Assignment 1 Rmarkdown

Jonathan Wai

2019-01-25

Read Table in R

```
Dat=read.table("GPA.txt", header=F)
names(Dat)<-c("GPA","ACT")
head(Dat)
```

```
##     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
```

Assign name

```
GPA=c(Dat$GPA)
ACT=c(Dat$ACT)
```

(a) Mean and Variance of GPA

```
summary(GPA)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.500   2.689   3.078   3.074   3.593   4.000
```

Mean of GPA is 3.074

```
var(GPA)
```

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

Variance of GPA is 0.4151719

(a) Mean and Variance of ACT

```
summary(ACT)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   14.00   21.00   25.00   24.73   28.00   35.00
```

Mean of ACT is 24.73

```
var(ACT)
```

```
## [1] 19.99937
```

Variance of ACT is 19.99937

(b) Correlation between ACT score and GPA.

```
cor.test(ACT,GPA)
```

```
##
##  Pearson's product-moment correlation
##
## data:  ACT and GPA
## t = 3.0398, df = 118, p-value = 0.002917
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.09482051 0.42804747
## sample estimates:
##       cor
## 0.2694818
```

Correlation between ACT and GPA is 0.2694818. Strong evidence of a a linear relationship.

A weak positive linear relationship between the ACT and GPA

(c) Fit a simple linear regression using ACT score as the explanatory variable, and GPA as the response variable.

```
GPA.lm <- lm(GPA ~ ACT, data=Dat)
```

(d) What is the estimated intercept and slope of the regression line?

```
GPA.lm
```

```
##
## Call:
## lm(formula = GPA ~ ACT, data = Dat)
##
## Coefficients:
## (Intercept)          ACT
##     2.11405      0.03883
```

Estimated intercept is 2.11405

Slope of regression line is 0.03883

(e) Write in words the interpretation of the slope.

For each 1 point increase in ACT score, GPA score increases by 0.03883

(f) What is the standard deviation around the regression line, i.e. estimate population variance ?

Residual standard error (RSE): 0.6231

Residual sum of squares (RSS): 0.6231^2 * 118 = rse^2 * df= 45.81393

(g) Use a t-test to determine whether or not there is a linear relationship between ACT score and GPA.

```
summary(GPA.lm)
```

```
##
## Call:
## lm(formula = GPA ~ ACT, data = Dat)
##
## 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 ***
## ACT          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
```

H0 :b1 =0 against HA :b1 != 0.

Null hypothesis: No statistical significance between GPA and ACT

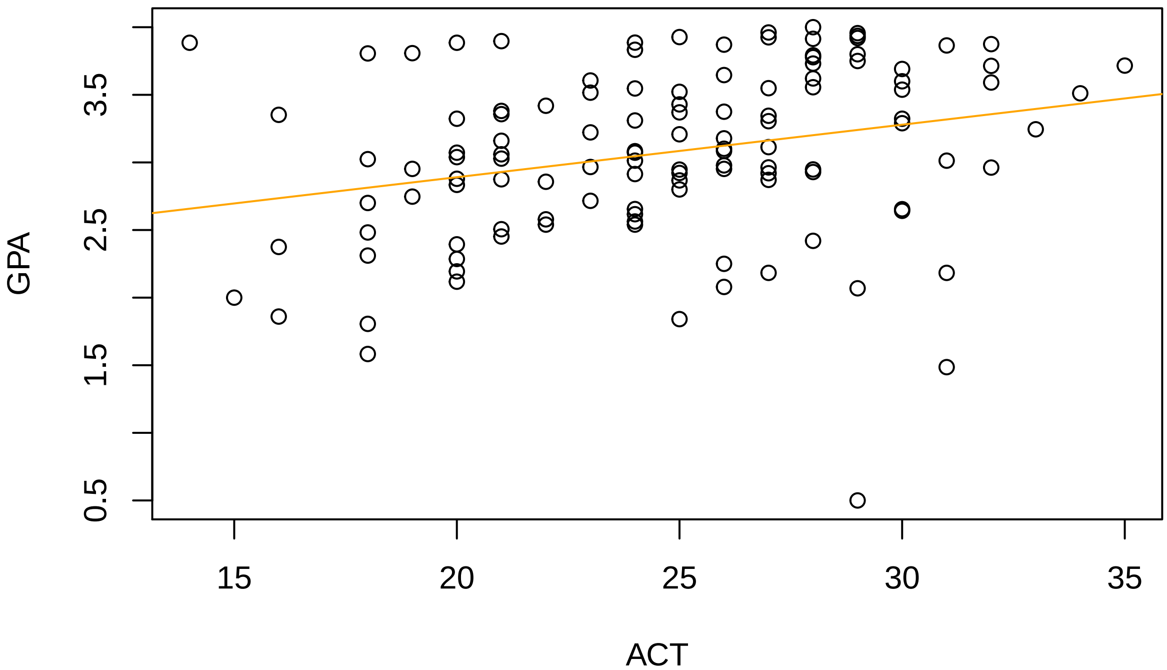
Alternative hypothesis: Statistical significance between GPA and ACT

