

**Homework Assignment #8**  
**Logit Regression**  
**DATS 6103 - Spring 2020**

**Titanic data**

**Question 1**

With the Titanic dataset, perform some summary statistics.

Variable	Definition	Key
survival	Survival	0 = No, 1 = Yes
pclass	Ticket class	1 = 1st, 2 = 2nd, 3 = 3rd
sex	Gender / Sex	
age	Age in years	
sibsp	# of siblings / spouses on the Titanic	
parch	# of parents / children on the Titanic	
ticket	Ticket number (for superstitious ones)	
fare	Passenger fare	
embarked	Port of Embarkation	C: Cherbourg, Q: Queenstown, S: Southampton

- a) Histogram on age. Maybe a stacked histogram on age with male-female as two series if possible
- b) proportion summary of male-female, survived-dead
- c) pie chart for "Ticket class"
- d) A single visualization chart that shows info of survival, age, pclass, and sex.

**Question 2**

Build a logistic regression model for survival. As we did before, include the features that you find plausible. Make sure categorical variables are use properly. If the coefficient(s) turns out insignificant, drop it and re-build.

**Question 3**

Interpret your result. What are the factors and how do they affect the chance of survival (or the survival odds ratio)? What is the predicted probability of survival for a 30-year-old female with a second class ticket, no siblings, 3 parents/children on the trip? Use whatever variables that are relevant in your model.

**Question 4**

Try three different cut-off values at 0.3, 0.5, and 0.7. What are the

- a) Total accuracy of the model
- b) The precision of the model
- c) The recall rate

## NFL field goal data

We would like to study the chances of making a “field goal kick” successful. If you do not know the game of American football enough to complete this analysis, please check with me or Jason. You should remove the obviously irrelevant columns in the dataset, keeping only the following. (There might still be some irrelevant ones here.)

Variable	Definition	Note
AwayTeam	Name of visiting team	
HomeTeam	Name of home team	
qtr	quarter	1, 2, 3, 4
min	Time: minutes in the game	
sec	Time: seconds in the game	
kickteam	Name of kicking team	
distance	Distance of the kick, from goal post (yards)	
timerem	Time remaining in game (seconds)	
GOOD	Whether the kick is good or no good	If not GOOD:
Missed	If the kick misses the mark	either Missed
Blocked	If the kick is blocked by the defense	or blocked

### Question 5

Build a model overall (not individual team or kicker) to predict the chances of a successful field goal. What variables do you have in your model?

### Question 6

Someone has a feeling that home teams are more relaxed and have a friendly crowd, they should kick better field goals. Can you build two different models, one for all home teams, and one for road teams, of their chances of making a successful field goal?

### Question 7

From what you found, do home teams and road teams have different chances of making a successful field goal? If one does, is that true for all distances, or only with a certain range?