

Math 150 - Methods in Biostatistics - Homework 7

your name here

Not Due

```
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, E11 And the Winnder Is...

(Lots to read in your text setting up the problem...)

- (a) Create a logistic regression model using all 17 explanatory variables. Which variables appear to be most significant? > The variables that appear to be the most significant are Actress in a Leading Role, Film Editing, and Adapted Screenplay.

```
oscars1 <- read_csv("~/Dropbox/teaching/math150/PracStatCD/Data Sets/Chapter 07/CSV Files/C7 Oscars.csv",
                  na="*")

oscars2 <- read_csv("~/Dropbox/teaching/math150/PracStatCD/Data Sets/Chapter 07/CSV Files/C7 Oscars2.csv",
                  na="*")

oscars <- full_join(oscars1, oscars2, by="X1")

## full_join: added 0 rows and added 7 columns (Year.x, Year.y, Golden Globe Nominations, Golden Globe W
oscars.glm<-glm(`Best Picture?`~`Actor in a Leading Role`+`Actress in a Leading Role`+
               `Actor in a Supporting Role`+`Actress in a Supporting Role`+`Art Direction`+
               `Cinematography`+`Costume Design`+`Directing`+`Film Editing`+`Makeup`+
               `Original Score`+`Original Song`+`Sound Editing`+`Sound Mixing`+
               `Visual Effects`+`Adapted Screenplay`+`Original Screenplay`, data=oscars,
               family="binomial")
```

- (b) Create and compare multiple logistic regression models. Submit the model with the fewest number of terms that best estimates the probability of winning the Best Picture award.
- (c) Academy Award for Best Picture in 2009 went to *Hurt Locker*. Use your final model in Part B to predict the likelihood that *Hurt Locker* would win the Best Picture award. *Avatar* and *The Blind Side* were also nominated. Use your final model to estimate the probability that each of these movies would win Best Picture.
- (d) Compare two models that you consider to be good using statistics of discordance / concordance. Comment on your results.
- (e) Provide ROC curves (ideally on the same plot) for your two best models. Comment on your graph.

1. Chp 7, E12 And the Winnder Is... 2

Again, there is a good amount to read in your text setting up the problem.

- (a) Create a logistic regression model using all four explanatory variables. Which variables appear to be most significant?
- (b) Using only the `Oscars2` data set, submit the model with the fewest number of terms that best estimates the probability of winning the `Best Picture` award.
- (c) Compare your model in Part (b) of Exercise 11 to the one from Part (b) above. Explain which model is better.
- (d) Combine the `Oscars` and `Oscars2` data sets to include a total of 21 explanatory variables. Create the model with only a few variables that best estimates the probability of winning the `Best Picture` award. Is this new model better than the one created in Part (b) of Exercise 11?