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

- **Binary responses are common in defense and aerospace testing.**
  - **e.g., hit or miss, detect or not detect, success or fail**
- **Objective is a statistical model to predict the probability a binary response's occurrence as a function of an explanatory variable(s).**
- **Often, there is more focus on modeling the data rather than building an efficient experimental design.**
- **Design constraints imposed by physics, ethics, resources, cost, time, and system capability restrict sample size, factor range, and distribution of factor levels.**
- **We propose a back-to-basics perspective that employs experimental design principles, emphasizing collaboration, for binary responses.**

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 true binary variables, not a 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
- **Generalized linear model (GLM) framework, logistic regression is the most common. Probability of a 1/0 as 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.**

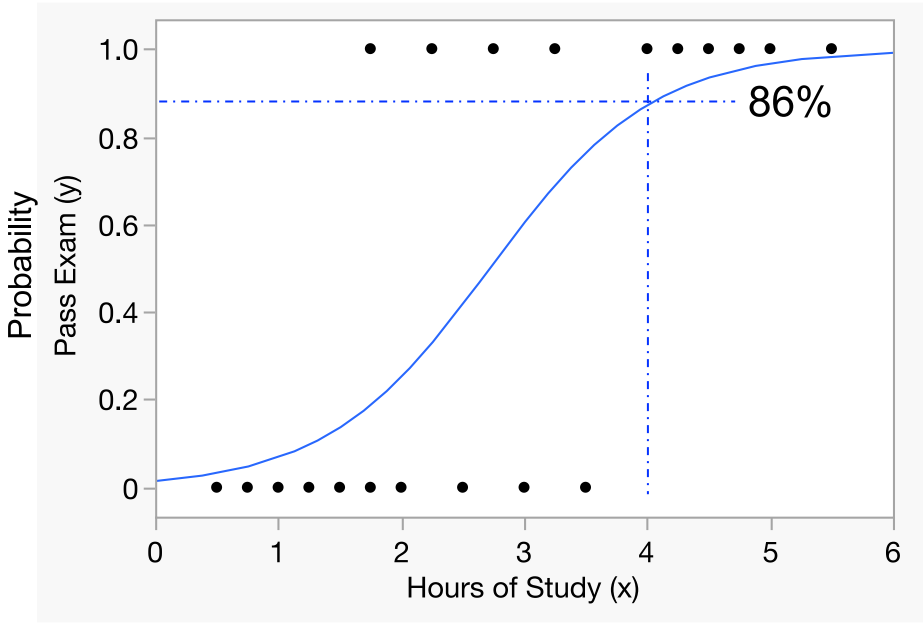
Simple Example of Logistic Regression

- **Probability a student passes an exam based on study hours.**
- **Data table for a student, 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

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

$$\beta_0 \approx -4.1$$
$$\beta_1 \approx 1.5$$



Using the model

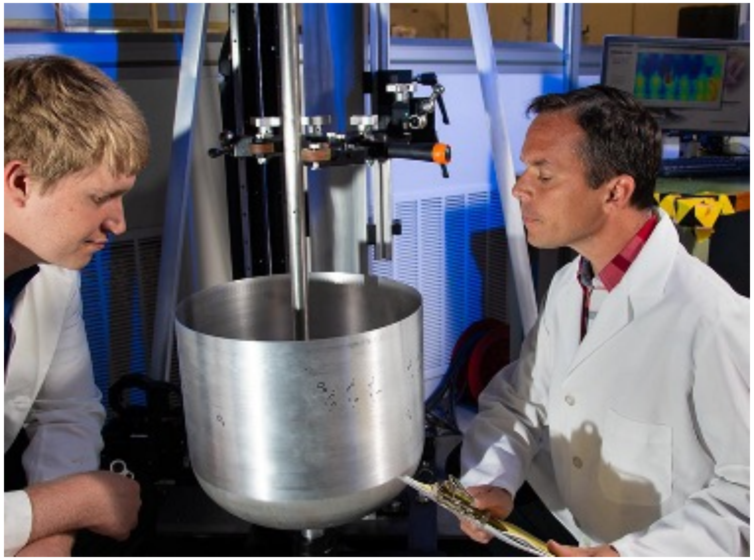
Probability of passing the exam with 4 hours of study?

