**Lab 3: Logistic Regression (Part 1)**

**To submit your work, insert screenshots of your code and outputs (both numeric outputs and graphs) under respective problem prompts. Many steps also require a written answer, and you should insert your written or typed answer below the prompt.**

*Suppose you are investigating allegations of gender discrimination in the hiring practices of a particular firm. An equal-rights group claims that females are less likely to be hired than males with the same background, experience, and other qualifications. You collected data on 28 former applicants. The variables in the dataset include:*

* *HIRE (1 = hired, 0 = not hired)*
* *Years of higher education (EDUC)*
* *Years of work experience (EXP)*
* *GENDER (1 = male, 0 = female).*

1. Download the data file “DISCRIM.csv” from Canvas.
2. Start R or R Studio. Load the “car” package.
3. Import the data into R. Name the imported data **hire.data**. View the data and make sure the data have been imported correctly.
4. Inspect the variables individually and pairwise. Describe your impression.

summary(hire.data)

table(hire.data$HIRE)

table(hire.data$GENDER)

pairs(hire.data)

1. Fit a logistic regression model.

m <- glm(HIRE ~ EDUC + EXP + GENDER, data=hire.data,   
 family=binomial)  
summary(m)

1. Write out the fitted model (original form) with the estimated coefficient values and meaningful variable names.
2. **Interpret model coefficients.** Interpret each of three slope coefficients in at least two ways. Refer to the lecture slides for different ways to interpret slope coefficients.  
   1. Higher education
   2. Work experience
   3. Gender
3. **Model prediction.** Try out the following code and Describe the difference between the two types. Utilize the R manual pages to understand the use of the function.

predict(m, type="link")

predict(m, type="response")

1. **Predict the probability of success of a given case.** Run the following code to predict the probability of being hired for a male candidate who has 6 years of higher education and 3 years of experience.  
     
   predict(m, newdata=data.frame(EDUC=6, EXP=3, GENDER=1),   
    type="response")
2. **Graphing.**

EXP.plot <- seq(0, 12, by=.1)

# GENDER == 0 & EDUC == 4 (Female, Bachelor's degree)

pi <- predict(m, newdata=data.frame(EDUC=4, EXP=EXP.plot,   
 GENDER=0), type="response")

plot(EXP.plot, pi, xlim=c(0, 12), ylim=c(0, 1),

xlab="Years of Working Experience",

ylab="Probability of Being Hired",

type='l', col='red', lty="solid", lwd=3,

cex.axis=1.5, cex.lab=1.5)

# GENDER == 1 & EDUC == 4 (Male, Bachelor's degree)

pi <- predict(m, newdata=data.frame(EDUC=4, EXP=EXP.plot,   
 GENDER=1), type="response")

lines(EXP.plot, pi, col='blue', lty="solid", lwd=3)

# GENDER == 0 & EDUC == 6 (Female, Master's degree)

pi <- predict(m, newdata=data.frame(EDUC=6, EXP=EXP.plot,   
 GENDER=0), type="response")

lines(EXP.plot, pi, col='red', lty="dashed", lwd=3)

# GENDER == 1 & EDUC == 6 (Male, Master's degree)

pi <- predict(m, newdata=data.frame(EDUC=6, EXP=EXP.plot,   
 GENDER=1), type="response")

lines(EXP.plot, pi, col='blue', lty="dashed", lwd=3)

legend(x="topleft", legend=c("Female, Bachelor's",   
 "Male, Bachelor's", "Female, Master's",   
 "Male, Master's"),

col=c("red", "blue", "red", "blue"),   
 lty=c("solid", "solid", "dashed", "dashed"))

1. **Derive the operational form of the logistic regression model from the original form (Slide #13).** You can hand-write it on a piece of paper, take a photo of it and insert it here.