STAT 511: HW #3 Q:1 & 2

Rumil Legaspi

15 February 2021

```r
setwd("C:/Users/RUMIL/Desktop/APU/STAT 511 - Millie Mao (Applied Regression Analysis)/week 4/Week 4")


gpa_data = read.table(file = "GPA.txt", header = FALSE, sep = "")

#Adding headers
names(gpa_data) <- c("GPA", "ACT")
head(gpa_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
```

```r
#Defining dependent and independent vars
ACT = gpa_data$ACT #X
GPA = gpa_data$GPA #Y

gpa_lm = lm(GPA ~ ACT, data = gpa_data)
summary(gpa_lm)
```

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

size.lm = lm(hours~size, data = lotdata) summary(size.lm)

#ANOVA decomp[osition aov(size.lm)

#ANOVA with F-test anova(size.lm)

#General linear test red.lm = lm(hours~1, data = lotdata) #hours~1 is reduced model anova(red.lm, size.lm)

1.Refer to the GPA problem in HW#1

(a).Setting up the ANOVA table

```r
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
```

(b). What does MSR measure in your ANOVA table? What does MSE measure? Under what condition do MSR and MSE estimate the same quantity?

To begin with MSR (Mean Squared of Regression) comes from the sum of squares (***of our X values) divided by the degrees of freedom from our regression model. In this case MSR measures our regression model's variability (separate from the variability of our error(MSE)).

MSE is the mean squared error and measures how close our regression line is to the data points. It measures the distances from the data points to our regression line. Those distances are the errors and the MSE is squaring those errors. **Ultimately** it measures the variability/spread of those errors.

Both the MSR and MSE estimate the same quantity when the slope β_1 is zero or not. (*** from the slide 8/39, the same quantity error is the same quantity measured when slope is zero or not)

(c). at $\alpha = 0.05$, conduct an F-test of whether or not $\beta_1 = 0$. State the null and alternative hypotheses, decision rule, and conclusion.

Null hypothesis: H0: $\beta_1 = 0$ (slope is horizontal/ no relationship)

Alternative hypothesis: H1: β_1 /= 0 (slope exists/ relationship exists)

Given our F-value is 9.2402

___note: 1. If the f statistic is larger than the critical value

2. or the p value is less than alpha then it is significant and we reject the null___

```r
#Using qt() to find our critical value we get
qt(0.995, 118)
```

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

```
#*** why do we use 0.995 and not 0.99 here?
```

The F-value is greater than our critical value (2.618137) **AND** our p-value (0.002917) is less than our $\alpha = 0.05$ we reject our null and therefore a non-zero slope exists and a significant relationship between ACT(X) and GPA(Y) scores.

##currently at 1:07:00 in class recording