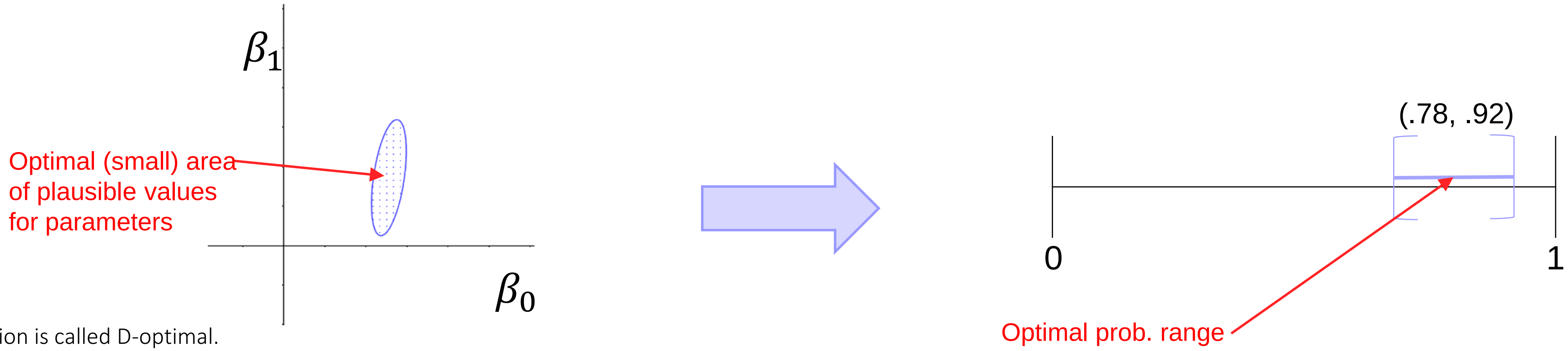
Both designs seek to maximize precision of parameter estimates according to their respective theories

- Recall β_0 and β_1 are fixed, unknown values that quantify the hit probability
  - β_0 relates to baseline hit probability
  - β_1 relates to the change in hit probability for the location (shallow/deep) of target boat

