**STT 811**

**In-Class Assignment 8**

This problem will use the OJ dataset.

1. Create a target based on the Purchase field with numerical values of 0 and 1.
   1. OJ$Purchase1 <- OJ$Purchase
   2. OJ$Purchase1[OJ$Purchase1 == “MM”] <- 1
   3. OJ$Purchase1[OJ$Purchase1 == “CH”] <- 0
2. Split the data into training and test datasets (with a 75/25 split).
   1. split\_pct <- 0.75
   2. n <- length(OJ$Purchase1)\*split\_pct # train size
   3. row\_samp <- sample(1:length(OJ$Purchase1), n, replace = FALSE)
   4. train <- OJ[row\_samp,]
   5. test <- OJ[-row\_samp,]
   6. OJ\_train\_mod <- glm(data = train, Purchase1 ~ PriceDiff + LoyalCH, family = binomial)
   7. test\_pred <- predict(OJ\_train\_mod,test, type = "response")
   8. train\_cm <- confusionMatrix(as.factor(as.integer(2\*OJ\_train\_mod$fitted.values)), reference = as.factor(train$Purchase1))
   9. test\_cm <- confusionMatrix(as.factor(as.integer(2\*test\_pred)), reference = as.factor(test$Purchase1))
3. Build a logistic regression model for your target based on PriceDiff and LoyalCH.
   1. How significant are the coefficients?
   2. Compute the confusion matrix for both the train and test datasets. How do they compare?
   3. Test the probability calibration for the training dataset.
4. Create bootstrapped 95% confidence intervals for the 2 coefficients (not Intercept)