Naomi Kaduwela
Oct. 19 2018
Chapter 3, Exercises

<u>3.3</u> -

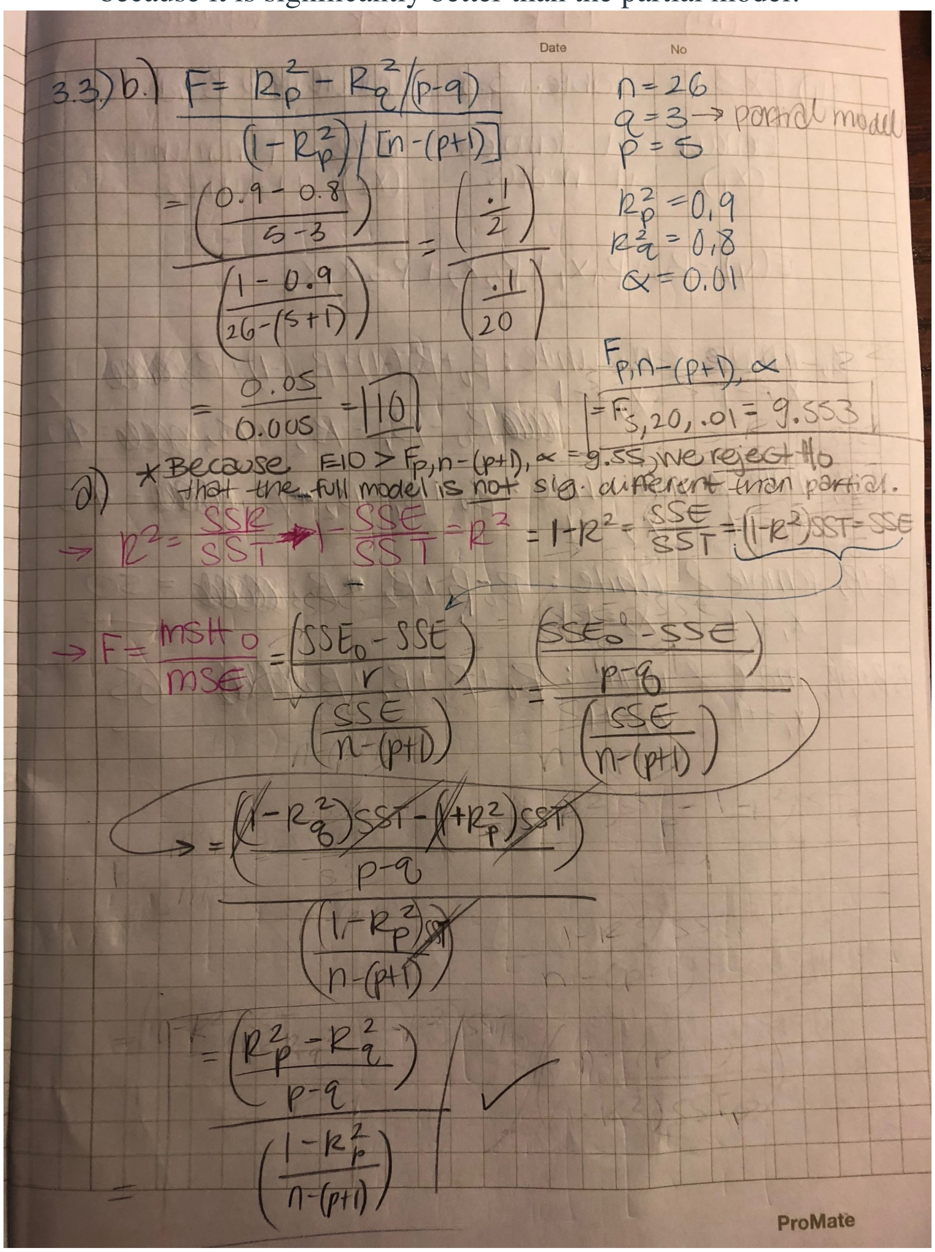
a.) Show that the extra SS F statistic equals

F = <u>Rp^2 - Rq^2/ p-q</u>

1-Rp^2/n-(p+1)

b.) & solve when *n= 26, q = 3, and p = 5*

- Because F = 10 > than the F test value, we reject the null hypothesis (that the model is not significantly different). Thus, we can conclude that the full model is good to use because it is significantly better than the partial model.



