Valeria Duran

Math 6357

Homework 1

1. I disagree with the model the student wrote. The simple linear regression model is written as: $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$. ε_i is a random error term with mean E$\{\varepsilon_i\}$=0; therefore, E$\{Y_i\}$ = E$\{\beta_0 + \beta_1 X_i + \varepsilon_i\} = \beta_0 + \beta_1 X_i + E\{\varepsilon_i\} = \beta_0 + \beta_1 X_i$, thus, making the mean of the probability distribution: $E\{Y_i\} = \beta_0 + \beta_1 X_i$

2.

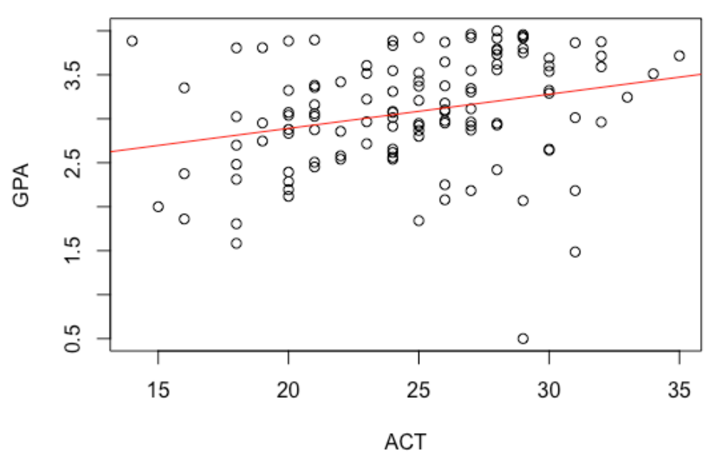
```
> CH01PR19 <- read.table("~/Downloads/CH01PR19.txt", quote="\"", comment.char="")
>  View(CH01PR19)
> Grades <- data.frame(CH01PR19)
> colnames(Grades) <- c("GPA", "ACT")
> lin.reg <- lm(Grades$GPA ~ Grades$ACT)
> lin.reg

Call:
lm(formula = Grades$GPA ~ Grades$ACT)

Coefficients:
(Intercept)   Grades$ACT
    2.11405      0.03883
```

a) The least squares estimate of β_0= 2.11405 and β_1= 0.03883. The estimated regression function is: $Y = 2.11405 + 0.03883X$

b) > plot(Grades$GPA~Grades$ACT, xlab= "ACT", ylab="GPA")
> abline(lin.reg, col="red")



Since the data is so spread out from the estimated regression function, the function does not fit the data well. This is due to a lot of variance in the data.

c) > Y = lin.reg$coefficients[[1]] + lin.reg$coefficients[[2]]*30
> Y
[1] 3.278863

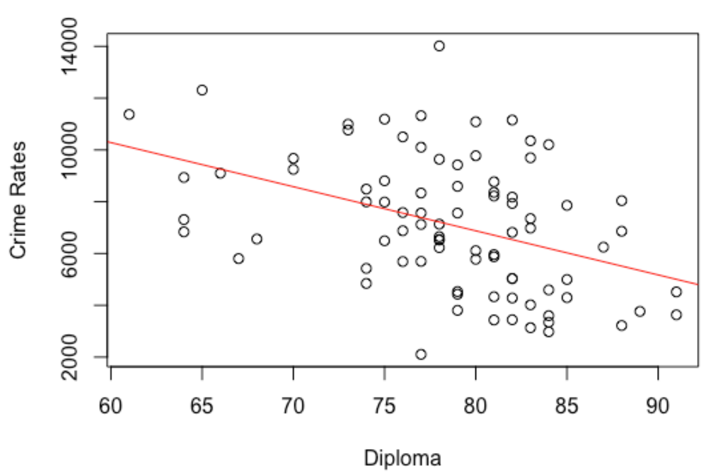
d) When the entrance test scores increases by one point, the mean estimate response increases by 0.03883 since β_1 is the slope of the estimated regression line and indicates the change of the mean response when X increases by one point.

3.
```
> CH01PR28 <- read.table("~/Downloads/CH01PR28.txt", quote="\"", comment.char="")
>   View(CH01PR28)
> Crimes <- CH01PR28
> colnames(Crimes) <- c("CrimeRates", "Diploma")
> crime.reg <- lm(Crimes$CrimeRates ~ Crimes$Diploma)
> crime.reg

Call:
lm(formula = Crimes$CrimeRates ~ Crimes$Diploma)

Coefficients:
  (Intercept)  Crimes$Diploma
     20517.6          -170.6
```

a) The estimated regression function is: $Y = 20517.6 - 170.6X$
> plot(Crimes$CrimeRates~Crimes$Diploma, xlab="Diploma", ylab = "Crime Rates")
> abline(crime.reg, col="red")



The function seems to represent the data fairly well. Although the data spreads out a bit towards the center of the graph, I believe most of the data points fall near the regression function.

b) i. The difference in the mean crime rate for two counties whose high school graduation rates differ by one percentage point is just the slope of the regression function, which is -170.6 (β_1).

ii. > crime.reg$coefficients[[1]] + crime.reg$coefficients[[2]]*80
[1] 6871.585

iii. $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$;
$Y_{10} = 7932 = 20517.6 - 170.69(82) + \varepsilon_{10}$
$\varepsilon_{10} = 7932 - 6530.434 = 1401.566$

iv.
> (summary(crime.reg)$sigma)^2
[1] 5552112
MSE = 5552112

4. $\widehat{B_0} = \bar{Y} - \widehat{B_1}\bar{X}$
$\widehat{B_0} = \beta_0 + \beta_1\bar{X} + \varepsilon - \widehat{B_1}\bar{X}$ since $\bar{Y} = \beta_0 + \beta_1\bar{X} + \varepsilon$
$= \beta_0 + (\beta_1 - \widehat{B_1})\bar{X} + \varepsilon$
E$[\widehat{B_0}] = E[\beta_0] + E[(\beta_1 - \widehat{B_1})\bar{X}] + E[\varepsilon]$
$= \beta_0 + \bar{X} * E[(\beta_1 - \widehat{B_1})] + E[\varepsilon]$ since β_0 is a constant
$= \beta_0 + \bar{X} * E[(\beta_1 - \widehat{B_1})]$ since $E[\varepsilon] = 0$
$= \beta_0 + \bar{X} * [E(\beta_1) - E(\widehat{B_1})]$
$= \beta_0 + \bar{X} * (\beta_1 - \beta_1)$ since E$[\beta_1] = \beta_1$ and $E\left[\widehat{B_1}\right] = \beta_1$
E$[\widehat{B_0}] = \beta_0$