T-test is 3.040 with p value 0.00292 P value is less than alpha, therefore we reject null hypothesis. There is a statistical significance between the two vairables

This means that there is a linear relationship between Act score and GPA GPA can be a perdictor of ACT score

(h) Construct a scatter plot of the length of the tibia against the length of the humerus. Superimpose the regression line.

```
plot(GPA ~ ACT, data=Dat)
abline(coef(GPA.lm), col="orange")
```



Regression line: y = 2.11405 + 0.3883(x)

(i) Compute the ANOVA table corresponding to the model.

```
anova(GPA.lm)
```

```
## Analysis of Variance Table
##
## Response: GPA
##            Df Sum Sq Mean Sq F value   Pr(>F)
## ACT         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
```

(j) From the table determine the mean square error(MSE).

```
mean(summary(GPA.lm)$residuals^2)
```

```
## [1] 0.3818134
```

Mean Square Error: 0.6231^2 * 118 / 120 = rss / n = 0.3818

Mean Square Error is 0.3818134

(k) Use the ANOVA F-test to determine whether or not there is a linear relationship between ACT score and GPA.

F value is 9.2402

F(0.95,1,118) = 3.92

Since F value 9.24 > 3.92 we conclude b1 != 0

Reject the null-hypothesis

There is a linear relationship between ACT score and GPA

(l) How do the results in (k) compare to those in (g)?

Rejecting null-hypothesis in both situations

(m) What proportion of the variation in GPA is explained by the regression model?

R^2 = SSreg / SST = 1 - RSS / SST

```r
summary(GPA.lm)$r.square
```

```
## [1] 0.07262044
```

This implies 75.283% of the variability have been accounted for and the remaining 21.717% of the variability is still unaccounted for.

Indication that the fit is good

(n) Construct a 95% confidence interval for the estimated mean GPA of students whose ACT test score was 28.

```r
predict(GPA.lm, data.frame(ACT=28), interval = "confidence", level = 0.95, se.fit = TRUE)
```

```
## $fit
##        fit      lwr      upr
## 1 3.201209 3.061384 3.341033
##
## $se.fit
## [1] 0.07060873
##
## $df
## [1] 118
##
## $residual.scale
## [1] 0.623125
```

95% of students with an ACT score of 28 will score a GPA between 3.06 and 3.34

(o) Construct a 95% prediction interval for a particular student whose ACT test score was 20.

```r
predict(GPA.lm, data.frame(ACT=20), interval = "prediction", level = 0.95, se.fit = TRUE)
```

```
## $fit
##        fit      lwr     upr
## 1 2.890592 1.645753 4.13543
##
```

```
## $se.fit
## [1] 0.08293451
##
## $df
## [1] 118
##
## $residual.scale
## [1] 0.623125
```

With a test score of 20 on the ACT, the student is perdicted to score between a GPA of 1.645753 and 4.13543. We have a 95% confidence interval.