ES/STT 7140: Final Exam

April 10, 2018

Note: The final is due by the end of class on 04/18/2018 (you may drop it off anytime during class on that day). The completed final must be delivered in person (email is not acceptable). Since this is an exam, working together is prohibited, but you are welcome to use the book, course notes, etc. Good luck!

Question 1 (15 points) Run the R script `hematology.R` and answer the following questions:

1. Based on a scatterplot, does a simple linear regression model seem reasonable for these data? Explain your answer using 2–3 complete sentences.

2. Fit a simple linear regression model and write out the estimated equation for the line.

3. Using $\alpha = 0.05$, test the hypothesis $H_0 : \beta_1 = 0$ vs. $H_1 : \beta_1 \neq 0$. Interpret the results of the test in the context of this problem using 2–3 complete sentences.

4. Based on residual diagnostic plots, does it look as though these data satisfy the assumptions of the simple linear regression model? Explain your answer using 2–3 complete sentences.

5. Compute a 95% confidence interval for the slope β_1 and interpret the results.

6. Summarize the results of the analysis using 2–3 complete sentences.

Question 2 (15 points) Korich et al. (2000) reported a study of how laboratory mice respond to the exposure of microscopic parasites of the genus *Cryptosporidium.* The objective of the study was to develop a *dose-response* model, that is, a model to predict the probability that a mouse will be infected if it is exposed to a certain number (i.e., dose) of parasites. The study is important because, according to the U.S. Center for Disease Control, *Cryptosporidium* is one of the most common causes of waterborne human diseases in the United States during the past two decades. Once an animal or person is infected, the parasite lives in the intestine and passes in the stool. The parasite is protected by an outer shell that allows it to survive outside the body for long periods of time and makes it very resistant to chlorine-based disinfectants. Both the disease and the parasite are commonly known as "crypto." The parasite may be found in drinking water and recreational water throughout the world.

The U.S. EPA in a 2006 publication requires surface water systems in the U.S. to provide treatment using ultraviolet light (UV) disinfection as one treatment option to meet treatment requirements. Many drinking water utilities in the U.S. are looking at UV disinfections as the best available technology for meeting their crypto inactivation requirements and goals. The term "inactivation" refers to the fact that UV radiation typically does not kill the parasite, but alters the parasite's nucleic acids, thus preventing replication and infection. As a result, the effectiveness of UV radiation must be evaluated using mouse-infectivity studies.

In a typical mouse-infection study, a dose-response relationship is first determined. In this step, a number of mice are inoculated with a known number of crypto oocysts (d) to observe the number of infected mice. The resulting infectivity data are used to fit a dose-response model, typically a log-linear logistic model of the form:

$$logit\,(p) = \beta_0 + \beta_1 \log_{10}(d)\,, \tag{1}$$

where p is the probability of infection, and d is the number of oocyst ingested.

The data can be loaded by running the following R code

```
url <- paste0("https://raw.githubusercontent.com/bgreenwell/",
              "eesR/master/R/Data/cryptoDATA.csv")
crypto <- read.table(url, header = TRUE)
```

1. Based on a scatterplot of the observed proportion of infections versus $\log_{10}(d)$, does there appear to be any relationship between the likelihood of infection and the number of oocyst ingested? Explain your answer using 2–3 complete sentences.

```
plot(Y/N ~ log10(Dose), data = crypto)
```

2. Use the code below to fit the simple dose-response model seen in Equation (1). For every 1-unit increase in $\log_{10}(d)$, what is the multiplicative effect on the estimated odds of infection? Construct a 95% confidence interval for this parameter and interpret (this is done for you in the code chunk below). (Do not worry if you get a warning after running th code!)

```
fit <- glm(cbind(Y, N-Y) ~ log10(Dose), data = crypto,
            family = binomial(link = "logit"))
summary(fit)  # print model summary
exp(confint(fit, parm = 2))  # 95% confidence interval
```

3. Using the fitted model from part 2), estimate the median effective dose, denoted Ld_{50}; that is, estimate the dose that will result in a probability of infection of 50%.