MA 4710 Homework 2

Benjamin Hendrick

Wednesday, January 27, 2016

Problem 1.18

The sum of the normal errors ϵ_i would equal zero. By the central limit theorem, the mean of the normally distributed errors is 0. If the mean of the errors is computed $\frac{1}{n}\sum\epsilon_i$, then the sum of the errors $\sum\epsilon_i$ must equal zero because n cannot be zero.

Problem 1.19

Load the data.

```
CH01PR19 <- read.table("~/GitHub/MA-4710/Homework 2/CH01PR19.txt",
                       quote="\"", comment.char="")
```

Part A

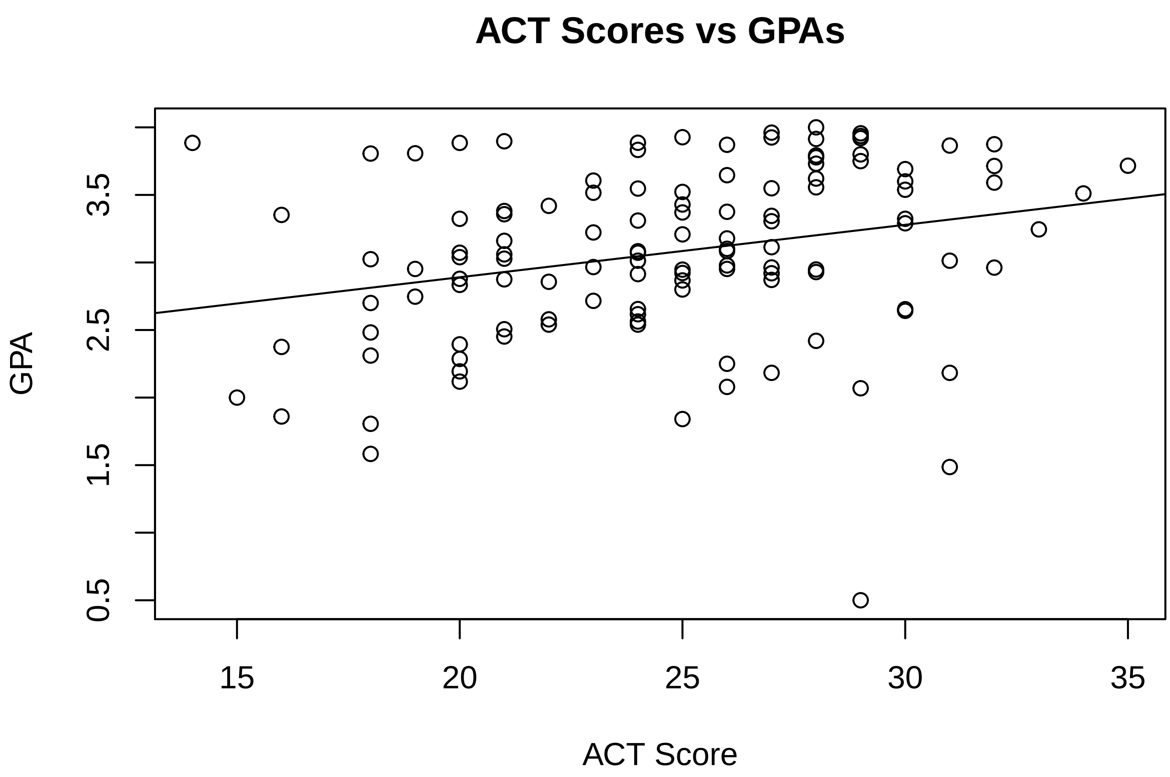
```
model <- lm(V1~V2, data = CH01PR19)
model
```

```
##
## Call:
## lm(formula = V1 ~ V2, data = CH01PR19)
##
## Coefficients:
## (Intercept)           V2
##     2.11405      0.03883
```

The least square estimate for β_0 is 2.11405. The least square estimate for β_1 is 0.3883. Therefore the estimated regression function is $\hat{Y}_i = 2.11405 + 0.03883 \times X_i + \epsilon_i$.

Part B

```
plot(x = CH01PR19$V2,
     y = CH01PR19$V1,
     xlab = "ACT Score",
     ylab = "GPA",
     main = "ACT Scores vs GPAs")
abline(model)
```



The regression appears to fit the data. The data is not very correlated and the variance of the response variable appears to be quite large.

