

Practical Experimental Design Strategies for Binary Responses under Operational Constraints

Peter Parker and Nathan Cruze

National Aeronautics and Space Administration

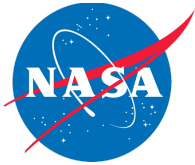
Langley Research Center

Hampton, Virginia

April 17, 2024

Defense and Aerospace Test and Analysis (DATAWorks) 2024

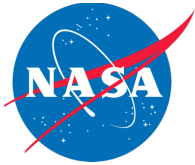
Alexandria, Virginia



Motivation for this Presentation

- **Defense and aerospace testing commonly include binary responses.**
 - **e.g., hit or miss, detect or not detect, success or fail**
- **A statistical model to predict the probability of occurrence of a binary response as a function of an explanatory variable(s) is the objective.**
- **Often, there is more focus on modeling the data rather than opportunities to build an efficient experimental design.**
- **Design constraints imposed by physics, ethics, resources, cost, and system operational capability lead to restrictions on sample size, factor range, and the distribution of factor levels.**
- **We propose a back-to-basics perspective that employs experimental design principles for binary responses that emphasizes collaboration.**

Experiments with Binary Responses

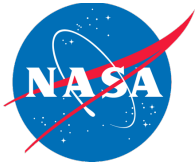


- **Binary response (y) has two possible outcomes, coded as 1 or 0**
 - Special case of categorical responses with multiple outcomes
 - Considering truly binary, not discretized continuous response
- **Designed experiment purposefully controls a factor (x) to deduce its relationships, if any, with the responses.**
 - Intentionally manipulate the factor, not observational data
 - Design specifies factor levels and order of execution
- **Modeling based on generalized linear model (GLM) framework, where logistic regression is the most common. The probability of a 1/0 response is a function of the factor level.**

$$p(x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$$

- **Dose-response experiment, where dose refers to factor level.**

Simple Example of Logistic Regression



- Probability a student passes an exam as a function of hours of study.
- Data table for a student: study hours and pass (1) or fail (0).

Hours (x_k)	0.50	0.75	1.00	1.25	1.50	1.75	1.75	2.00	2.25	2.50	2.75	3.00	3.25	3.50	4.00	4.25	4.50	4.75	5.00	5.50
Pass (y_k)	0	0	0	0	0	0	1	0	1	0	1	0	1	0	1	1	1	1	1	1

