**Regression Models 13 - Quiz 3**

***Toan Xuan Mai***

**Question 1**

Consider the mtcars data set. Fit a model with mpg as the outcome that includes number of cylinders as a factor variable and weight as confounder. Give the adjusted estimate for the expected change in mpg comparing 8 cylinders to 4.

**Answers**

mtcars$cyl <- factor(mtcars$cyl)

mtcars$am <- factor(mtcars$am)

levels(mtcars$am) <- c('-auto', '-manual')

fit <- lm(mpg ~ cyl + wt, mtcars)

summary(fit)$coef[3, 1]

## [1] -6.07086

**Question 2**

Consider the mtcars data set. Fit a model with mpg as the outcome that includes number of cylinders as a factor variable and weight as a possible confounding variable. Compare the effect of 8 versus 4 cylinders on mpg for the adjusted and unadjusted by weight models. Here, adjusted means including the weight variable as a term in the regression model and unadjusted means the model without weight included. What can be said about the effect comparing 8 and 4 cylinders after looking at models with and without weight included?.

**Answers**

fit\_adjusted <- lm(mpg ~ cyl + wt, mtcars)

summary(fit\_adjusted)$coef

## Estimate Std. Error t value Pr(>|t|)

## (Intercept) 33.990794 1.8877934 18.005569 6.257246e-17

## cyl6 -4.255582 1.3860728 -3.070244 4.717834e-03

## cyl8 -6.070860 1.6522878 -3.674214 9.991893e-04

## wt -3.205613 0.7538957 -4.252065 2.130435e-04

fit\_unadjusted <- lm(mpg ~ cyl, mtcars)

summary(fit\_unadjusted)$coef

## Estimate Std. Error t value Pr(>|t|)

## (Intercept) 26.663636 0.9718008 27.437347 2.688358e-22

## cyl6 -6.920779 1.5583482 -4.441099 1.194696e-04

## cyl8 -11.563636 1.2986235 -8.904534 8.568209e-10

*# --> Holding weight constant, cylinder appears to have less of an impact on mpg than if weight is disregarded.*

**Question 3**

Consider the mtcars data set. Fit a model with mpg as the outcome that considers number of cylinders as a factor variable and weight as confounder. Now fit a second model with mpg as the outcome model that considers the interaction between number of cylinders (as a factor variable) and weight. Give the P-value for the likelihood ratio test comparing the two models and suggest a model using 0.05 as a type I error rate significance benchmark.

**Answers**

suppressMessages(**library**(lmtest))

fit\_non\_interaction <- lm(mpg ~ cyl + wt, mtcars)

summary(fit\_non\_interaction)$adj.r.squared

## [1] 0.8200146

fit\_interaction <- lm(mpg ~ cyl + wt + cyl:wt, mtcars)

summary(fit\_interaction)$adj.r.squared

## [1] 0.8349382

lrtest(fit\_interaction, fit\_non\_interaction)

## Likelihood ratio test

##

## Model 1: mpg ~ cyl + wt + cyl:wt

## Model 2: mpg ~ cyl + wt

## #Df LogLik Df Chisq Pr(>Chisq)

## 1 7 -70.741

## 2 5 -73.311 -2 5.1412 0.07649 .

## ---

## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

*# --> The P-value is larger than 0.05. So, according to our criterion, we would fail to reject, which suggests that the interaction terms may not be necessary.*

**Question 4**

Consider the mtcars data set. Fit a model with mpg as the outcome that includes number of cylinders as a factor variable and weight inlcuded in the model as

lm(mpg ~ I(wt \* 0.5) + factor(cyl), data = mtcars)

How is the wt coefficient interpretted?

**Answers**

*# The estimated expected change in MPG per one ton increase in weight for a specific number of cylinders (4, 6, 8).*

**Question 5**

Consider the following data set

x <- c(0.586, 0.166, -0.042, -0.614, 11.72)

y <- c(0.549, -0.026, -0.127, -0.751, 1.344)

Give the hat diagonal for the most influential point

**Answers**

x <- c(0.586, 0.166, -0.042, -0.614, 11.72)

y <- c(0.549, -0.026, -0.127, -0.751, 1.344)

fit <- lm(y ~ x)

max(hatvalues(fit))

## [1] 0.9945734

**Question 6**

Consider the following data set

x <- c(0.586, 0.166, -0.042, -0.614, 11.72)

y <- c(0.549, -0.026, -0.127, -0.751, 1.344)

Give the slope dfbeta for the point with the highest hat value.

**Answers**

x <- c(0.586, 0.166, -0.042, -0.614, 11.72)

y <- c(0.549, -0.026, -0.127, -0.751, 1.344)

fit <- lm(y ~ x)

influence.measures(fit)$infmat[5, 'dfb.x']

## [1] -133.8226

**Question 7**

Consider a regression relationship between Y and X with and without adjustment for a third variable Z. Which of the following is true about comparing the regression coefficient between Y and X with and without adjustment for Z.

**Answers**

*# It is possible for the coefficient to reverse sign after adjustment. For example, it can be strongly significant and positive before adjustment and strongly significant and negative after adjustment.*