# Lab 6

*Maria Barouti*

*2/11/2020*

## Exercise 1

The director of admissions of a small college selected 120 students at random from the new freshman class in a study to determine whether a student's grade point average (GPA) at the end of the freshman year (Y) can be predicted from the ACT test score (X). The full data set is available on our course blackboard site. To read text (ASCII) file, you can use an R command `read.table("CH01PR19.txt")`

1. Obtain the least squares estimates of $\beta_0$ and $\beta_1$ and state the estimated regression function.

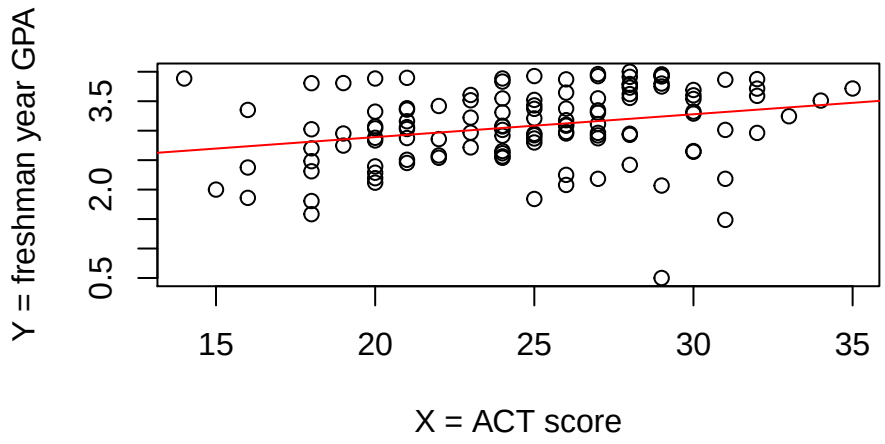
```
GPA = read.table("CH01PR19.txt")
attach(GPA)
X <- V2
Y <- V1
```

2. Plot the estimated regression function and the data. Does the estimated regression function appear to fit the data well?

```
reg = lm(Y ~ X)
reg
```

```
##
## Call:
## lm(formula = Y ~ X)
##
## Coefficients:
## (Intercept)            X
##     2.11405      0.03883
```

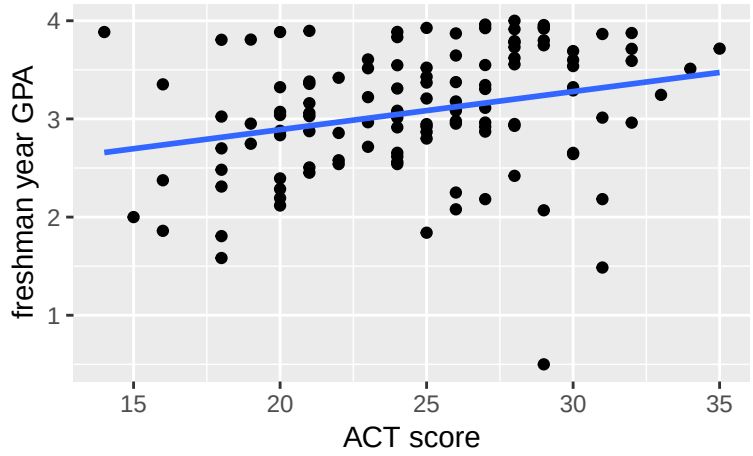
```
plot(X,Y,xlab="X = ACT score",ylab="Y = freshman year GPA")
abline(reg,col="red")
```



```
# Another way to plot
suppressPackageStartupMessages(library(tidyverse))
```

```
## Warning: As of rlang 0.4.0, dplyr must be at least version 0.8.0.
## x dplyr 0.7.6 is too old for rlang 0.4.1.
## i Please update dplyr to the latest version.
## i Updating packages on Windows requires precautions:
##   <https://github.com/jennybc/what-they-forgot/issues/62>
```

```
ggplot(GPA, mapping = aes(x = V2, y = V1)) +
        geom_point() +
        geom_smooth(se = FALSE, method = lm) +
        xlab("ACT score") +
        ylab("freshman year GPA")
```




There is a lot of variation of responses around the regression line. A positive trend is apparent, and so, regression explains captures the trend, but it cannot produce an accurate prediction of the GPA. Points are relatively far from the regression line, so no, it does not fit the data well.

3. Obtain a point estimate of the mean freshman GPA for students with ACT test score $X = 30$.

```
predict(reg, data.frame(X=30))
```

```
##        1
## 3.278863
```

4. What is the point estimate of the change in the mean response when the entrance test score increases by one point?

```
reg$coefficients[2]
```

```
##          X
## 0.03882713
```

5. Obtain the residuals $e_i$ and the sum of the squared residuals $\sum e_i^2$.

```
e = Y - fitted.values(reg)
sum(e^2)
```

```
## [1] 45.81761
```

6. Obtain point estimates of $\sigma^2$ and $\sigma$. In what units is each of them expressed?

```r
n <- length(X)
var_est <- sum(e^2)/(n-2)
var_est #measured in squared units of the GPA
```

```
## [1] 0.3882848
```

```r
sqrt(var_est) #measured in the original units of the GPA
```

```
## [1] 0.623125
```

```r
summary(reg) #A direct way of estimating Std(Y)
```

```
##
## Call:
## lm(formula = Y ~ X)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.74004 -0.33827  0.04062  0.44064  1.22737
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.11405    0.32089   6.588  1.3e-09 ***
## X            0.03883    0.01277   3.040  0.00292 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6231 on 118 degrees of freedom
## Multiple R-squared:  0.07262,    Adjusted R-squared:  0.06476
## F-statistic:  9.24 on 1 and 118 DF,  p-value: 0.002917
```

## Exercise 2

1. Obtain a 99% confidence interval for $\beta_1$. Does it include zero? Why might the director of admissions be interested in whether the confidence interval includes zero?

```r
# A 99% confidence interval for the slope
confint(reg,level=0.99)
```

```
##                   0.5 %     99.5 %
## (Intercept) 1.273902675 2.95419590
## X           0.005385614 0.07226864
```

It does not include 0, and therefore, at the 1% level of significance, the slope $\beta_1$ is found significant. So, the director of admissions will conclude that the ACT score is an important variable predicting success of students (their GPA) during their freshman year.