p(4 hours) = 0.86 => 86%

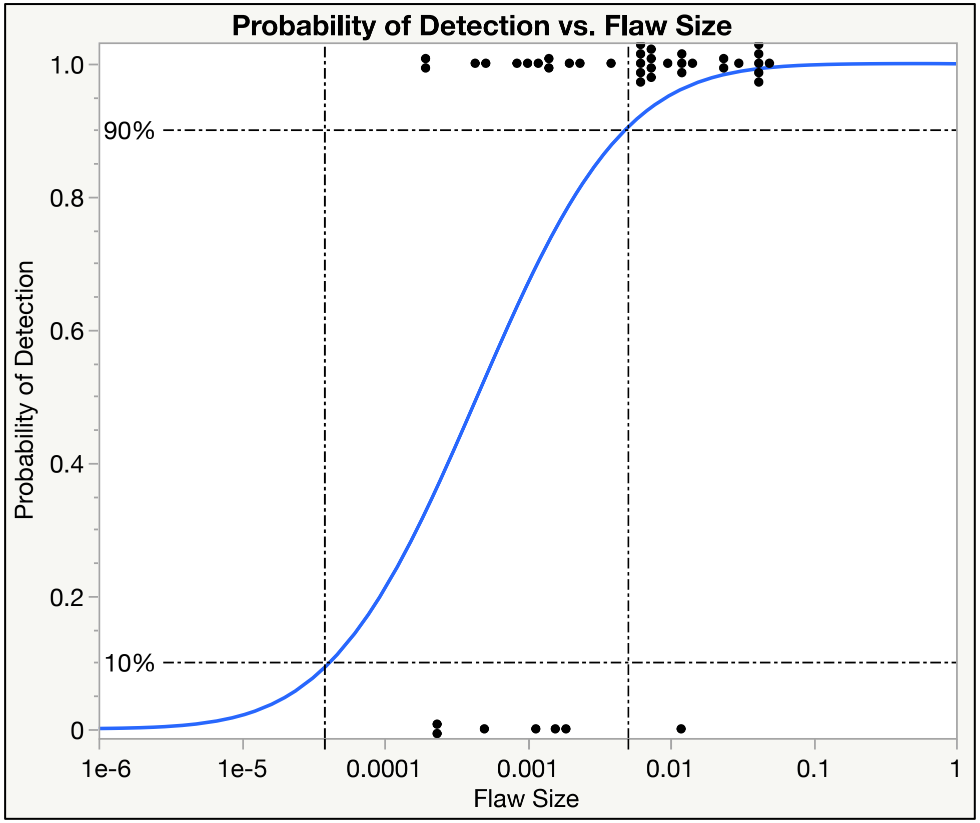
Wikipedia, https://en.wikipedia.org/wiki/Logistic_regression, accessed on April 1, 2024.

Probability of Detection

- Model 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).



Predict the flaw size with 90% POD at 95% conf.



Probability of Annoyance

- **Model 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 with varying noise levels and surveying participants for their level of annoyance, highly-annoyed (1) or not (0).**

