Homework 7

Math 445, Spring 2017

Due Wednesday, May 31 by 4:30 p.m.

Instructions

- Please complete problems that require the use of R in R markdown. Theoretical problems may be handwritten. **Please knit your .Rmd file to either Word or PDF and print it for submission** with the remainder of the problems.
- All assignments should be stapled.

Assignment

Chapter 8, Exercise 24

Chapter 8, Exercise 30

Chapter 8, Exercise 32

Chapter 8, Exercise 37

Problem 5

Clouding seeding is the practice of using an airplane to spray certain chemicals into a cloud. The hope is to increase the amount of rain that falls from the cloud. It is difficult to evaluate the effectiveness of cloud seeding. The following data are from one of the best studies.

This study was conducted in southern Florida between 1968 and 1972. The weather was watched daily. If the cloud conditions were considered suitable, then that day was included in the experiment. A total of 52 days were considered suitable. Each suitable day, one target cloud was arbitrarily chosen and randomly assigned to one of two treatments:

seed: fly an airplane through the cloud and spray the seed chemical

control: fly an airplane through the cloud but the sprayer was not loaded with the seed chemical.

The total rain that fell from the target cloud was measured. This experiment was double blind (neither the pilots nor rainfall observers knew which treatment was used). The file `rainfall.csv` on the class web site contains two variables: treatment and rain for each of the 52 clouds in the study.

(a) Do you expect a problem with the assumption of independent groups?
(b) Consider the untransformed rainfall values by treatment. Is a normal distribution reasonable? Explain why or why not.
(c) For the untransformed rainfall values, use a graphical method to assess the assumption of equal variances. Is the assumption reasonable? Explain why or why not.
(d) Create a new variable in the data set, `lograinfall`, that applies a natural log transformation to the rainfall values. Is a normal distribution reasonable for the log transformed values?
(e) It is reasonable to assume that the log-transformed values have equal variances? You can use a graphical approach to make this assessment.
(f) Use an appropriate two-sample t-test on the log-transformed rainfall to determine whether seeding effects rainfall.

Problem 6

You are interested in determining who sends more text messages: graduate students or undergraduate students. Results of a survey that measured the variables $Y =$ texts, the number of texts sent per day, and $X =$ grad, a binary (0 or 1) variable indicating whether a student is a grad students (grad=1) or an undergrad (grad=0) for $n = 91$ Harvard students are in the file `textsHW7.csv`. Let $Y_i \overset{\text{iid}}{\sim} \text{LogNormal}(\mu = \beta_0 + \beta_1 X_i,\ \sigma^2)$. The lognormal distribution has PDF

$$f(y|\mu, \sigma^2) = \frac{1}{y\sqrt{2\pi\sigma^2}} \exp\left\{\frac{-(\log y - \mu)^2}{(2\sigma^2)}\right\},\ y \in (-\infty, \infty),\ \mu \in (-\infty, \infty),\ \sigma > 0,$$

and $\text{E}(Y) = \exp\left\{\mu + (\sigma^2/2)\right\}$.

(a) Derive the log-likelihood function.

(b) Write a function in R (call it something like `lognormal.loglik` that computes the log-likelihood for a lognormal distribution given three parameters (call it `theta` where `theta[1]` is β_0, `theta[2]` is β_1, and `theta[3]` is σ^2) and the data vectors `y` and `x`.

(c) In R, read in the data set and calculate the mean and variance of the number of texts for each group (undergraduate students and graduate students). How many of each type of students are in the sample?

(d) Use your function from part (b) and the `optim` function to calculate the maximum likelihood estimates for $\beta_0,\ \beta_1$, and σ^2. What is the value of the log-likelihood function at the ML estimates?

(e) Based on this model, what is the estimated average number of text messages for undergraduates from this model? How about for graduate students?

(f) In preparation to test $H_0 : \beta_1 = 0$ vs. $H_a :\ \beta_1 \neq 0$, create a new R function, similar to the one in part (b), which passes just two parameters in `theta` (now, `theta[1]` is β_0 and `theta[2]` is σ^2). Notice that we no longer include β_1, as we will assume $\beta_1 = 0$ under the null hypothesis. Name this function `lognormal.loglik.null`.

(g) Test $H_0 : \beta_1 = 0$ against $H_a :\ \beta_1 \neq 0$ using a likelihood ratio test. Be sure to clearly show your steps in calculating the test statistic and the p-value (using R will be necessary here), and state your conclusion for the test. The results from part (d) and the function in part (f) will be helpful.

(h) Interpret the results of the hypothesis test in part (g). What does this mean for the difference in text messaging between undergraduate and graduate students?