

STA 712 Exam 2

Due: Friday, November 18, 12:00pm (noon) on Canvas.

Instructions: You have until the due date to complete this exam. There are no other time restrictions.

Submit your work as a single PDF. For this exam, you may include written work by scanning it and incorporating it into the PDF. Include all R code needed to reproduce your results in your submission.

For this exam, you may:

- Use any resources from the course (the textbook, the course website, class notes, previous assignments, etc.)
- Use the internet for R help
- Email me, or come to office hours, with specific questions (I may be somewhat less helpful than for regular assignments)

You may *not*:

- Discuss the exam with anyone else
- Use the internet or other textbooks (except for R help and to access the course materials)

Part I: Modeling a Geometric response

1. If $Y \sim \text{Geometric}(p)$, then Y takes values $y = 1, 2, 3, \dots$ with probabilities

$$f(y; p) = p(1 - p)^{y-1},$$

where $p \in [0, 1]$.

- (a) Show that the geometric distribution is an EDM by identifying θ , $\kappa(\theta)$, and ϕ .
 - (b) Use (a) to find $\mathbb{E}[Y]$ and $\text{Var}(Y)$.
 - (c) Deduce the canonical link function for the geometric distribution.
 - (d) Find the variance function $V(\mu)$ for the geometric distribution.
 - (e) Derive the unit deviance for the geometric distribution, and write the probability function in dispersion model form.
2. Suppose that the Deacon OneCard Office is tired of Wake Forest students losing their ID card and having to get a replacement. The office posits a relationship between the number of credits a student has completed at the university, and the number of ID cards the student has been issued (all students have been issued at least one ID card). They propose the following model:

$$\begin{aligned} \text{cards}_i &\sim \text{Geometric}(p_i) \\ \log\left(\frac{1}{p_i} - 1\right) &= \beta_0 + \beta_1 \text{credits}_i \end{aligned} \tag{1}$$

We have data on a sample of 1000 Wake students. You can load the data into R by

```
id_cards <- read.csv("https://sta712-f22.github.io/exams/ids.csv")
```

In base R, the `glm` function supports families which include the `binomial`, `gaussian`, `poisson`, `Gamma`, and `quasipoisson`. The `MASS` package also allows the `negative.binomial` family (with a specified r , which is referred to as `theta` in R), and the `glm.nb` function fits a negative binomial model with unknown r .

- (a) You will note that the Geometric distribution is not listed among the families supported in base R or in `MASS`. Explain how you could still fit the model in Equation (1) using one of the supported families.
- (b) Using your answer to (a), fit the model proposed by the Deacon OneCard Office, and report the estimated coefficients.

Part II: Count regression with insects

In the second part of this exam, you will work with a real dataset from researchers interested in studying how insects are attracted to artificial lights. In the original study, the researchers set up light traps containing different bulbs, and captured any insect attracted by the light. The traps were set up each evening at sunset and then checked in the morning after sunrise, and the insects collected identified and counted.

You will work with a subset of this insect data, consisting of data on 190 traps with the following variables:

- **Light.Type**: Type of light bulb used to attract insects. ‘No’ means no bulb at all; ‘A’, ‘B’, and ‘C’ are all Philips light bulbs whose color can be adjusted by the user; ‘LED’ is a commercial white LED bulb; and ‘CFL’ is a commercial compact fluorescent bulb.
- **Standardized.Moon**: Standardized variable representing the percentage of the moon visible on the night chosen (i.e., the moon phase).
- **Standardized.Mean.Temp**: Temperature on the night, standardized.
- **Standardized.Mean.Humidity**: Humidity on the night, standardized.
- **Total**: Total number of insects captured in the trap

You can load the data into R by

```
bugs <- read.csv("https://sta712-f22.github.io/exams/bugs_small.csv")
```

Let μ_i denote the mean number of insects captured in trap i . The researchers propose the following model:

$$\log(\mu_i) = \beta_0 + \beta_1 \text{Light}B_i + \beta_2 \text{Light}C_i + \beta_3 \text{Light}CFL_i + \beta_4 \text{Light}LED_i + \beta_5 \text{Light}No_i + \beta_6 \text{Moon}_i + \beta_7 \text{Temp}_i + \beta_8 \text{Humidity}_i$$

3. The researchers have specified a relationship for the mean μ_i , but have not specified a distribution for the response $Total_i$. Choose between a Poisson, quasi-Poisson, and negative binomial model by fitting models in R, creating appropriate diagnostic plots, and assessing goodness of fit and overdispersion.
4. The researchers are interested in how the amount of natural light (the percentage of the moon visible) is related to the number of insects attracted by artificial light. Using your model from Question 3:
 - (a) Interpret the estimated coefficient $\hat{\beta}_6$ for the standardized percentage of the moon visible on the night.
 - (b) Calculate and interpret a 90% confidence interval for the parameter β_6 in (a).

5. Now the researchers want to test whether different types of lights are associated with different numbers of insects attracted to artificial lights. Using your model from Question 3:
- (a) Test whether there is a relationship between the type of artificial light and the number of insects caught in the trap, after accounting for the percentage of moon visible, the temperature, and the humidity. You should state the hypotheses in terms of one or more model parameters, calculate the test statistic and p-value, and make a conclusion.
 - (b) Test whether the relationship between the amount of natural light (the moon) and the number of insects caught in traps depends on the type of light bulb used to attract insects, after accounting for temperature and humidity. (*Hint: you may need to modify your model from Question 3.*) You should state the hypotheses in terms of one or more model parameters, calculate the test statistic and p-value, and make a conclusion.