Date No

3.3) b.) $F = \dfrac{R_p^2 - R_q^2/(p-q)}{(1-R_p^2)/[n-(p+1)]}$

$= \dfrac{\left(\frac{0.9-0.8}{5-3}\right)}{\left(\frac{1-0.9}{26-(5+1)}\right)} = \dfrac{\left(\frac{.1}{2}\right)}{\left(\frac{.1}{20}\right)}$

$= \dfrac{0.05}{0.005} = \boxed{10}$

$n=26$
$q=3 \rightarrow$ partial model
$p=5$
$R_p^2 = 0.9$
$R_q^2 = 0.8$
$\alpha = 0.01$

$F_{p,n-(p+1),\alpha}$
$= F_{5,20,.01} = 9.553$

a) *Because $F=10 > F_{p,n-(p+1),\alpha} = 9.55$, we reject H_0 that the full model is not sig. different than partial.

$\rightarrow R^2 = \frac{SSR}{SST} \rightarrow 1 - \frac{SSE}{SST} = R^2 = 1+R^2 = \frac{SSE}{SST} = (1-R^2)SST = SSE$

$\rightarrow F = \dfrac{MSH_0}{MSE} = \dfrac{\left(\frac{SSE_0 - SSE}{r}\right)}{\left(\frac{SSE}{n-(p+1)}\right)} = \dfrac{\left(\frac{SSE_0 - SSE}{p-q}\right)}{\left(\frac{SSE}{n-(p+1)}\right)}$

$= \dfrac{\left(\frac{(1-R_q^2)\cancel{SST} - (1-R_p^2)\cancel{SST}}{p-q}\right)}{\left(\frac{(1-R_p^2)\cancel{SST}}{n-(p+1)}\right)}$

$= \dfrac{\left(\frac{R_p^2 - R_q^2}{p-q}\right)}{\left(\frac{1-R_p^2}{n-(p+1)}\right)}$ ✓

ProMate

<u>3.10</u> - Matrix Calculations: Fit a line to y = B0 + B1x

x = 1, 2, 3, 4, 5

y = 2, 6, 7, 9, 10

a.) Write x matrix and y vector

b.) Calculate X'X and its inverse. Check that the product of the original matrix and its inverse = the identity matrix

c.) Calculate the X'y vector

d.) Calculate the LS estimates B0 and B1

B0 = (X1'X1)^-1 X'y

B1 = (X2'X2)^-1 X2'y

3.1)

x	1	2	3	4	5
y	2	6	7	9	10

$y = X\beta + \varepsilon$

a)

$$\begin{bmatrix} 2 \\ 6 \\ 7 \\ 9 \\ 10 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

$$y = X \quad \beta + \varepsilon$$

b.) $X'X$ & inverse (X^{-1}):

↳ Transpose $x = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 & 5 \end{bmatrix} = X'$

$$X'X = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 & 5 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \end{bmatrix} = \begin{array}{c} \begin{array}{cc} a & c \end{array} \\ \begin{bmatrix} 5 & 15 \\ 15 & 55 \end{bmatrix} \\ \begin{array}{cc} d & b \end{array} \end{array}$$

$$\begin{bmatrix} a & c \\ d & b \end{bmatrix}^{-1} = X^{-1} = \frac{1}{ab-cd} \begin{bmatrix} b & -c \\ -d & a \end{bmatrix}$$

$$= \frac{1}{(5)(55)-(15)(15)} \begin{bmatrix} 55 & -15 \\ -15 & 5 \end{bmatrix}$$

$$= \frac{1}{275-225} \begin{bmatrix} 55 & -15 \\ -15 & 5 \end{bmatrix} = \frac{1}{50} \begin{bmatrix} 55 & -15 \\ -15 & 5 \end{bmatrix}$$

$$(X'X)^{-1} = \begin{bmatrix} 1.1 & -0.3 \\ -0.3 & 0.1 \end{bmatrix} = \begin{bmatrix} \frac{55}{50} = \frac{11}{10} & \frac{-15}{50} = \frac{-3}{10} \\ \frac{-15}{50} = \frac{-3}{10} & \frac{5}{50} = \frac{1}{10} \end{bmatrix} = \boxed{\begin{bmatrix} 11/10 & -3/10 \\ -3/10 & 1/10 \end{bmatrix}}$$

