Westat

Improving Lives Through Research®

Modeling approaches to estimate community annoyance due to sonic booms using data from repeated surveys

Robyn Ferg

Jean Opsomer

FCSM 2022

Introduction

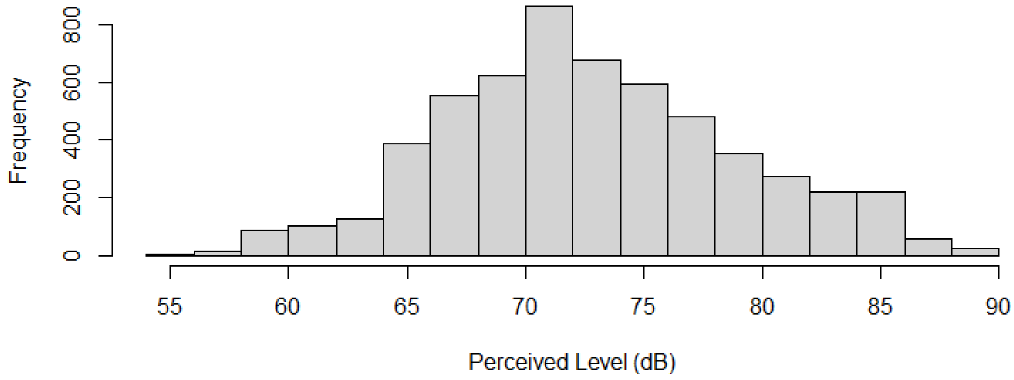
- NASA is developing the X-59 aircraft, for researching the effects of low-magnitude sonic booms (thumps)
- Westat is part of the project team, responsible for conducting community surveys to estimate community annoyance levels based on noise level from X-59 supersonic flights
- Use data from previous flight test (QSF18) to develop and apply candidate approaches
- QSF18 study used different plane and different maneuver, but methods still applicable to X-59

QSF18 Study

- Flights took place over a two-week period in November 2018 in Galveston, Texas
- 52 sonic thumps delivered over the test period
- Noise monitors used to measure sound levels across the survey area

QSF18 Data

- 5634 total observations from 373 participants
- Each observation has an estimated noise level (PL) based on the participant's location at the time of the flight



QSF18 Data

- Annoyance is measured on a five-point scale:
  - Not at all annoyed
  - Slightly annoyed
  - Moderately annoyed
  - Very annoyed
  - Extremely annoyed
- Top two categories are considered ‘highly annoyed’
- Challenge: only 1.1% of responses were highly annoyed

QSF18 Data

- Demographic data
  - Gender
  - Age
  - Household size
  - Kids under 6
  - Geographic quadrant
  - Education
- No weights included in the data set
- Since X-59 study will include weights calibrated to Census Bureau totals, simulate weights for QSF18 data ~unif(50, 250)

Quadrants

C
20
D
116
B
212
A
148
29.395192°, -94.943149°
29.330446°, -94.870243°
29.222446°, -95.004079°

Multilevel Logistic Regression

- Multilevel logistic regression model with participant-level random effects is a common method of creating a dose-response curve for this type of data

- Model log-odds of high annoyance (HA) based on noise level PL_{ij} and covariates X_j, with random effects for each participant j:

$$HA_{ij} \sim Binom(1, p_{ij})$$

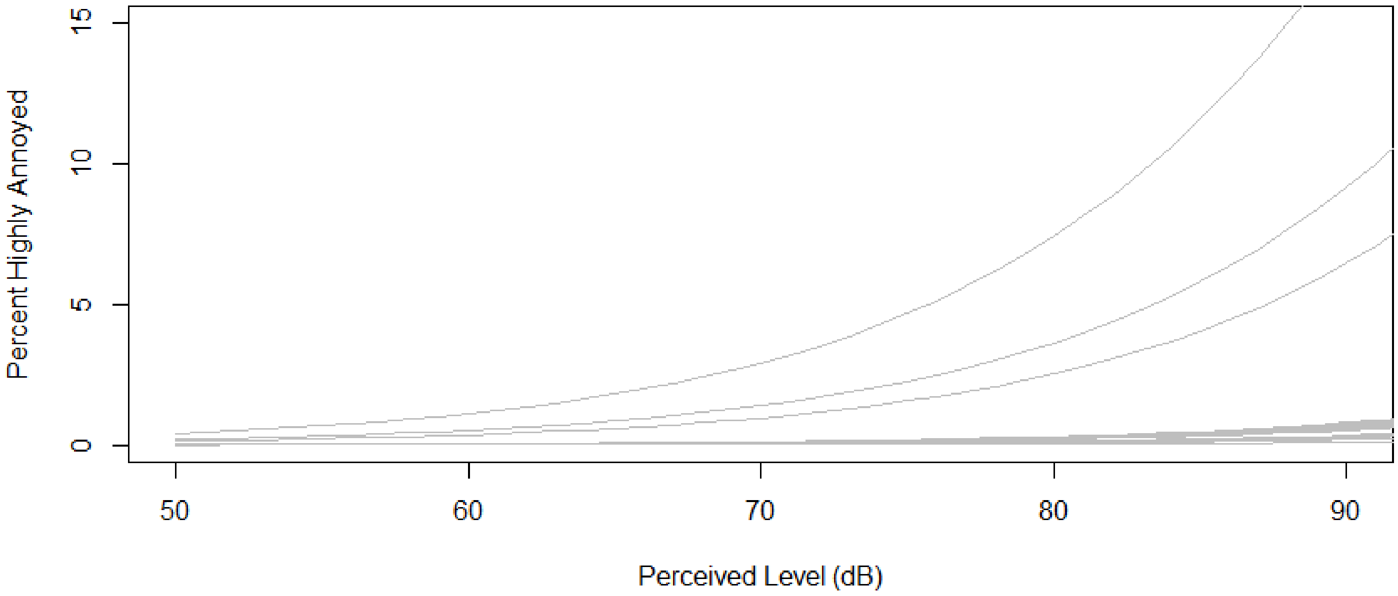
$$p_{ij} = logit^{-1}(\beta_0 + \beta_1 * PL_{ij} + \beta_2 X_j + u_j) = \frac{1}{1 + exp[-(\beta_0 + \beta_1 * PL_{ij} + \beta_2 X_j + u_j)]}$$

$$u_j \sim N(0, \sigma^2)$$

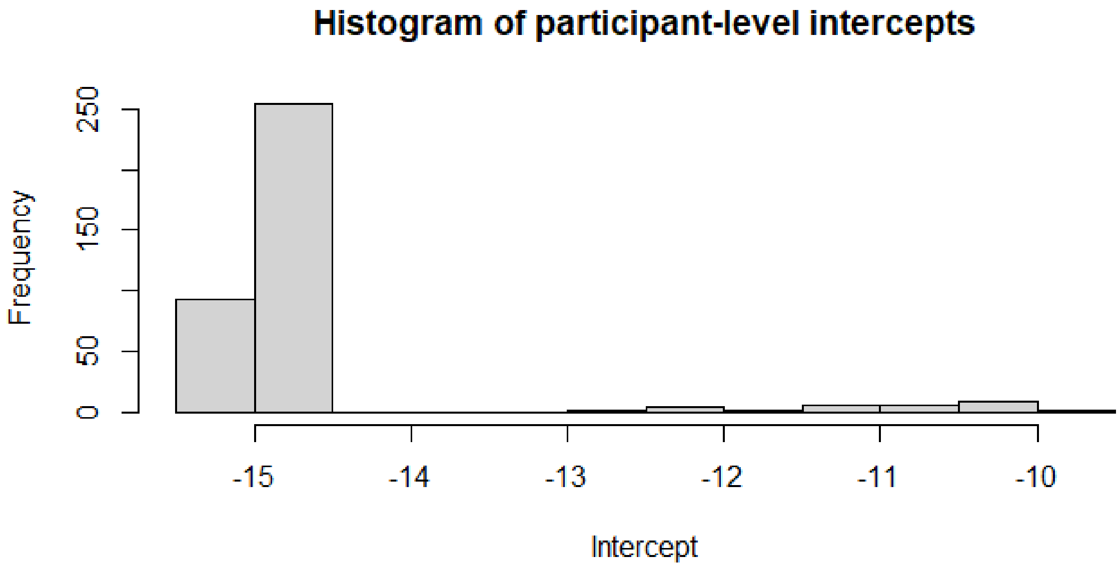
- For demonstration, use gender and geographical quadrant as covariates