<u>Plausible values for parameters</u>

<u>Notional: Plausible range of hit probs. against a shallow target</u>

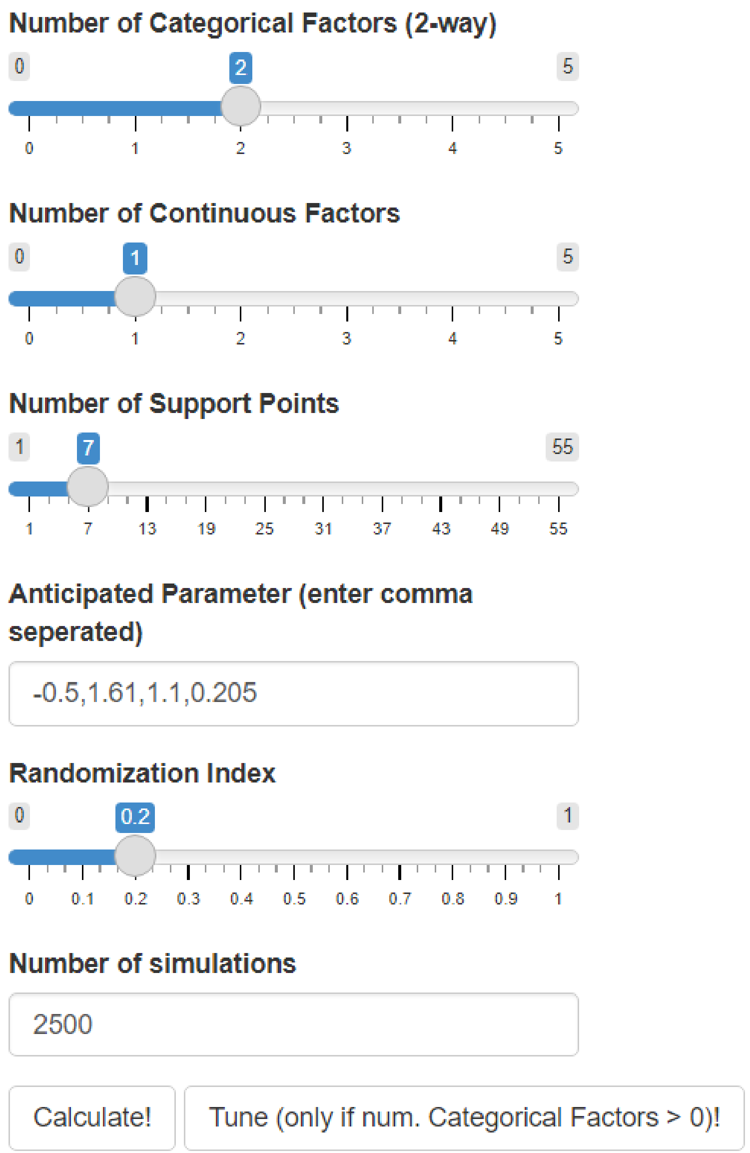




*This type of optimization is called D-optimal.