Part C

The point estimate where $X = 30$ is expressed as $Y_{30} = 2.11405 + 0.03883(30) = 3.2794$

Part D

The point estimate of the change in the mean response when the entrance test score increases by one point is $\beta_1 = 0.03883$.

Problem 1.23

We will use the data set `CH01PR19` from problem 1.19.

Part A

```
model <- lm(V1~V2, data = CH01PR19)
resids <- residuals(model)
sum(resids)
```

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

The sum of the residuals is approximately zero. By a rounding error, we can say that the sum of the residuals is zero.

Part B

We will find the variance σ^2 and standard deviation σ for the responce variable, GPA. In the case of the data set `CH01PR19`, the responce is `V1`.

```
var(CH01PR19$V1)
```

```
## [1] 0.4151719
```

```
sd(CH01PR19$V1)
```

```
## [1] 0.6443383
```

The variance is $\sigma^2 = 0.4151719$. The stand deviation $\sigma = 0.6443383$. The standard deviation is expressed in the same units as the response variable, GPA.

Problem 2.24

We will use the data set `CH01PR19` from problem 1.19.

Part A

Use `R` to create the linear model and confidence interval.

```
model <- lm(V1~V2, data = CH01PR19)
confint(model, level=0.99)
```

```
##                   0.5 %     99.5 %
## (Intercept) 1.273902675 2.95419590
## V2          0.005385614 0.07226864
```

From the output above, we see that the 99% confidence interval for β_1 is between 0.005385614 and 0.07226864. We can bet 99% confident that the expected GPA increase between 0.005385614 and 0.07226864 for each unit increase in the ACT score.

Zero is not in the confidence interval. If zero was included in the confidence interval, the director of admissions could assume (with 99% confidence) that there is not positive linear increase in GPA with respect to ACT score.

Part B

Establish the hypothesis.

$$H_0 : \beta_1 = 0$$
$$H_1 : \beta_1 \neq 0$$

Calculate t^* using the linear regression model `model` from part A. Set $\beta_1 = 0$ under H_0.

```
beta1=0   # under H0
tstar=(summary(model)$coefficients[2,1]-beta1)/summary(model)$coefficients[2,2]
tstar
```

```
## [1] 3.039777
```

Calculate the degrees of freedom.

```
df=length(CH01PR19$V1)-2
df
```

```
## [1] 118
```

Find the t-value with two sided 0.01 significance.

```
abs(qt(0.01/2, df = length(CH01PR19$V1)-2))
```

```
## [1] 2.618137
```

Because t^* is greater than the critical value, we reject the null hypothesis that $\beta_1 = 0$.

Part C

```
pvalue=2*(1-pt(tstar,df))
pvalue
```

```
## [1] 0.002916604
```

The p-value from part B is 0.002916604. This is less than 0.01, which supports the rejection of the null hypothesis.

Problem 2.7

Load the data

```
CH01PR22 <- read.table("~/GitHub/MA-4710/Homework 2/CH01PR22.txt", quote="\"", comment.char="")
```

Part A

```
model <- lm(V1~V2, data = CH01PR22)
model
```

```
##
## Call:
## lm(formula = V1 ~ V2, data = CH01PR22)
##
## Coefficients:
## (Intercept)           V2
##     168.600        2.034
```

```r
confint(model, level=0.99)
```

```
##                  0.5 %     99.5 %
## (Intercept) 160.690457 176.509543
## V2            1.765287   2.303463
```

The mean hardness when the elapsed time increases by one hour is 2.034 Brinell units. The 99% confidence interval is between 1.765287 and 2.303463. This means that we can be 99% confident that the increase in hardness over an hour will be between 1.765 and 2.303.

Part B

State the hypothesis:

$$H_0 : \beta_1 = 2$$
$$H_1 : \beta_1 \neq 2$$

```r
t.test(CH01PR22$V1, alternative = "two.sided", mu = 2, conf.level = 0.99)
```

```
##
##  One Sample t-test
##
## data:  CH01PR22$V1
## t = 46.941, df = 15, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 2
## 99 percent confidence interval:
##  211.5283 239.5967
## sample estimates:
## mean of x
##  225.5625
```

We will determine the test if the p-value is less than the 0.01. In this test, the p-value is 2.2e-16 which is much less than 0.01. Therefore, we reject the null hypothesis. This means that the mean hardness does not increase by 2 Brinell units per hour.