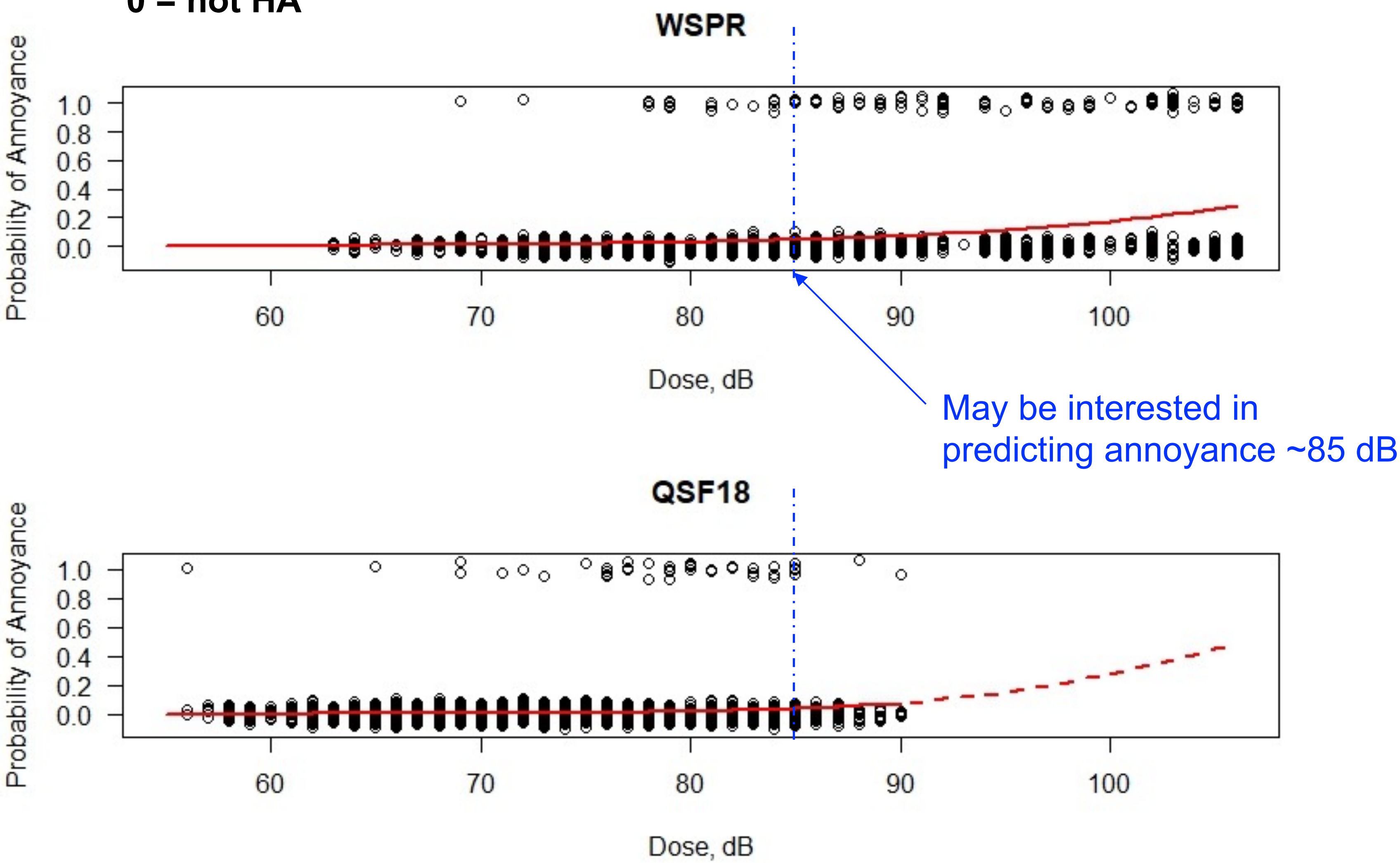
May be interested in predicting the noise level that provides:

5% POA with 95% confidence

Noise-Annoyance Experimental Data

1 = highly annoyed (HA)

0 = not HA



Lee, Rathsam, Wilson (2019) "Statistical Modeling of Quiet Sonic Boom Community Response Survey Data," NASA TM
Lee, Rathsam, Wilson (2020), "Bayesian Statistical Models for Community Annoyance Survey Datta," Journal of Acoustical Society of America

Noise-Annoyance Design Constraints

- **Each noise level requires an experimental aircraft flight. Total number of design points (noise levels) is limited by flights per day and flights per operational deployment at a community site.**
  - **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 side (quiet) by the aircraft's performance.**
  - **Narrow range of noise levels to detect onset of annoyance**
- **Ability to achieve design noise setpoints is influenced by uncertainty in the aircraft performance modeling and atmospheric conditions.**
  - **Noise levels have relatively coarse setpoint control relative to range. They are measured during experiment, with error.**

