Asn2 Q2

Sumanth Donthula

2022-09-20

2.a.

The values of Beta0 and Beta1 are 2.114,0.038

The model is #Y=2.11+0.03883X

```
Data =read.delim("AS_2_Q_2_Data.txt", header = FALSE, sep = " ")
Y=Data[,2]
X=Data[,6]

Model=lm(Y~X)

Beta0=as.numeric(Model$coefficients["(Intercept)"])
Beta0
```

```
## [1] 2.114049
```

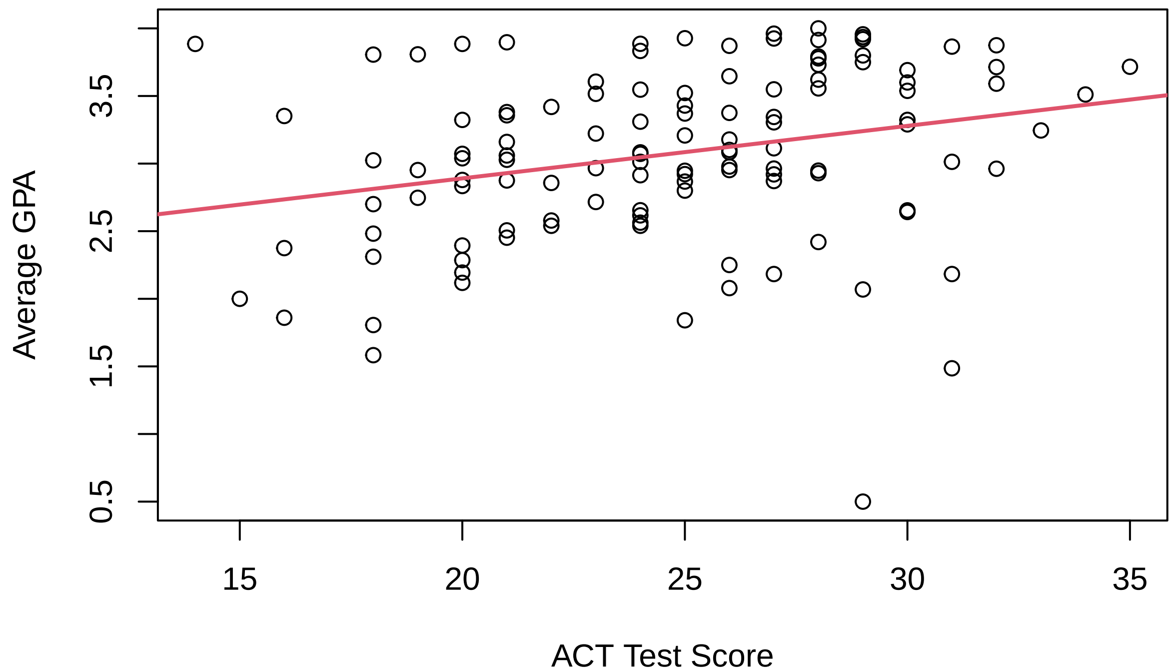
```
Beta1=as.numeric(Model$coefficients["X"])
Beta1
```

```
## [1] 0.03882713
```

2.b.

The plot of the Data and the Model Fit is as shown. The regression function fit the model well with minimum outliers.

```
plot(X,Y,xlab="ACT Test Score",ylab="Average GPA")
abline(Beta0,Beta1,col=2,lwd=2)
```



2.c.

The point estimate of mean freshman GPA of ACT score=30 is 3.2788

```
Y1=Beta0+Beta1*30
Y1
```

```
## [1] 3.278863
```

2.d.

The point estimate of change in mean response when entrance score is increased by one point is Beta1 which is 0.0388

Problem 3

3.a.