$$p(x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$$

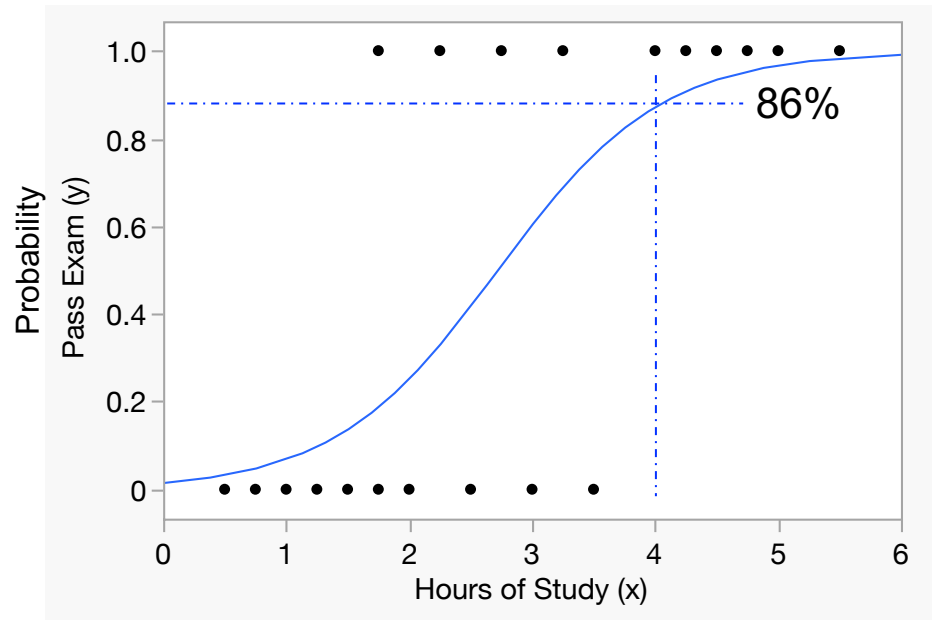
$$\beta_0 \approx -4.1$$

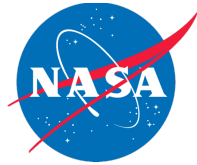
$$\beta_1 \approx 1.5$$

Applying the estimated model

Probability of passing the exam with 4
hours of study

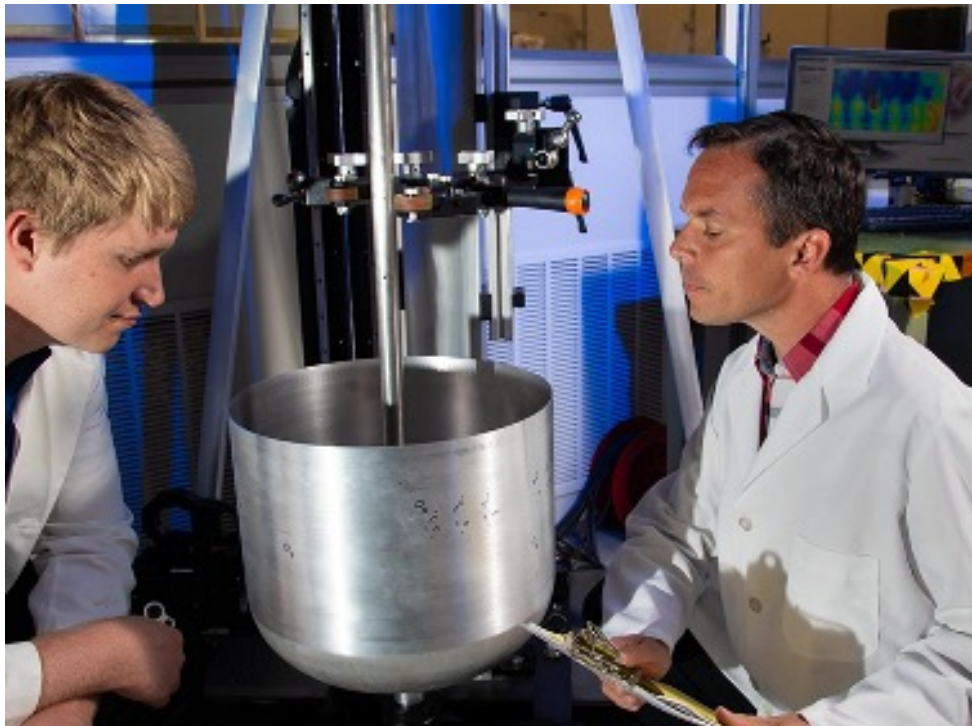
$$p(4 \text{ hours}) = 0.86 \Rightarrow 86\%$$



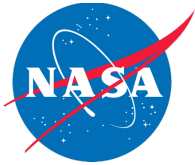


Probability of Detection

- **Model the probability of detection (POD) of a nondestructive evaluation technique as a function of flaw size for fracture-critical human spaceflight components.**
- **Experiment is conducted by presenting flawed and unflawed specimens to an inspector, and they report their indications (hit/miss).**



**Interested in predicting a
flaw size that provides:
90% POD with 95%
confidence**



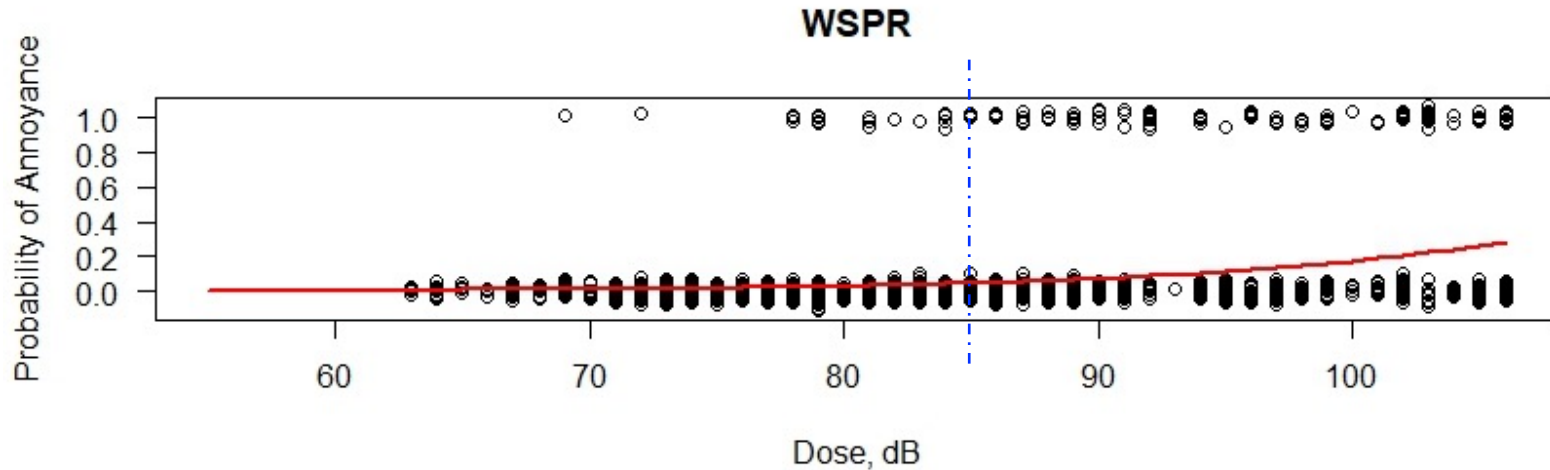
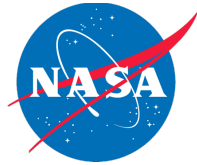
Probability of Annoyance

