**Homework Week 4**

**PSC 204B**

Install and load the ‘titanic’ package into your R library using the following code:

install.packages('titanic')

library(titanic)

You will also need to combine the test and training datasets in order to get the complete dataset for this example:

data('titanic\_train')

data('titanic\_test')

titanic\_test$Survived = NA

complete\_data = rbind(titanic\_train, titanic\_test)

And then subsetting the variables we will be using for this homework assignment:

titanic\_dat = titanic\_dat = complete\_data[, c('Survived', 'Sex', 'Age',

'Fare')]

You will use this dataset to answer the following questions.

1. Report the summary statistics for the titanic\_dat dataset (i.e., report means, standard deviations, number of missing values). (Hint: this may be important for a later question on this homework!)
2. Run a logistic regression to predict whether passengers’ survival (Survived [levels: 1 = survived; 0 = perished]) is predicted by cost of fare paid by passengers (Fare). Write out the logistic regression model that you have just fit.
3. Interpret each coefficient from the analysis above in terms of both odds-ratios and probabilities.
4. Plot the observed data with the logistic curve overlaid on the scatterplot.
5. Evaluate the fit of your logistic regression model (hint: see slides 50-54).
6. Imagine that you have just remembered that women and children were given priority in boarding lifeboats. Repeat your logistic regression, this time adding in additional predictors of Age and Sex [levels: male; female]. Include an appropriate interaction term.
7. Interpret all coefficients from the analysis above in terms of both odds-ratios and probabilities.
8. Use the predict() function to determine the odds of your own survival had you been a passenger on the Titanic (you can use the mean value for Fare). Note that predict() returns the predicted values at the linear scale, i.e., on the logit scale.
9. Evaluate whether the multiple logistic regression model is a better fitting model than the simple logistic regression model using fare as the only predictor via a likelihood ratio test. (Hint: Look up the lrtest() function from the ‘lmtest’ package.)

BONUS 1: Create a plot to illustrate the interaction effect found in question 6.