- Note: Unweighted, so resulting curve will not be representative of the population

Sample of 100 Participant-Level Dose-Response Curves



Participant-Level Intercepts



Proposed Two-Stage Model

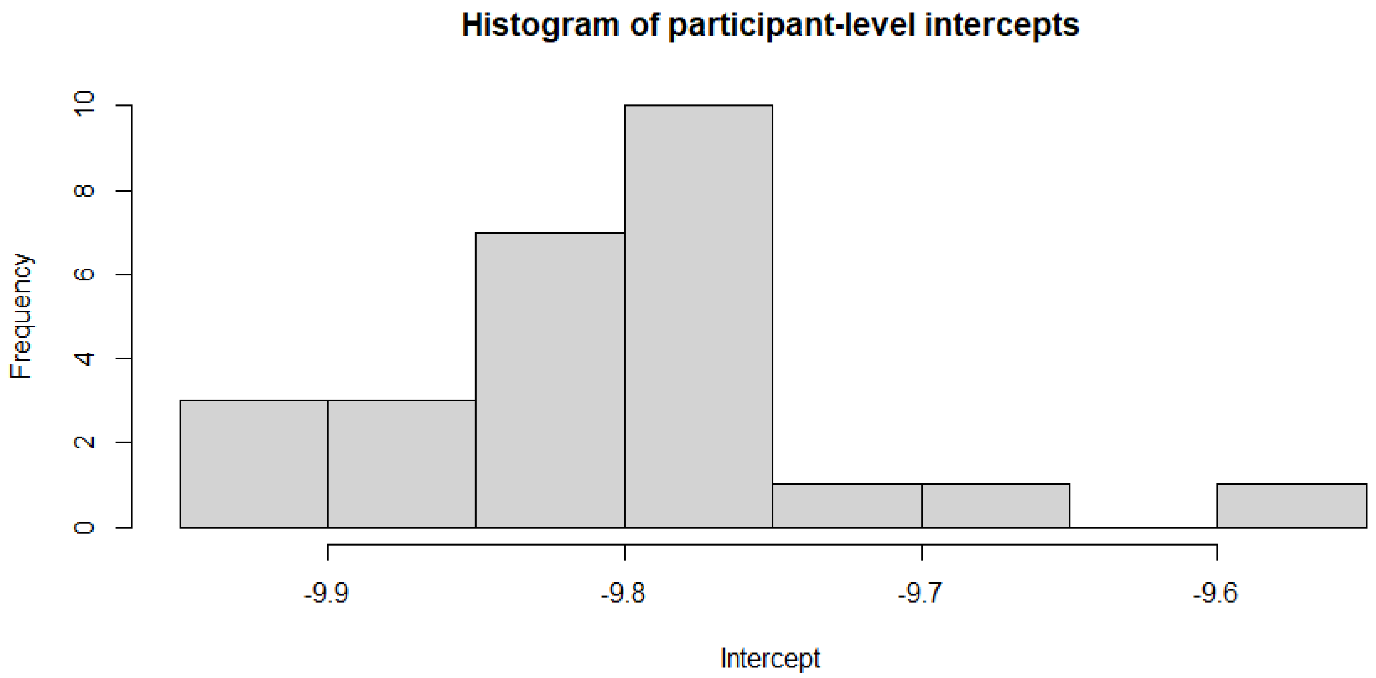
- Break up modeling into two stages
- Stage 1: Model probability of ever being highly annoyed given demographics of participant j
  - Logistic regression:

$$p^{(1)}(X_j) = logit^{-1}(\beta_0 + \beta_1 X_j) = \frac{1}{1 + \exp[-(\beta_0 + \beta_1 X_j)]}$$