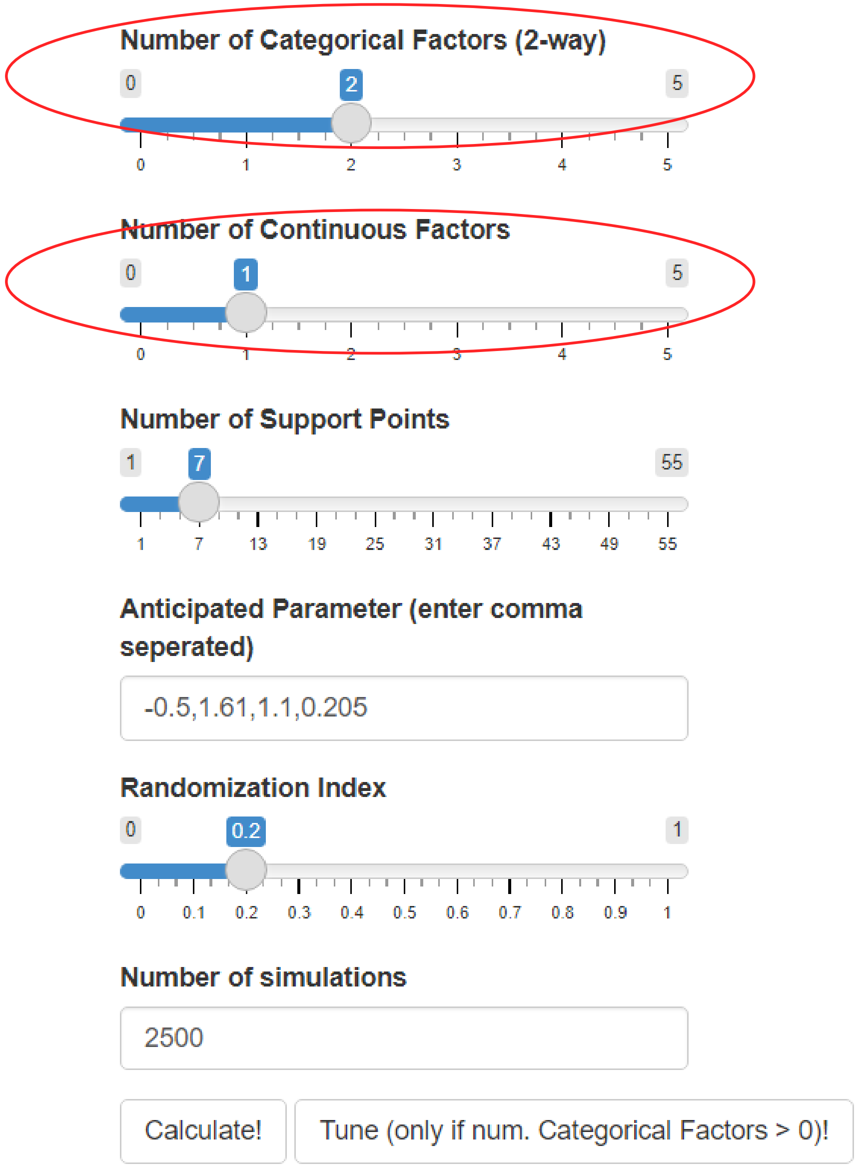
Finding the optimal logistic design

Application generates optimal logistic designs



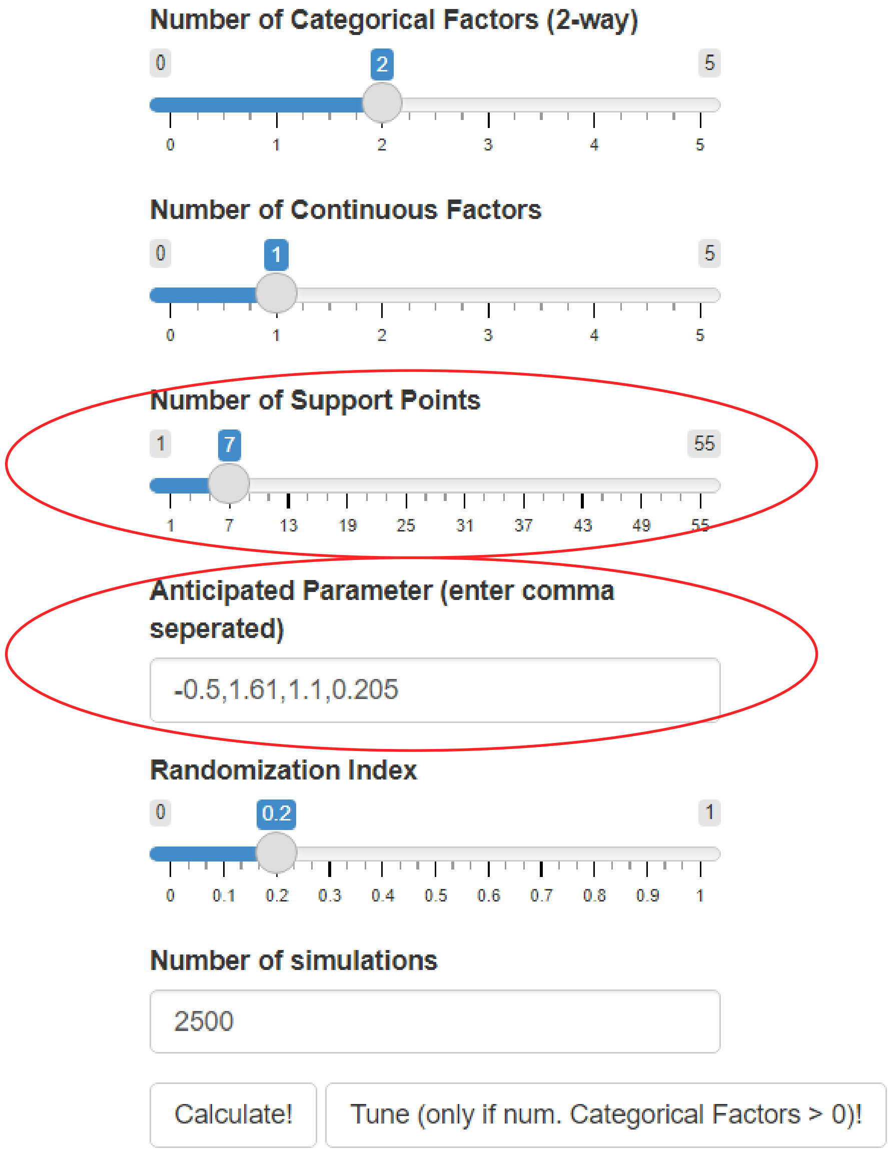
Finding the optimal logistic design

- Select the number of categorical and continuous factors



Finding the optimal logistic design

- Support points are the unique factor configurations
- Anticipated Parameters: What one expects to be the truth
  - Empirical Issue



Finding the optimal logistic design

Output of the app

Design

Optimality criterion

Optimality check

```
                    deswts
[1,] 0 1 -0.99998 0.16417
[2,] 0 0  0.99999 0.16580
[3,] 1 0 -0.99997 0.16115
[4,] 1 1 -0.99881 0.04400
[5,] 0 1  1.00000 0.16228
[6,] 1 0  0.99983 0.13540
[7,] 0 0 -1.00000 0.16720
```

```
[1] "Information Matrix Determinant: 7.9647143135043e-05"
```