$XX^{-1} = I$ check: $\begin{bmatrix} 5 & 15 \\ 15 & 55 \end{bmatrix} \begin{bmatrix} 11/10 & -3/10 \\ -3/10 & 1/10 \end{bmatrix}$

$$= \begin{bmatrix} \frac{55}{10} + \frac{-45}{10} & \frac{-15}{10} + \frac{15}{10} \\ \frac{165}{10} - \frac{165}{10} & \frac{-45}{10} + \frac{55}{10} \end{bmatrix} = \boxed{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}}$$

c.) $X'y$

$$= \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 & 5 \end{bmatrix} \begin{bmatrix} 2 \\ 6 \\ 7 \\ 9 \\ 10 \end{bmatrix}$$

$$= \begin{bmatrix} 34 \\ 2+12+21+36+50 \\ 121 \end{bmatrix} = \begin{bmatrix} 34 \\ 121 \end{bmatrix}$$

d.) LS estimate $\hat{\beta}_0$ & $\hat{\beta}_1$

$$\hat{\beta}_1 = (X'X)^{-1}X'y$$

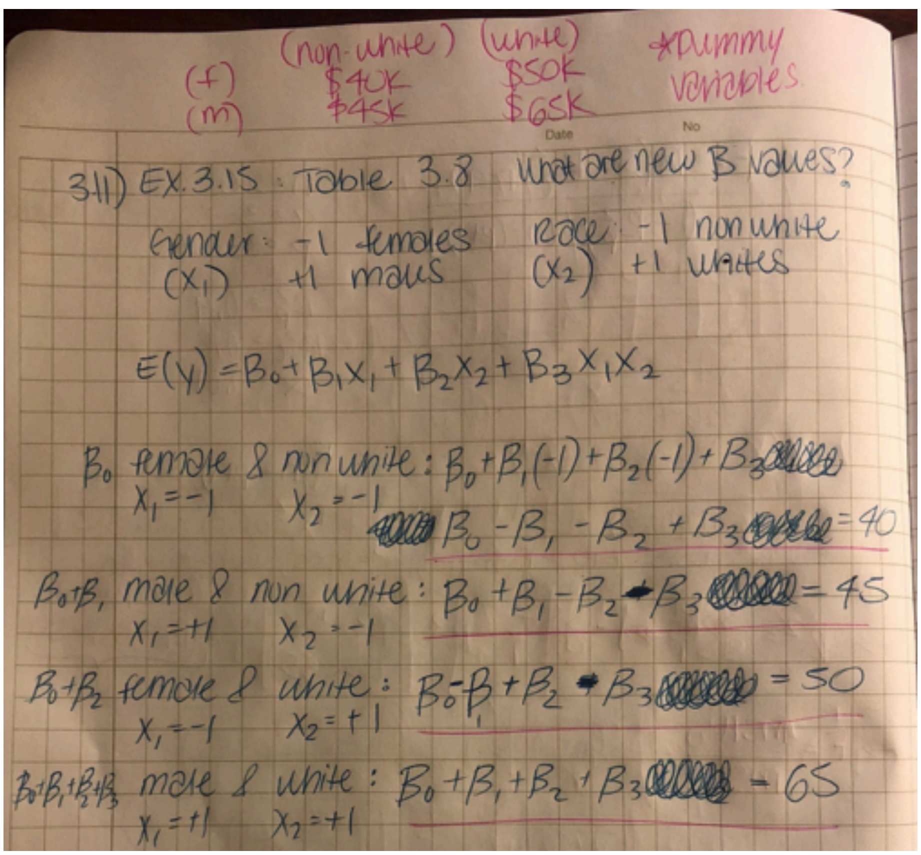
$$= \begin{bmatrix} 11/10 & -3/10 \\ -3/10 & 1/10 \end{bmatrix} \begin{bmatrix} 34 \\ 121 \end{bmatrix} = \frac{1}{10} \begin{bmatrix} 11 & -3 \\ -3 & 1 \end{bmatrix} \begin{bmatrix} 34 \\ 121 \end{bmatrix}$$

$$= \frac{1}{10} \begin{bmatrix} 374-363 = 11 \\ -102+121 = 18 \end{bmatrix}$$

$$\boxed{\hat{\beta}'s = \begin{bmatrix} 11/10 \\ 18/10 \end{bmatrix}} \quad \boxed{\hat{\beta}_0 = 11/10 \quad \hat{\beta}_1 = 15/10}$$

3.11 - Suppose Gender is encoded as x1 = -1 for females and x1 = 1 for males. Similarly, Race is encoded as x2 = -1 for nonwhites and x2 = 1 for whites. What are the new values for B0, B1, B2, B3? Interpret them.