- Stage 2: Model probability of high annoyance given demographics of participant j and noise level PL_{ij} given participant j was ever highly annoyed
  - Multilevel logistic regression:

$$p^{(2)}(PL_{ij}, X_j) = logit^{-1}(\alpha_0 + \alpha_1 X_j + \alpha_2 PL_{ij} + u_j) = \frac{1}{1 + \exp[-(\alpha_0 + \alpha_1 X_j + \alpha_2 PL_{ij} + u_j)]}$$

2nd Stage Participant-Level Intercepts



Marginal Dose-Response Curve

- To accurately account for the participant-level random effects in predicting the dose-response curve for the population, we integrate over the distribution of random effects (Pavlou et al. 2015)
- For a given PL noise level, the predicted fraction of people highly annoyed among those ever highly annoyed with covariates X is

$$\hat{p}^{(2)}(PL, X) = \int_{-\infty}^{\infty} logit^{-1}(\hat{\beta}_0 + \hat{\beta}_1 * PL + \hat{\beta}_2 X + u)\, f(u)\, du$$

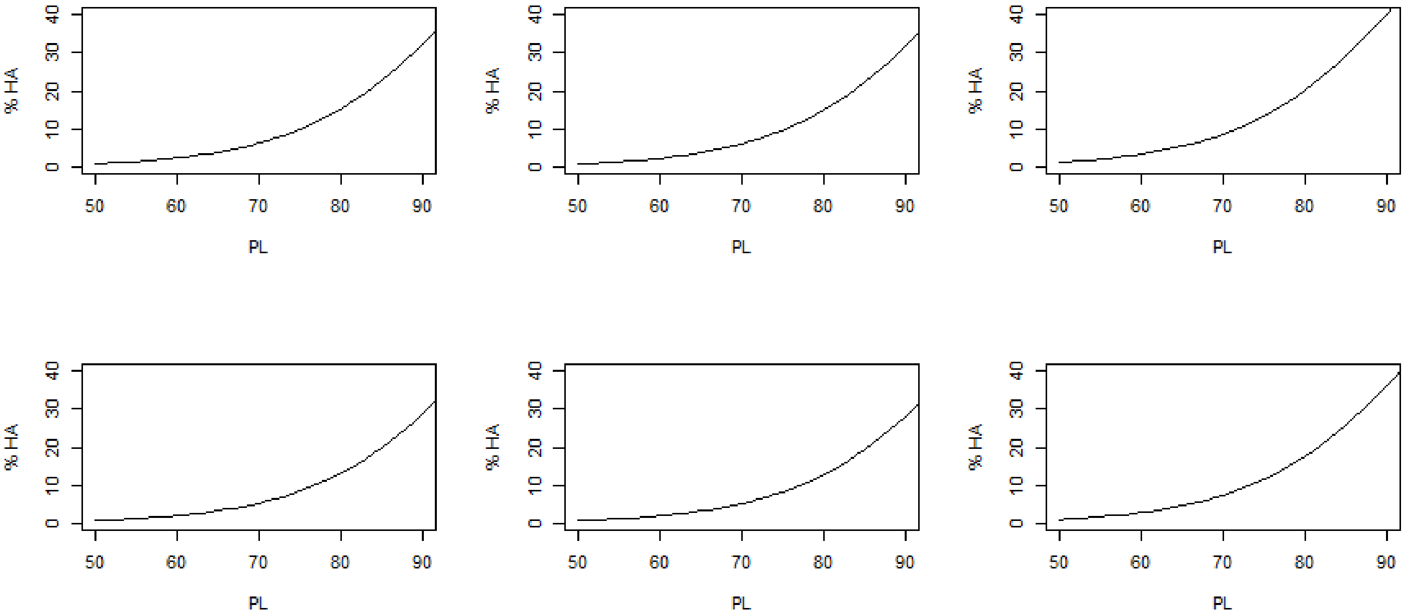
where $f(u)$ is a normal distribution with mean 0 and variance equal to the estimated random effect variance