The 95 % confidence interval of test score which is 28 is (3.06,3.34)

```
xnew = data.frame(X=28)

Y2=Beta0+Beta1*28

SXY = sum(X *Y ) - length(X) * mean(X) * mean(Y)
SYY = sum( Y * Y) - length(Y) * (mean(Y))^2
SXX = sum( X * X) - length(X) * (mean(X))^2
RSS = SYY - Beta1 * SXY
SIGMA=(RSS/(length(X)-2))^0.5
```

```
mu=mean(X)

n=0.05
t=qt(n/2, length(X)-2, lower.tail=TRUE)
#Method 1 from Function
confInt = predict(Model,xnew, interval = "confidence", level = 0.95, se.fit = TRUE)
conflwr=confInt$fit[2]
confupr=confInt$fit[3]
conflwr
```

```
## [1] 3.061384
```

```
confupr
```

```
## [1] 3.341033
```

```
#Method 2 from Formula
ConfUb=Y2+t*SIGMA*((1/length(X))+((xnew-mu)^2/SXX))^0.5
ConfLb=Y2-t*SIGMA*((1/length(X))+((xnew-mu)^2/SXX))^0.5
ConfUb
```

```
##          X
## 1 3.061384
```

```
ConfLb
```

```
##          X
## 1 3.341033
```

3.b.

The 95 % prediction interval of test score which is 28 is (1.95,4.44)

```
#Method 1 from Function
predInt=predict(Model,xnew, interval = "prediction", level = 0.95, se.fit = TRUE)
predlwr=predInt$fit[2]
predlwr
```

```
## [1] 1.959355
```

```
predupr=predInt$fit[3]
predupr
```

```
## [1] 4.443063
```

```
#Method 2 Derivating with Formula
PredUb=Y2+t*SIGMA*(1+(1/length(X))+((xnew-mu)^2/SXX))^0.5
PredLb=Y2-t*SIGMA*(1+(1/length(X))+((xnew-mu)^2/SXX))^0.5
PredUb
```

```
##          X
## 1 1.959355
```

```r
PredLb
```

```
##          X
## 1 4.443063
```

3.c.

Yes the Prediction interval is wider than confidence interval. Because the prediction model is used to find estimates of the random samples rather than the confidence interval which is an inference on sample data.

3.d.

The range of confidence band is (3.02,3.37).

Yes the confidence band is little wider because it represents the entire regression model line not only the sample at Xh=28.

```r
Weight = sqrt(2*qf(0.95,2,length(X)-2))

cbandupper = confInt$fit[,1]+Weight*confInt$se.fit
cbandlower = confInt$fit[,1]-Weight*confInt$se.fit
cbandlower
```

```
## [1] 3.026159
```

```r
cbandupper
```

```
## [1] 3.376258
```

4.a.

Anova Table

```r
Anova =anova(Model)
Anova
```

```
## Analysis of Variance Table
##
## Response: Y
##            Df Sum Sq Mean Sq F value   Pr(>F)
## X           1  3.588  3.5878  9.2402 0.002917 **
## Residuals 118 45.818  0.3883
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

4.b.

MSR is the sum of squares due to regression taking degrees of freedom into account.

MSE is Mean Squared Error which is not a biased estimator of Standard Deviation (Variance Squared).

When Beta1=0 or the slope of the regression equation is 0.

4.c.

Let's Start Null Hypothesis Testing

Assumption: H0: Beta1 is zero Beta1=0

Ha: Beta1 is not Zero Beta1!=0

Decision rule: The value of Fscore is less than FStarScore so we reject the null hypothesis so the Alternative is true which is Beta1 is not equal to 0.

Conclusion: Beta1 is not Zero Beta1!=0

```
alpha=0.01
Fscore = qf((1-alpha),1,length(X)-2)
Fscore
```

```
## [1] 6.854641
```

```
FStarScore=9.242 #From Anova Table
FStarScore
```

```
## [1] 9.242
```

4.d.

The absolute magnitude of reduction in variance of Y when X is introduced in regression model is R Squared(R^2)

4.e.

The sign of R is positive as the data has a positive correlation from the graph.

```
R=sqrt(0.07262) # from summary
R
```

```
## [1] 0.269481
```

4.f.

R Squared gives clear cut interpretation. It describes the percent of variance of Y with respect to X. R square is usually used to represent the relation. It takes values from 0 to 1.