- Calculated Beta Values:
  - B0 = 50
  - B1 = 5
  - B2 = 7.5
  - B3 = 2.5
- Since we changed the encoding of x's from 0 and 1 to -1, we know that the beta values will change. However, we cannot interpret conclusions on the magnitude of change, as it is not comparable.
  - Note Beta values when using 0 and 1 encoding:
    - B0=40k
    - B1 = 5k
    - B2 = 10
    - B3 = 10k



	(non-white)	(white)
(f)	\$40K	\$50K
(m)	\$45K	\$65K

*Dummy variables

3.11) EX.3.15 Table 3.8 What are new B values?

Gender: -1 females; (x_1) +1 maus

Race: -1 non white; (x_2) +1 whites

$E(y) = B_0 + B_1x_1 + B_2x_2 + B_3x_1x_2$

B_0 female & non white: $B_0 + B_1(-1) + B_2(-1) + B_3$
$x_1 = -1$ $x_2 = -1$ $B_0 - B_1 - B_2 + B_3 = 40$

$B_0 + B_1$ male & non white: $B_0 + B_1 - B_2 - B_3 = 45$
$x_1 = +1$ $x_2 = -1$

$B_0 + B_2$ female & white: $B_0 - B_1 + B_2 - B_3 = 50$
$x_1 = -1$ $x_2 = +1$

$B_0 + B_1 + B_2 + B_3$ male & white: $B_0 + B_1 + B_2 + B_3 = 65$
$x_1 = +1$ $x_2 = +1$

Equations:

1.) $\beta_0 - \beta_1 - \beta_2 + \beta_3 = 40$

2.) $\beta_0 + \beta_1 - \beta_2 - \beta_3 = 45$

3.) $\beta_0 - \beta_1 + \beta_2 - \beta_3 = 50$

4.) $\beta_0 + \beta_1 + \beta_2 + \beta_3 = 65$

Solve:

1.) $\beta_0 - \cancel{\beta_1} - \cancel{\beta_2} + \beta_3 = 40$

4.) $+\beta_0 + \cancel{\beta_1} + \cancel{\beta_2} + \beta_3 = 65$

$= \dfrac{2\beta_0 + 2\beta_3}{2} = \dfrac{105}{2}$

$= \beta_0 + \beta_3 = 52.5$

$= \beta_0 = 52.5 - \beta_3$

$= \beta_0 = 52.5 - 2.5$

$= \boxed{\beta_0 = 50}$

2.) $\beta_0 + \cancel{\beta_1} - \cancel{\beta_2} - \beta_3 = 45$

3.) $+\beta_0 - \cancel{\beta_1} + \cancel{\beta_2} - \beta_3 = 50$

$= \dfrac{2\beta_0 - 2\beta_3}{2} = \dfrac{95}{2}$

$= \beta_0 - \beta_3 = 47.5$

$= 52.5 - \beta_3 - \beta_3 = 47.5$

$= -2\beta_3 = -5$

$= \boxed{\beta_3 = 2.5}$

1.) $\beta_0 - \beta_1 - \beta_2 + \beta_3 = 40$

$50 - \beta_1 - \beta_2 + 2.5 = 40$

$-\beta_1 - \beta_2 = -12.5$

2.) $\beta_0 + \beta_1 - \beta_2 - \beta_3 = 45$

$50 + \beta_1 - \beta_2 - 2.5 = 45$

$\beta_1 - \beta_2 = -2.5$

$-\cancel{\beta_1} - \beta_2 = -12.5$

$+\cancel{\beta_1} - \beta_2 = -2.5$

$-2\beta_2 = -15$

$\boxed{\beta_2 = 7.5}$

$= \beta_1 - 7.5 = -2.5$

$\boxed{= \beta_1 = 5}$

3.12 -

a.) Fit the Cobb Douglas prediction function y = B0 x1^B1 x2^B2

```
Call:
lm(formula = (log(cobbDouglas$output)) ~ (log(cobbDouglas$capital) +
    log(cobbDouglas$labor)), data = cobbDouglas)

Residuals:
    Min      1Q  Median      3Q     Max 
-1.7604 -0.2665 -0.0694  0.1926  3.7975 

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)    
(Intercept)               -1.71146    0.09671  -17.70   <2e-16 ***
log(cobbDouglas$capital)   0.20757    0.01719   12.08   <2e-16 ***
log(cobbDouglas$labor)     0.71485    0.02314   30.89   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4781 on 566 degrees of freedom
Multiple R-squared:  0.8378,	Adjusted R-squared:  0.8373 
F-statistic:  1462 on 2 and 566 DF,  p-value: < 2.2e-16
```

b.) Find Var(B1), Var(B2), Cov(B1) and test the null hypothesis for the constant returns to scale (H0: B1 + B2 = 1)

- Var(B1) = 0.00295
- Var(B2) = 0.00053
- Cov(B1, B2) = -0.0027
- As our T Calculated value is < the T critical Value, we cannot reject the Null hypothesis (Ho: B1 + B2 = 1)

```
> vcov(cobbModel)
                           (Intercept) log(cobbDouglas$capital) log(cobbDouglas$labor)
(Intercept)               0.0093529294             0.0009477719          -0.0021749139
log(cobbDouglas$capital)  0.0009477719             0.0002954123          -0.0002674770
log(cobbDouglas$labor)   -0.0021749139            -0.0002674770           0.0005355372
```

- Looking at the absolute value of T calculated value, we see that the absolute value of it is greater than the T critical value, thus we reject the Null Hypothesis (Ho: B1 + B2 = 1)

```
[1] "Calculated T value: -4.50945035247538"
[1] "T Critical Value: 1.96414191000173"
```

```
# b.) Check Hypothesis: B1 + B2 = 1

#Use vcov function to get estimates for Var(β1), Var(β2), and Cov(β1, β2)
vcov(cobbModel)
vcovCobbModel = vcov(cobbModel)

#store values to use in t-test
#Var(β1), Var(β2),
varB1 <- as.numeric(vcovCobbModel[2,2])
paste("vcov coeff 1",varB1)
varB2 <- as.numeric(vcovCobbModel[3,3])
paste("vcov coeff 2",varB2)
#calcualte covariabnce: Cov(β1, β2)
covarB1B2 <- as.numeric(vcovCobbModel[2,3])
paste("vcov coeff 2",covarB1B2)

#Calculate t value :  t = |β1 - βj|/standard(βj)
t_value <- ((b1+b2)-1)/sqrt(varB1 + varB2 + (2*covarB1B2))
paste("Calculated T value:", t_value) #t = -107.5672

#Look up t statistic value, sig = 0.5/2 = 0.025, n
paste("T Critical Value:", abs(qt(0.975, 569))) #t = 1.964142
```

c.) F statistic

- Since our F calculated value > than the F critical value we reject the Null Hypothesis (Ho: B3 = 0)
- In part b.) the t value confirmed that we should reject B1 + B2 = 1. Now, in part c.) we see the f value confirmed that we should reject B3 =0. Thus, overall we can conclude that B3 = B1 + B2 -1 is not valid.

```
#c.)
#Partial model
partial <- lm(I(log(cobbDouglas$output) - (log(cobbDouglas$labor))) ~
I(log(cobbDouglas$capital) - log(cobbDouglas$labor)), cobbDouglas)
summary(partial)

#Full model
full <- lm(I(log(cobbDouglas$output) - (log(cobbDouglas$labor))) ~
I(log(cobbDouglas$capital) - log(cobbDouglas$labor)) + log(cobbDouglas$labor), cobbDouglas)
summary(full)

#ANOVA
anova(partial)
anova(full)

#get SSE value for anova formula
SSEpartial <- anova(partial)$`Sum Sq`[2]
SSEfull <- anova(full)$`Sum Sq`[3]
SSEpartial
SSEfull

#calculate F value = partial = [ (full/p-q) / (SSE/n-(p+1)) ]
f = ((SSEpartial - SSEfull)/(3-2)) / (SSEfull/(569-(3+1)))
paste("F calculated value:", f) #calculated F value
paste("F critical value:", qf(.95,1,565)) #F criical value=
```
```