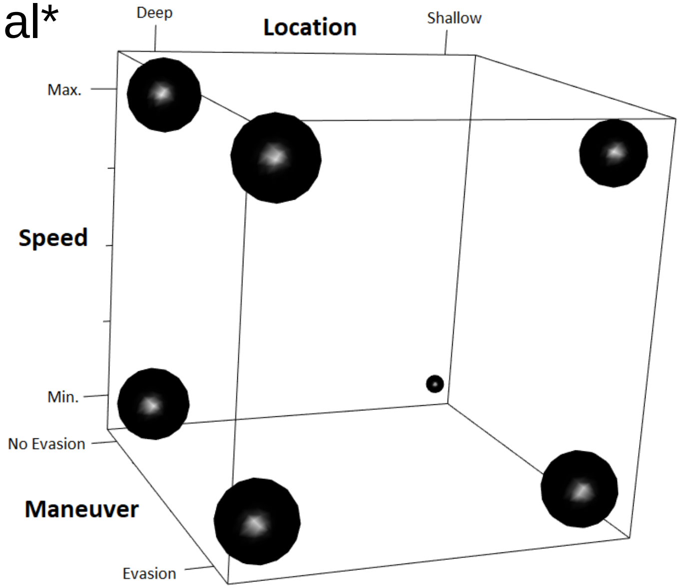
```
For support point  1
 Maximum std var is  4.016498  at
 0 1 -1
For support point  2
 Maximum std var is  3.999306  at
 0 0 1
For support point  3
 Maximum std var is  4.011516  at
 1 0 -1
For support point  4
 Maximum std var is  4.024509  at
 1 1 -1
For support point  5
 Maximum std var is  3.98413  at
 0 1 1
For support point  6
 Maximum std var is  4.025801  at
 1 0 1
For support point  7
 Maximum std var is  3.962165  at
 0 0 -1
```

The optimal logistic design for submarine example

Optimal Logistic Design

Location	Maneuver	Speed	Design Weight	Replication (18)	Replication (30)
deep	no evasion	minimum	16.42%	x3	x5
deep	evasion	maximum	16.58%	x3	x5
shallow	evasion	minimum	16.12%	x3	x5
shallow	no evasion	minimum	4.40%	x1	x1
deep	no evasion	maximum	16.23%	x3	x5
shallow	evasion	maximum	13.54%	x2	x4
deep	evasion	minimum	16.72%	x3	x5

- The design is optimal*



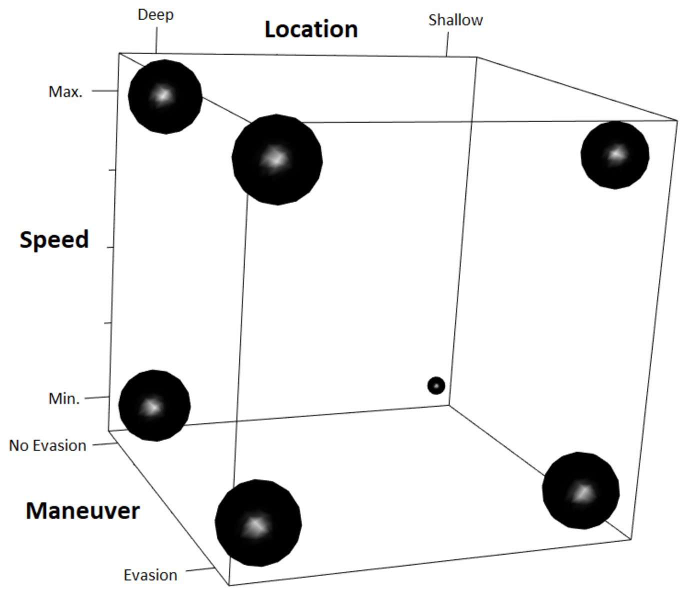
*D-optimal

Results and Next Steps

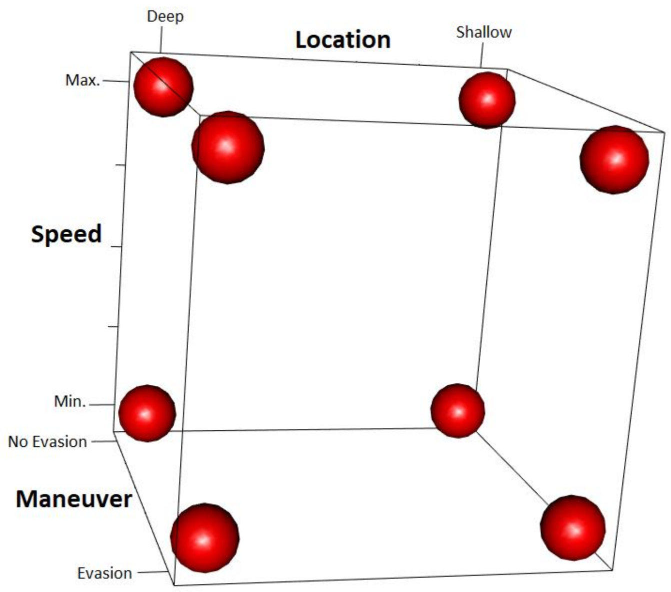
Theoretical Results favor Logistic Design