- **Model the probability of human annoyance (POA) due to noise levels induced by a supersonic aircraft to inform regulators and designers.**
- **Experiment is conducted by flying an experimental aircraft over communities, over a range of noise levels, and surveying participants for their level of annoyance, highly-annoyed or not.**

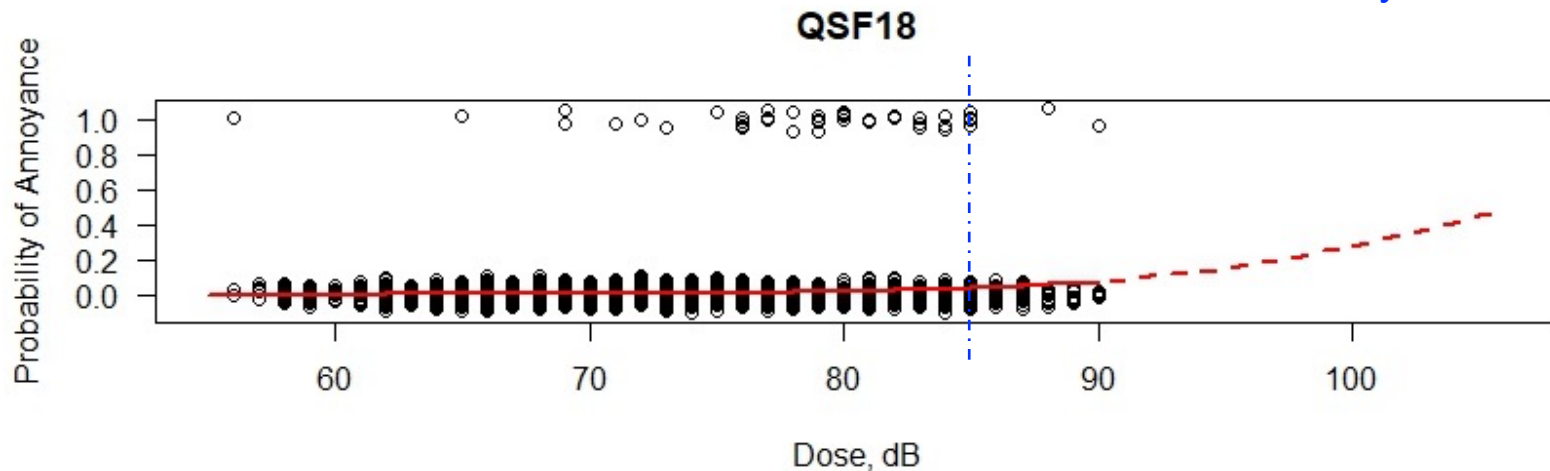


**May be interested in predicting
the noise level size that provides:
5% POA with 95% confidence**

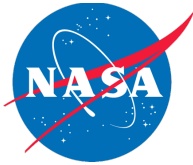
Noise-Annoyance Experimental Data



Interested in predicting
annoyance ~85 dB

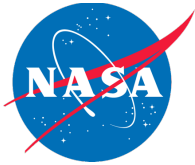


Noise-Annoyance Design Constraints

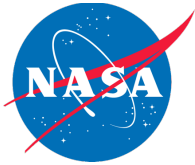


- **Each noise level requires a flight of the experimental aircraft. Total number of design points (noise levels) is limited by flights per day and flights per aircraft's operational deployment in a community.**
 - **Small sample size ~80 flights over 5 weeks**
- **Administering too high of a noise level can have negative human consequences, painful and traumatic (ethical concerns)**
 - **Noise levels are constrained to be less than ~15% highly-annoyed**
- **Range of noise levels is constrained on low (quiet) side by the aircraft's performance.**
 - **Narrow range of noise levels to detect onset of annoyance**
- **Ability to achieve design noise setpoints is influenced by aircraft performance modeling, control, and atmospheric conditions.**
 - **Noise levels have relatively coarse setpoint control**

Limitations of Available Guidance on Binary Experimental Design Strategies

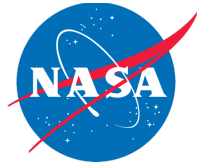


- **Design of experiments (DOE) for binary responses is an active area of research and practitioner-friendly literature is much less available than for continuous responses in traditional DOE.**
 - **Practical constraints is even more sparse in literature.**
- **Textbook approaches for dose-response experimentation usually expect a large sample size with dose levels spread across the full range of responses that solicit clear separation of the binary response.**
 - **Operational constraints and objectives drive a small sample size over a narrow, focused tail region of the dose range.**