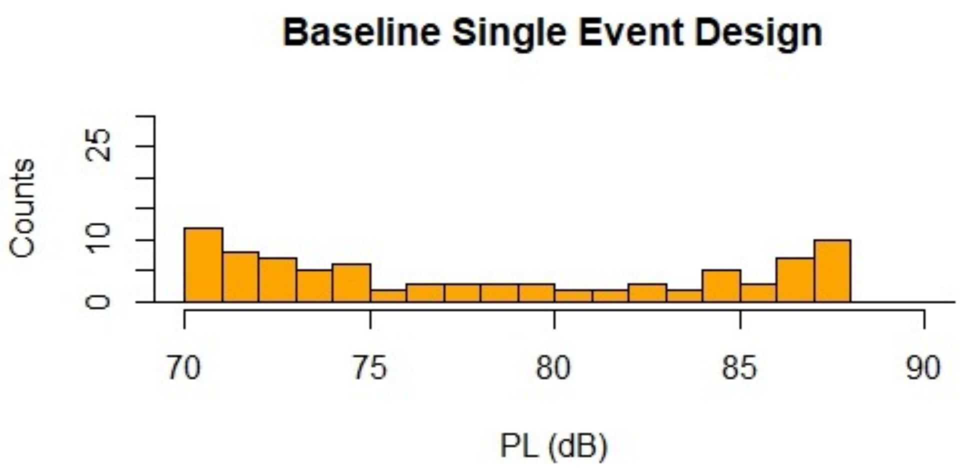
Limitations of Available Guidance on Binary Experimental Design Strategies

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

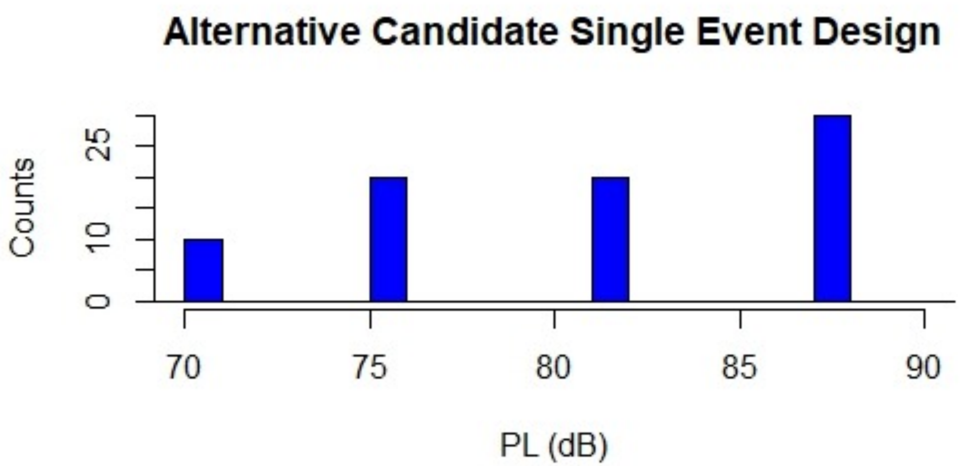
Elements of Design Building

- Collaboration with subject matter expert is essential to appreciate the expected binary response characteristics, operational constraints, existing literature, presuppositions, and colloquial terminology.
  - Explaining statistical design concepts is helpful in realizing the impact of practical constraints and manages expectations.
- Basic philosophy of design of experiments remains foundational.
  - How many design points, which ones, and in what order?
  - Replication, randomization, and blocking still apply.
- Blocking may be an operational mitigation strategy for potential disruptions and robustness to lurking systematic effects.
  - Especially in developmental systems (experimental aircraft).
- Challenging the “desirements” vs. “requirements” of the test objectives is often necessary, especially for binary responses.

Candidate Experimental Designs



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



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

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 feature the same distribution of design levels**

Dose Level (dB)	Weekly Doses
70	2
75	4
81	4
87	6

Option #2: Pragmatic

- **Robustness to campaign disruption**
- **Maximize information early by focusing on higher levels to solicit rare responses**
- **"Bookend" with high dose weeks to assess acclimation effects**

Dose 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.**
- **Experimental design may be underappreciated or overlooked due to an over-emphasis on modeling.**
- **Collaboration among statisticians and subject-matter experts is always essential to developing robust and efficient designs.**

Foundational DOE concepts provide an efficient framework to build designs with operational constraints for binary responses