- Theoretical results show the logistic design is preferable (it was optimal whereas the normal design was not)
  - 1.06x efficiency
- Logistic design generally requires fewer support points than the normal design

Logistic Design



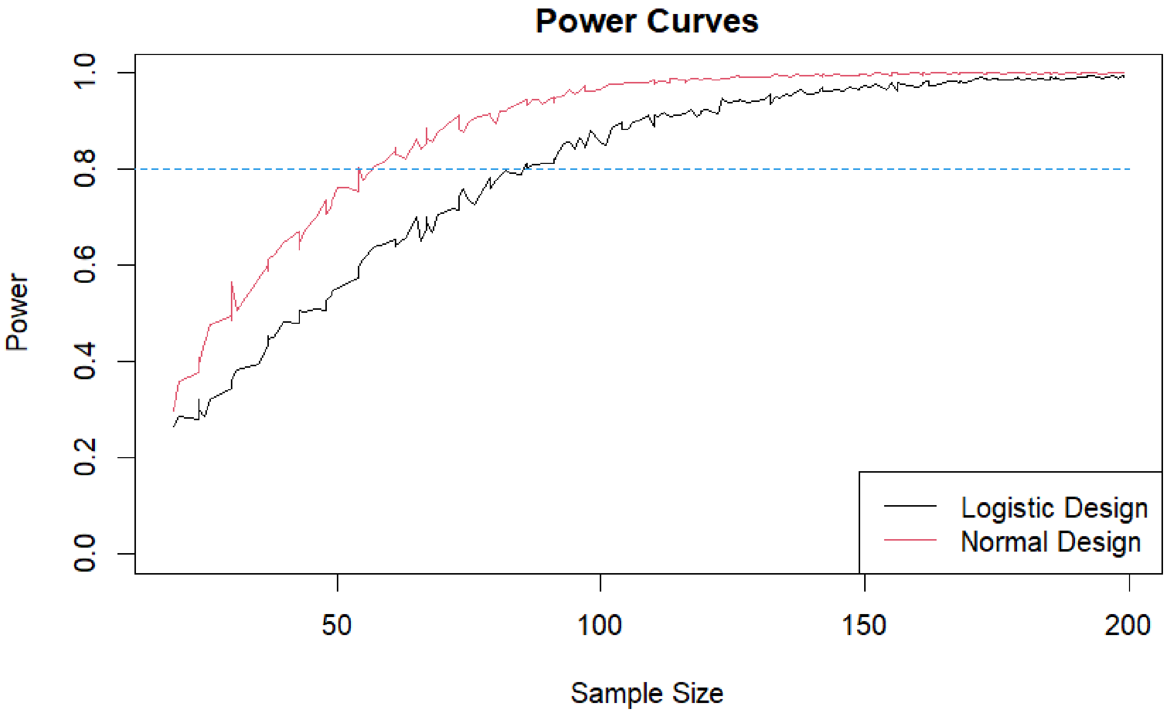
Normal Design



Logistic Design underperforms in power analysis

- Power: The probability of detection for at least one factor*
- Logistic design underperforms normal design in power analysis

- Until further work is done, recommendation is to use the normal design



* Model utility test hypothesis with likelihood ratio statistic

Primary Findings and Next Steps

Primary Findings

- Designs we currently construct (under normal theory) outperform* the investigated designs (using logistic theory)
- Theoretical results favor new approach when we have a binary response
  - This suggests there may be more work to be done
- The design generating tool will be made available to IDA staff

Next Steps

- Consider a new optimality criterion
- Consider more non-normal responses
  - Failure/Reliability data
  - Count data

Thank You!

- Curtis Miller
- Rebecca Medlin
- John Haman
- Jason Sheldon
- Hank Donnelly
- Tyler Morgan-Wall
- Lamarr Colvin
- Heather Neff

- Test Science
- OED

Questions

?