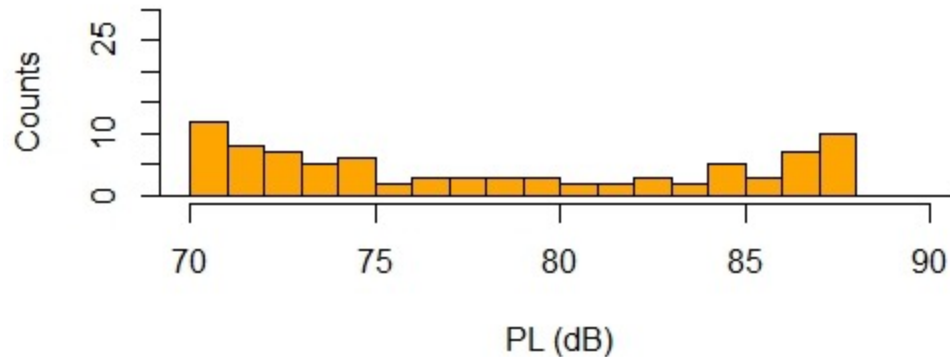
Elements of Design Building

- **Collaboration with subject matter expert is essential to appreciate the expected binary response characteristics, operational constraints, existing literature, presuppositions, and colloquial terminology.**
- **Teaching statistical design and modeling concepts enables self-realization of the impact of practical constraints and manages expectations.**
- **Challenging the “desirements” vs. “requirements” is often necessary.**
- **Basics of design of experiments remain foundational.**
 - **How many design points, which ones, and in what order?**
 - **Replication, randomization, and blocking still apply.**
- **Blocking can be an operational mitigation strategy that addresses potential disruptions, along with robustness to lurking systematic effects.**
 - **Especially for an experimental aircraft deployment.**



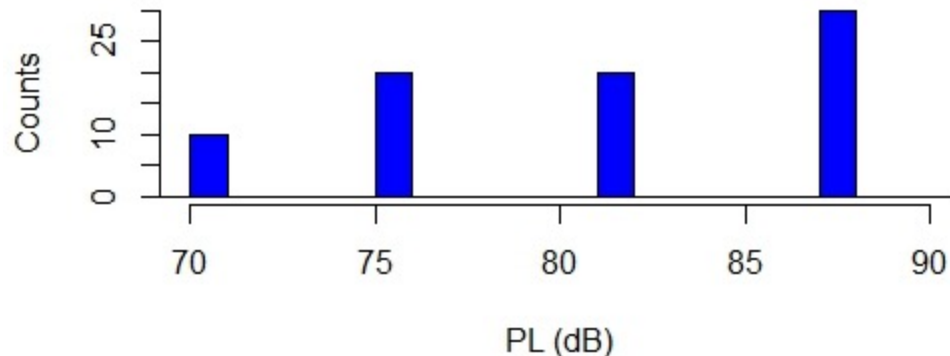
Candidate Experimental Designs

Baseline Single Event Design

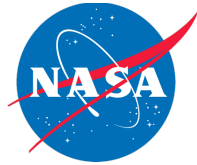


Baseline design proposed many discrete levels and with a concentration of noise levels at the low end of the range.

Alternative Candidate Single Event Design



Alternative design features replicates at 4 nominal levels and concentrates more design points at the upper end of the range to solicit rare responses.



Design Blocking Strategies

Considering the Alternative Design (previous slide) with 4 discrete levels.
What are candidate blocking strategies over 5 weeks of testing?

Option #1: Optimistic

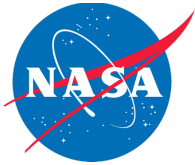
- Expect to complete test campaign
- All weeks have the same distribution of design levels

Level	Weekly
70	2
75	4
81	4
87	6

Option #2: Pragmatic

- Robust to campaign disruption
- Maximize early information, focus on higher levels to solicit rare responses

Level	Wk1	Wk2	Wk3	Wk4	Wk5
70	0	3	3	1	1
75	3	3	5	3	3
81	5	4	4	4	4
87	8	6	4	8	8



Concluding Remarks

- **Experiments with binary responses are common in aerospace and defense testing, and they always involve practical constraints.**
- **Experimental design guidance is limited for binary experiments, however, focusing on the foundational DOE principles remains an efficient framework to build designs within the operational constraints.**
- **Opportunities to improve the experimental design may be underappreciated or overlooked due to an over-emphasis on the data analysis and modeling.**
- **Collaboration among statisticians and subject-matter experts is always essential to developing robust and efficient designs.**