2nd-Stage Marginal Poststrata Curves

Quadrant A | Quadrants B&C | Quadrant D

Male

Female

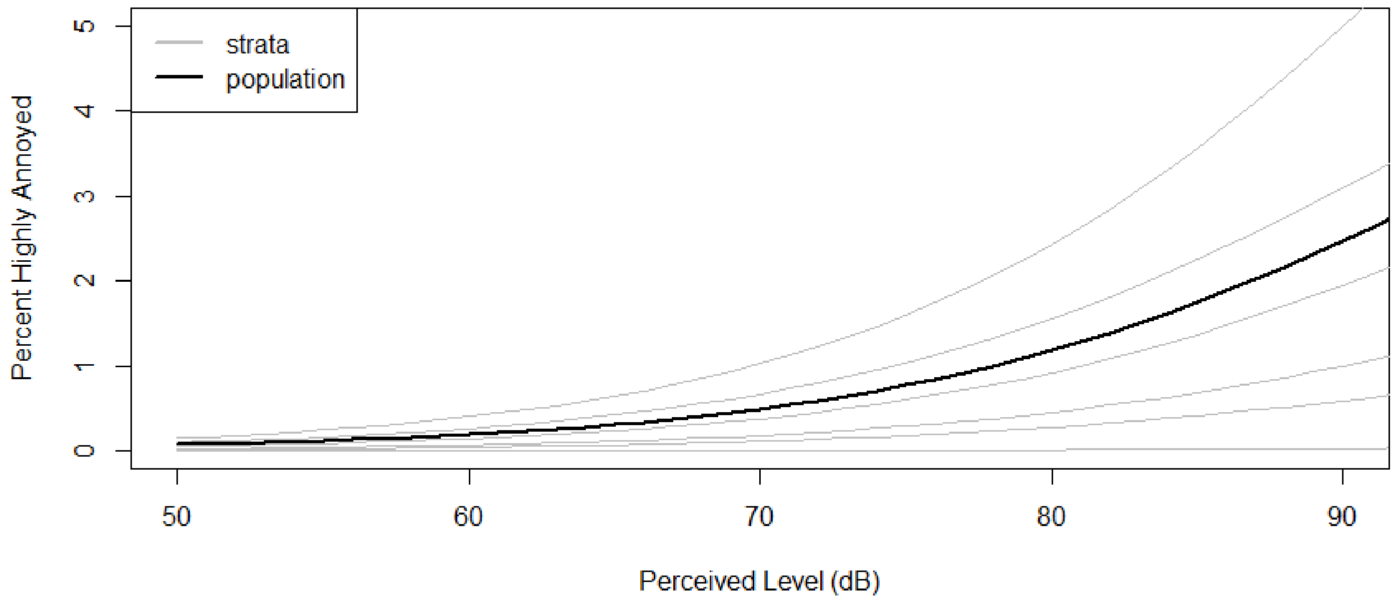


Multilevel Regression and Poststratification

- To generalize the dose-response curve to the population, we use Multilevel Regression and Poststratification (MRP) (Gelman & Little, 1997)
- Using the fitted multilevel regression model, create marginal dose-response curves for each poststratum
- Create a weighted average of these curves based on the proportion of the population within each poststratum
- Let $X^{(g)}$ be the g^{th} combination of categorical variables in the study, where $g = 1, \ldots, G$ and $N = \sum_{g=1}^{G} N_g$. Then the estimated probability of high annoyance for the population is

$$\hat{p}(PL) = \frac{1}{N}\sum_{g=1}^{G} N_g \hat{p}^{(1)}\left(X^{(g)}\right)\hat{p}^{(2)}\left(PL, X^{(g)}\right)$$