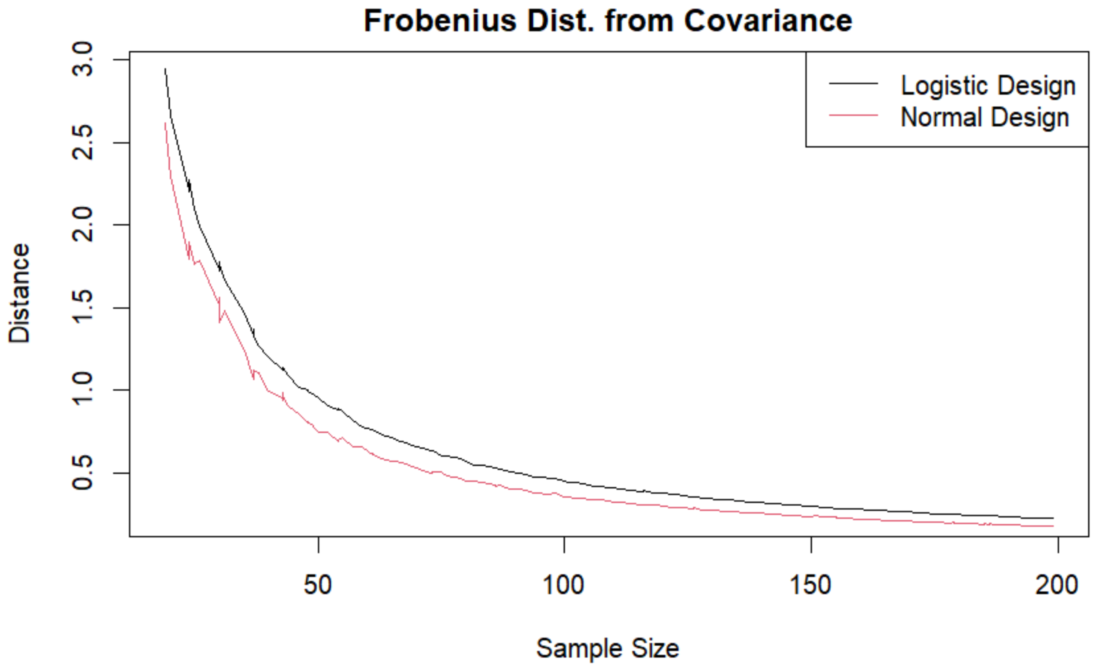
References

- Russell, K. G. (2018). *Design of experiments for generalized linear models*. Chapman and Hall/CRC.
- Box, G. E., Hunter, W. H., and Hunter, S. (1978). *Statistics for experimenters* (Vol. 664). New York: John Wiley and sons.
- Johnson, R. T., and Montgomery, D. C. (2009). Choice of second-order response surface designs for logistic and Poisson regression models. *International Journal of Experimental Design and Process Optimization*, *1*(1), 2-23.
- Morgan-Wall, T., and Khoury, G. (2021). Optimal Design Generation and Power Evaluation in R: The skpr Package. *Journal of Statistical Software*, *99*, 1-36.
- Hothorn, T., Zeileis, A., Farebrother, R. W., Cummins, C., Millo, G., Mitchell, D., and Zeileis, M. A. (2015). Package ‘lmtest.’ *Testing linear regression models. https://cran. r-project. org/web/packages/lmtest/lmtest. pdf. Accessed*, *6*.