2. Test whether or not a linear association exists between student's ACT score (X) and GPA at the end of the freshman year (Y). Use a level of significance of 0.01. State the alternatives, decision rule, and conclusion.

As noted above, $H_0$ is rejected at the 1% level. We conclude that there is significant evidence of a linear association between the ACT score and the freshman GPA.

3. Obtain a 95 percent confidence interval for the mean freshman GPA for students whose ACT test score is 28. Interpret your confidence interval.

```r
# Confidence interval for the mean response
predict(reg, data.frame(X=28), interval="confidence")
```

```
##        fit      lwr      upr
## 1 3.201209 3.061384 3.341033
```

In a long run of samples, 95% of confidence intervals constructed this way will contain the actual population mean response $\mu(28) = E\{Y|X = 28\}$.

4. A student obtained a score of 28 on the ACT. Predict her freshman GPA using a 95 percent prediction interval. Interpret your prediction interval.

```r
# Prediction interval for the individual response
predict(reg, data.frame(X=28), interval="prediction")
```

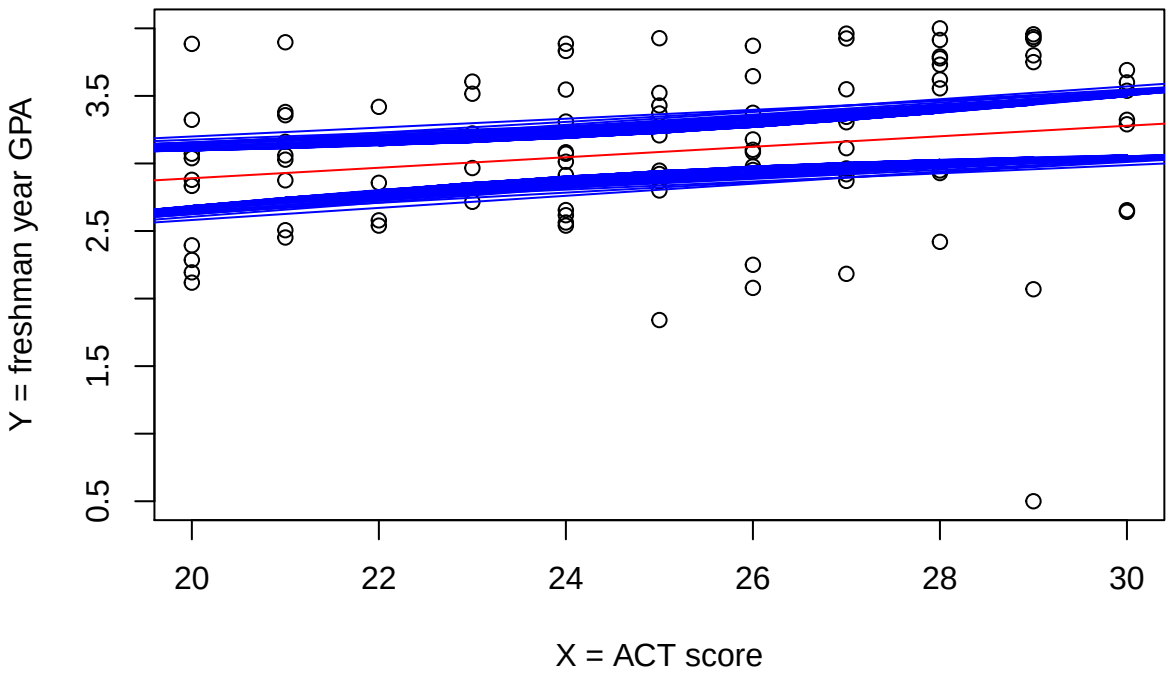
```
##        fit      lwr      upr
## 1 3.201209 1.959355 4.443063
```

In a long run of samples and students with ACT=28, 95% of prediction intervals constructed this way will contain the actual response $Y$.

5. On the same graph, plot the data, the least squares regression line for ACT scores, the 95 percent confidence band for the true regression line for ACT scores between 20 and 30. Does the confidence band suggest that the true regression relation has been precisely estimated? Discuss.

```r
# Confidence band for the entire population regression line
n <- length(X)
W <- sqrt(2*qf(0.95,2,n-2))
Sxx <- (n-1)*var(X)
e <- reg$residuals
s <- sqrt( sum(e^2)/(n-2) )
margin <- W*s*sqrt(1/n + (X - mean(X))^2/Sxx)
upper.band <- predict(reg) + margin
lower.band <- predict(reg) - margin

# Plots
plot(X,Y,xlab="X = ACT score",ylab="Y = freshman year GPA",xlim=c(20,30))
abline(reg,col="red")
lines(X,upper.band,col="blue")
lines(X,lower.band,col="blue")
```


Y = freshman year GPA
0.5
1.5
2.5
3.5
20
22
24
26
28
30
X = ACT score