```
[1] "F calculated value: 20.299214667869"
[1] "F critical value: 3.85796988014781"
```
```

__3.15__ -

a.) Fit prediction model and ensure that it matches

$$\log(10)(\mathit{Salary}) = 4.429 + 0.0075\ \mathit{YrsEm} + 0.0017\mathit{PriorYr} + 0.0170\ \mathit{Educ}$$
$$+\ 0.0004\ \mathit{Super} + 0.0231\ \mathit{Female} - 0.0388\ \mathit{Advert} - 0.00573\ \mathit{Egg} - 0.0938\ \mathit{Sales}$$

```
Call:
lm(formula = (log10(salaries$Salary)) ~ (salaries$YrsEm + salaries$PriorYr +
    salaries$Education + salaries$Super + salaries$Female + salaries$Advert +
    salaries$Eng + salaries$Sales), data = salaries)

Residuals:
      Min        1Q    Median        3Q       Max
-0.089659 -0.024036 -0.004498  0.028587  0.089410

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)        4.4287934  0.0213399 207.535  < 2e-16 ***
salaries$YrsEm     0.0074788  0.0011931   6.269 2.72e-07 ***
salaries$PriorYr   0.0016839  0.0019568   0.861 0.395039
salaries$Education 0.0170345  0.0033360   5.106 1.02e-05 ***
salaries$Super     0.0003901  0.0008056   0.484 0.631115
salaries$Female1   0.0230683  0.0142917   1.614 0.115002
salaries$Advert1  -0.0387774  0.0249146  -1.556 0.128124
salaries$Eng1     -0.0057292  0.0197703  -0.290 0.773597
salaries$Sales1   -0.0937783  0.0225745  -4.154 0.000185 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.04586 on 37 degrees of freedom
Multiple R-squared:  0.8634,    Adjusted R-squared:  0.8338
F-statistic: 29.22 on 8 and 37 DF,  p-value: 9.629e-14
```

b.) Use Female and Sales as reference categories, what is the new coefficient for Male and the other 3 departments

```
lm(formula = (log10(salaries$Salary)) ~ (salaries$YrsEm + salaries$PriorYr +
    salaries$Education + salaries$Super + salaries$Male + salaries$Advert +
    salaries$Eng + salaries$Purchase), data = salaries)

Residuals:
      Min        1Q    Median        3Q       Max
-0.089659 -0.024036 -0.004498  0.028587  0.089410

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)        4.3580834  0.0248414 175.436  < 2e-16 ***
salaries$YrsEm     0.0074788  0.0011931   6.269 2.72e-07 ***
salaries$PriorYr   0.0016839  0.0019568   0.861 0.395039
salaries$Education 0.0170345  0.0033360   5.106 1.02e-05 ***
salaries$Super     0.0003901  0.0008056   0.484 0.631115
salaries$Male1    -0.0230683  0.0142917  -1.614 0.115002
salaries$Advert1   0.0550009  0.0230111   2.390 0.022045 *
salaries$Eng1      0.0880491  0.0180562   4.876 2.07e-05 ***
salaries$Purchase1 0.0937783  0.0225745   4.154 0.000185 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.04586 on 37 degrees of freedom
Multiple R-squared:  0.8634,	Adjusted R-squared:  0.8338
F-statistic: 29.22 on 8 and 37 DF,  p-value: 9.629e-14
```

c.) The coefficient of Engg is highly nonsignificant with a P-value = 0.774 in the regression. If Sales is used as the reference category, the coefficient of Engg is highlight significant with a P-value < 0.01. Interpret this result. If the coefficient of a dummy variable insignificant, what does it tell you?

- When Sales is used in the model, Engg is -0.0057. When Sales is not used (i.e. is the reference point) in the model, Engg is 0.088. Thus, we can see that Engg has a bigger Beta value when Sales is not included. This could be because there are many records with Sales, thus, when Sales is included, it makes the engineering Beta value less. Overall though, looking at the R^2 of the model, we see it is the same. Thus, the model isn't changing, just the Beta value, which is expected when we change dummy reference variables.