Coursera Regression Models Quiz 3

Cheng-Han Yu

August 11, 2015

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.

Solution:

```
data(mtcars)
fit <- lm(mpg ~ factor(cyl) + wt, data = mtcars)
summary(fit)$coefficient
##                Estimate Std. Error   t value     Pr(>|t|)
## (Intercept)  33.990794  1.8877934 18.005569 6.257246e-17
## factor(cyl)6 -4.255582  1.3860728 -3.070244 4.717834e-03
## factor(cyl)8 -6.070860  1.6522878 -3.674214 9.991893e-04
## wt           -3.205613  0.7538957 -4.252065 2.130435e-04
```

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?.

Solution:

```
fit1 <- lm(mpg ~ as.factor(cyl), data = mtcars)
summary(fit1)$coef[3]
## [1] -11.56364
summary(fit)$coef[3]
## [1] -6.07086
```

Note that 11.564 > 6.071, and so 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.

Solution:

```
fit_inter <- lm(mpg ~ factor(cyl) * wt, data = mtcars)
anova(fit, fit_inter, test = "Chisq")
## Analysis of Variance Table
##
## Model 1: mpg ~ factor(cyl) + wt
## Model 2: mpg ~ factor(cyl) * wt
##   Res.Df    RSS Df Sum of Sq Pr(>Chi)
## 1     28 183.06
## 2     26 155.89  2     27.17   0.1038
```

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?

Solution:

Since the unit of `(wt * 0.5)` is (lb/2000), and one (short) ton is 2000 lbs, the wt coefficient is interpretted as the estimated expected change in MPG per one ton increase in weight for a specific number of cylinders (4, 6, 8).

```
fit4 <- lm(mpg ~ I(wt * 0.5) + factor(cyl), data = mtcars)
summary(fit4)$coefficient
##                  Estimate Std. Error   t value     Pr(>|t|)
## (Intercept)     33.990794   1.887793 18.005569 6.257246e-17
## I(wt * 0.5)     -6.411227   1.507791 -4.252065 2.130435e-04
## factor(cyl)6 -4.255582   1.386073 -3.070244 4.717834e-03
## factor(cyl)8 -6.070860   1.652288 -3.674214 9.991893e-04
```

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.

Solution:

```
x <- c(0.586, 0.166, -0.042, -0.614, 11.72)
y <- c(0.549, -0.026, -0.127, -0.751, 1.344)
fit5 <- lm(y ~ x)
```

```
hatvalues(fit5)
##         1         2         3         4         5
## 0.2286650 0.2438146 0.2525027 0.2804443 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.

Solution:

```
x <- c(0.586, 0.166, -0.042, -0.614, 11.72)
y <- c(0.549, -0.026, -0.127, -0.751, 1.344)
fit6 <- lm(y ~ x)
dfbetas(fit6)[, 2]
##            1            2            3            4            5
##  -0.37811633  -0.02861769   0.00791512   0.67253246 -133.82261293
```

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.

Solution:

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.