Appendix

Why might theoretical results not match with practical

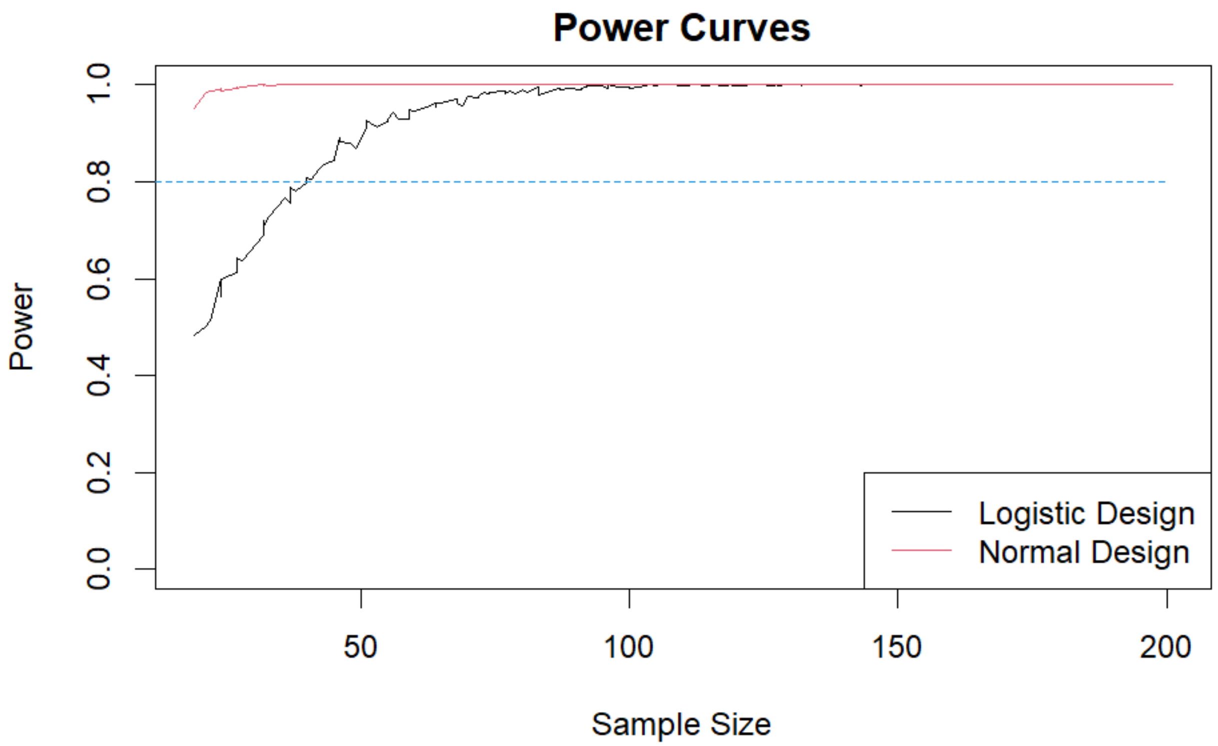
- Although the logistic design is D-optimal and the normal is not, the normal design approaches its theoretical covariance matrix much faster than the logistic design
- We optimize on the theoretical covariance matrix but we might want to consider a different optimization criterion
  - Integrated Mean Square Error (IMSE)



- Frobenius Distances from asymptotic covariance matrix

Categorical and Continuous

- 1 categorical (0,1) and 2 continuous (-1,1)
- Logistic design is 1.48x more efficient
  - Only has 6 support points
- Catches normal design sooner



Ideas

- Gamma response with IMSE optimality
  - Some bias reduction fitting technique (maybe Firth 1993)
- How to take into account the non-asymptotics?
- Literature space does not use practical power studies – only theoretical optimality criteria to determine performance of design.
- Is the BR method effectively reducing bias?
- Theoretical Covariance matrix better approximates the covariance between parameters for logistic design. Does this translate in power curves for more than main effect models, i.e. interactions?

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 nonzero
- Test Statistic is Likelihood Ratio (R package lmtest)
- Parameter estimation is done using the firth correction
  – R package mbest to modify the glm object
- Additionally, Wald and Likelihood Ratios were assessed with ANOVA type III and produced similar results

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
01/08/2022

2. REPORT TYPE
Final

3. DATES COVERED *(From - To)*
N/A

4. TITLE AND SUBTITLE
Design of Experiments for Generalized Linear Models

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

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Curtis Miller
Addison Adams

5d. PROJECT NUMBER
C9082

5e. TASK NUMBER

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
D-33208
H 2022-000348

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Institute for Defense Analyses
730 East Glebe Road
Alexandria, Virginia 22305

10. SPONSOR/MONITOR'S ACRONYM(S)

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

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for Public Release. Distribution Unlimited.

13. SUPPLEMENTARY NOTES
Project Leader: Rebecca Medlin

14. ABSTRACT
This presentation discusses how researchers should design experiments when the outcome of interest consists of either success or failure. This involves exploring designs of experiments when the intended analysis involves estimation of a generalized linear model (in this case, logistic regression). Tools were developed to generate and evaluate these designs. Power studies suggest that D-optimal designs targeting logistic regression do not result in statistical designs with better power than conventional D-optimal designs (which do not assume the use of logistic regression). While these findings are surprising, further research into this topic is needed.

15. SUBJECT TERMS
Design of Experiments; D-Optimal Designss; Logistic Regression; Power Analysis; 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
Unclassified	Uncalssified	Unclassified	Unlimited	35	Dr. Rebecca Medlin
					19b. TELEPHONE NUMBER *(Include area code)* 703-845-6731