GU4205 - HW1 Problem 3

Mathieu Sauterey - UNI: mjs2364

September 21, 2017

a) Obtain least squares estimates and state estimated regression function

First, we import the data using read.table() function and then we use the lm function to compute a linear regression model named “lm1” from the data.

```
data <- read.table("Data_HW1_Problem3.txt", header = FALSE, as.is =TRUE)
names(data) <- c("GPA","ACT")
head(data)
```

```
##     GPA ACT
## 1 3.897  21
## 2 3.885  14
## 3 3.778  28
## 4 2.540  22
## 5 3.028  21
## 6 3.865  31
```

```
lm1 <- lm(data$GPA ~ data$ACT)
```

The estimated regression coefficients β_0 and β_1 are computed by the function lm. We retrive them in the list lm1 and respectively print them below:

```
beta_0 <- lm1$coefficients[1]
beta_1 <- lm1$coefficients[2]
names(lm1$coefficients) <- c("Intercept","Slope")
lm1$coefficients[1]
```

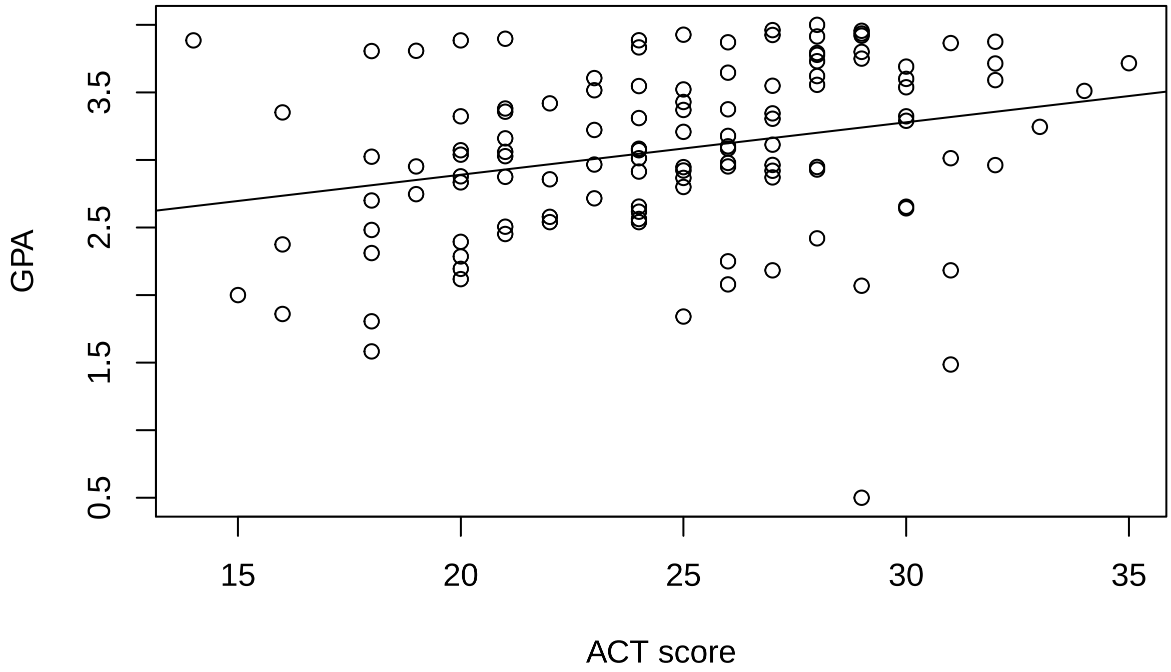
```
## Intercept
##  2.114049
```

```
lm1$coefficients[2]
```

```
##      Slope
## 0.03882713
```

b) Plot the estimated regression function and the data

```
plot(data$ACT,data$GPA,xlab="ACT score",ylab="GPA")
abline(lm1$coefficients[1],lm1$coefficients[2])
```



The regression line of the estimated regression function appear to fit the data quite well.

c) Obtain a point estimate of the mean freshman GPA for students with ACT score X = 30

Using the formula

$$\hat{Y_i} = \beta_0 + \beta_1 * X_i$$

we can compute the point estimate for X=30.

```
Y_30 <- lm1$coefficients[1]+lm1$coefficients[2]*30
names(Y_30)="Estimated Y for X=30"
Y_30
```

```
## Estimated Y for X=30
##             3.278863
```

d) What is the point estimate of the change in the mean response when the entrance test score increases by 1 point?

Using

$$\beta_1 * X_i$$

we can compute the point estimate for $X = 1$. This will tell us what the increment of the GPA mean response is when SAT score increases by 1 point.

```
Y_1 <- lm1$coefficients[2]*1
names(Y_1)="Estimated Y for increment of 1 SAT point"
Y_1
```

```
## Estimated Y for increment of 1 SAT point
##                                0.03882713
```

e) Obtain the residuals e_i. Do they sum to zero in accord with property 1 in Problem 1?

Per property 1 below, the sum of the residuals of a regression model must sum up to zero. To obtain the residuals, we call and sum the residuals element e_i in the list lm1.

$$\sum_{i=1}^{N} e_i = 0$$

```
sum(lm1$residuals)
```

```
## [1] -2.942091e-15
```

While not exactly equal to zero due to rounding errors, the sum of residuals is extremely small (to the order of E-15) and can reasonably be considered equal to 0.

f) Estimate σ and σ^2. In what units is σ expressed?

Estimate σ and σ^2

We can estimate the variance σ^2 using the unbiased estimator MSE (Mean Square Error) or s^2. The sum of squared residuals can be obtained similarly to e) and the degrees of freedom n can be obtained by calling the df.residuals element of the lm1 list.

$$s^2 = \frac{\sum_{i=1}^{N} (e_i^2)}{n-2}$$

```
sigma2 <- sum(lm1$residuals^2)/lm1$df.residual
sigma2
```

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

Finally, the standard deviation σ can be estimated by calculating s the positive square root of s^2.

```
sigma <- sqrt(sigma2)
sigma
```

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

In what units is σ expressed?

We know that σ^2 is estimated by the below equation that we rewrite to show the estimated mean response $\hat{Y}_i$ and dependent variable Y_i:

$$s^2 = \frac{\sum_{i=1}^{N} (e_i^2)}{n-2} = \frac{\sum_{i=1}^{N} ((Y_i - \hat{Y}_i)^2)}{n-2}$$

We then consider the square root of this expression in order to obtain that of s:

$$s = \sqrt{\frac{\sum_{i=1}^{N} (e_i^2)}{n-2}} = \sqrt{\frac{\sum_{i=1}^{N} ((Y_i - \hat{Y}_i)^2)}{n-2}}$$

In order to determine the unit of σ we can analyze the unit of its estimator s. Looking at the leg furthest right in the equation above, we conclude that the unit will be contained in the numerator since the denominator only contains the unitless degrees of freedom n. The sum $\sum$ has no influcence on units, and square root and squared numerator cancel out so we are left looking at $Y_i - \hat{Y}_i$. We know that the observed value Y_i and its fitted value $\hat{Y}_i$ have the same unit because the lattest is an estimate of the former. Given that Y_i was given as the GPA, we conclude that σ has unit [GPA point].