

Math 150 - Methods in Biostatistics - Homework 5

your name here

Due: Wednesday, February 27, 2019, in class

```
knitr::opts_chunk$set(message=FALSE, warning=FALSE, fig.height=3, fig.width=5,
                        fig.align = "center")
library(tidyverse)
library(broom)
library(tidylog)
```

Note: there are two places to check for hints on R code. One is the class notes (<http://st47s.com/Math150/Notes/>, see R Examples) and the other is the R manual associated with the textbook which is on Sakai.

1. Chp 7, E1 Bird Nest study

The file `Birdnest` contains data for 99 species of North American passerine birds. Passerine are “perching birds” and include many families of familiar small birds (e.g., sparrows and warblers), as well as some larger species like crows and ravens, but do not include hawks, owls, water fowl, wading birds, and woodpeckers. One hypothesis of interest was about the relationship of body size to type of nest. Body size was measured as average length of the species. Although nests come in a variety of types (see the `Nesttype` variable), in this data set nest type was categorized into either closed or open. “Closed” refers to nests with only a small opening to the outside, such as the tree cavity nest of many woodpeckers or the pendant-style nest of an oriole. “Open” nests include the cup-shaped nest of the American robin. (Note: `Closed?` = 1 for closed nests; `Closed?` = 0 for open nests.)

```
birdnest <- read_csv("~/Dropbox/teaching/math150/PracStatCD/Data Sets/Chapter 07/CSV Files/C7 Birdnest.csv",
                    na="*")
glm(`Closed?` ~ Length, data=birdnest, family="binomial") %>% tidy()
```

```
## # A tibble: 2 x 5
##   term      estimate std.error statistic p.value
##   <chr>      <dbl>    <dbl>    <dbl>   <dbl>
## 1 (Intercept)  0.457    0.753     0.607   0.544
## 2 Length      -0.0677  0.0425    -1.59   0.112
```

- Create a logistic regression model using bird length (`Length`) to estimate the probability that a bird species has a closed net type. Interpret the model in terms of the odds ratio.
- Use the Wald statistic to create a 95% confidence interval for the odds ratio.
- Test $H_0 : \beta_1 = 0$ vs. $H_a : \beta_1 \neq 0$ using both Wald’s test and the likelihood ratio test. State your conclusions based on these tests.

skip (d), (e), (f) for now

2. Chp 7, E2 Donner Party: Logistic Regression and Chi-Square tests

In 1846, a group of 87 people (called the Donner Party) were heading west from Springfield, Illinois, for California. The leaders attempted a new route through the Sierra Nevada and were stranded there throughout the winter. The harsh weather conditions and lack of food resulted in the death of many people within the group. Social scientists have used the data to study the theory that females are better able than men to survive harsh conditions.

- Create a logistic regression model using `Gender` to estimate the probability of `Survival`.

```
donner <- read_csv("~/Dropbox/teaching/math150/PracStatCD/Data Sets/Chapter 07/CSV Files/C7 Donner.csv",
                  na="*")
```

```
names(donner) <- c("name", "gender", "age", "survived", "familysize", "X6", "X7", "X8", "X9",
                  "adultname", "adultgender", "adultage", "adultsurvived", "adultfamilysize")
```

- (b) Interpret the model in terms of the odds ratio. Use the Wald statistic to create a 95% confidence interval for the odds ratio.
- (c) Calculate and interpret the likelihood ratio test.
- (d) Test $H_0 : \beta_1 = 0$ vs. $H_a : \beta_1 \neq 0$ using both Wald's test and the likelihood ratio test. State your conclusions based on these tests.
- (e) No measures of association (yet).
- (f) Create a two-way contingency table using **Gender** and **Survival** as row and column variables. Conduct a chi-square test for equal proportions (e.g., is the proportion of survival the same for males and females?). In addition, use this two-way table to create the odds ratio. How does the analysis of the two-way table compare to the logistic regression analysis?
- (g) D.K. Grayson states, "The differential fate of the members of the Donner Party lends strong support to the argument that females are better able than males to withstand conditions marked by famine and extreme cold." While there is some evidence that **Gender** is associated with **Survival**, explain why the data cannot be used to show that being female *causes* a higher probability of survival.

3. Chp 7, E3 Drug Treatment and Criminal Conviction

A study was conducted to determine if a relationship existed between criminal conviction and years of education. Sixty people who had taken part in a drug rehabilitation program were classified by years of education. Each person was categorized as having a "short" education (15 years or less) or a "long" education (more than 15 years); also recorded was whether or not they had a post-treatment conviction.

```
convict <- read_csv("~/Dropbox/teaching/math150/PracStatCD/Data Sets/Chapter 07/CSV Files/C7 Convict.csv",
                   na="*")
```

- (a) Create a logistic regression model using years of education to estimate the probability of conviction. Interpret the model in terms of the odds ratio.
- (b) Interpret the results of Wald's test and the LRT.
- (c) Conduct Fisher's exact test and a chi-square test of independence using the **Convict** data. How do these tests compare to the logistic regression model?
- (d) No measures of association yet.