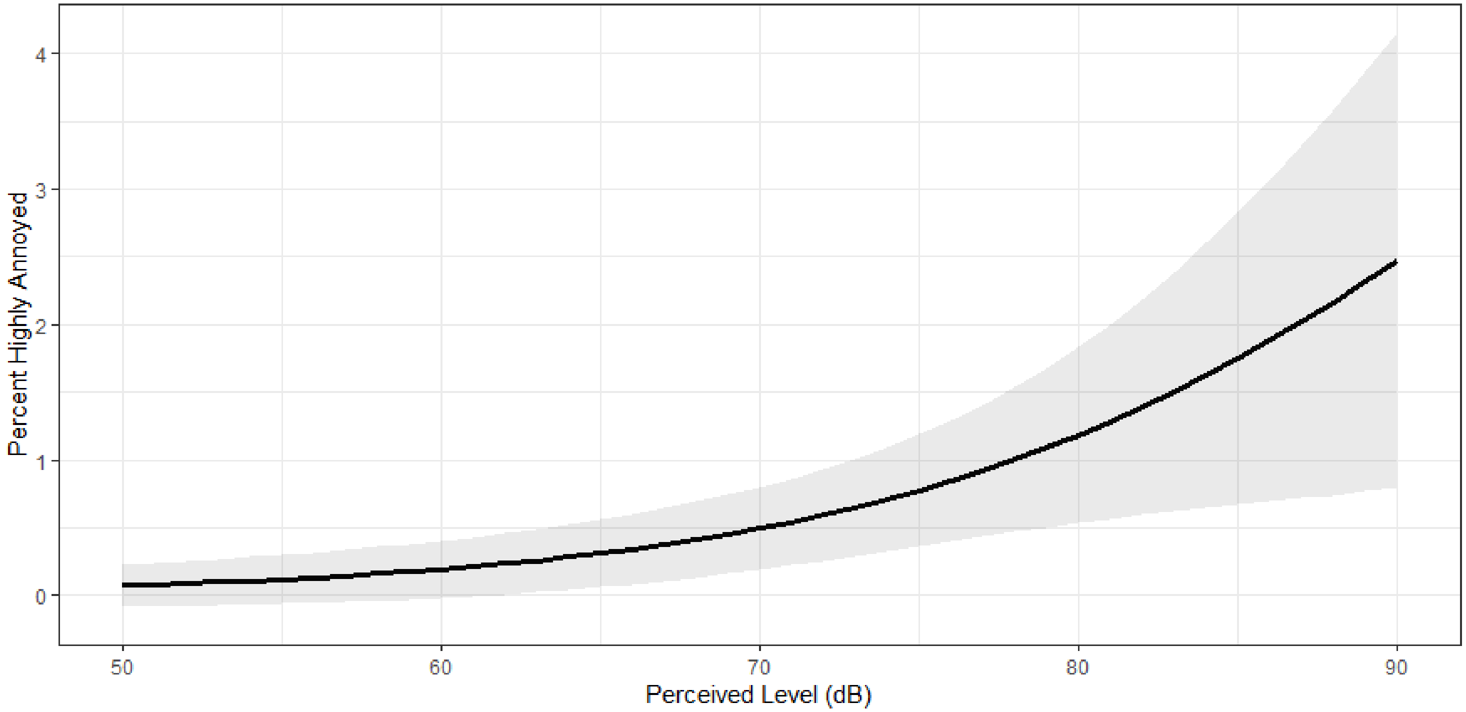
Population Dose-Response Curve



Inference

- Confidence intervals and standard errors can be estimated through bootstrapping
- Take 100 samples with replacement from the data
- Repeat the exact same process of fitting a multilevel logistic regression model and use MRP to estimate the study region population-level curve for the subsample
- 100 resulting study region population-level curves
- At each PL level, take the standard deviation of the 100 curves

Population-Level Dose Response Curve with 95% Confidence Interval



Summary

- Two-Stage Model
- Multilevel regression and Poststratification (MRP)
- Bootstrap inference
- Not included, but possible extensions:
  - Multilevel ordinal regression or monotone regression model
  - Bayesian hierarchical model to account for covariate (noise level) uncertainty