

# DATA 603: Assignment 3

*Deadline: Nov. 29th, 2019 by 4 pm. Submit to Dropbox via D2L.*

## Logistic Regression

**Problem 1.** A study investigated characteristics associated with  $y$ = whether a cancer patient achieved remission (1=yes, 0=no). An important explanatory variable was a labeling index (LI=percentage of “labeled” cells) that measures proliferative activity of cells after a patient receives an injection of tritiated thymidine. Fit a logistic regression model in order to answer the following questions. The data is provided in **remission.scv** file.

- Estimate the probabilities of a cancer patient achieved remission when  $LI=15$  and  $LI=37$ . Comment on your results.
- Interpret the effect of LI in terms of the odds.
- Compute 95% confidence intervals for the logistic regression coefficient ( $\hat{\beta}$ ) and its associated odds. Give an interpretation.
- Use the Wald Z test and Likelihood ratio test to confirm that the LI predictor is associated with remission of cancer at  $\alpha=0.05$ .

**Problem 2.** The RMS Titanic was a British passenger liner that sank in the North Atlantic Ocean in the early morning hours of 15 April 1912, after it collided with an iceberg during its maiden voyage from Southampton to New York City. The RMS Titanic was the largest ship afloat at the time it entered service and was the second of three Olympic-class ocean liners operated by the White Star Line. The Titanic was built by the Harland and Wolff shipyard in Belfast. Thomas Andrews, her architect, died in the disaster. The training-dataset with 891 examples is provided in **titanic.csv** file and the list of Variables with a short description is provided:

survival: Survival

PassengerId: Unique Id of a passenger.

pclass: Ticket class

sex: Sex

Age: Age in years

- Test if the chances of survival of passengers in Titanic depends on those variables at  $\alpha = 0.05$
- From part a), use the likelihood ratio test to check whether the variable age should be in the full model.
- Write the logit and logistic regression model for predicting the chances of survival of passengers in Titanic.
- Interpret the effect of Class and Sex in terms of the odds ratio from the logistic regression model in part c)

- e. Report Deviance, AIC, ROC with AUC (the model fit) from the model in part c)
- f. Build the logistic regression with interaction terms and write the logit model. Use the likelihood ratio test to confirm your result.
- g. Report Deviance, AIC, ROC with AUC (the model fit) from the model in part e) and compare the result with the model in part c).
- h. From the model in part e), predict the probability of survival for a 35 years old man who got the third class ticket. Show your work by substituting the effect values and use R command to confirm your result. Comment on your result.
- i. From the model in part e), predict the probability of survival for a 20 years old man who got the first class ticket. Show your work by substituting the effect values and use R command to confirm your result. Comment on your result.