Name: Jiahong Hu   ID: jh3561
HW7 Stat 4315

7.1

(1) df=1

(2) df=1

(3) df=2

(4) df=3

7.3

a.

Using R, we can obtain the following code and result:

```
data<-read.table("http://www.stat.ufl.edu/~rrandles/sta4210/Rclassnotes/data/te
names(data)<-c("Y","X1","X2")
attach(data)
fit1<-lm(Y~X1)
fit2<-lm(Y~X2)
fit3<-lm(Y~X1+X2)

anova(fit1)
anova(fit2)
anova(fit3)
```

```
> anova(fit3)
Analysis of Variance Table

Response: Y
          Df  Sum Sq Mean Sq F value    Pr(>F)
X1         1 1566.45 1566.45 215.947 1.778e-09 ***
X2         1  306.25  306.25  42.219 2.011e-05 ***
Residuals 13   94.30    7.25
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

SSR (X_1, X_2) = 1566.45+306.25 = 1872.7

SSR(X1) = 1566.45 $SSR(X_2 \mid X_1) = 306.25$

b.

H0: $\beta_2 = 0$

Ha: $\beta_2 \neq 0$

$$F^* = \frac{SSR(X_2|X_1)/(3-2)}{SSE(X_1,X_2)/(16-3)} = \frac{306.25/1}{94.30/13} = 42.22$$

$F^* = 42.219 \quad F^{-1}_{1,13}(0.99) = 9.073806$

Since $F^* > F^{-1}_{1,13}(0.99)$, we can conclude $H_\alpha : \beta_2 \neq 0$.

The p-value is 2.011×10^{-5}, close to 0

7.7

a.

```
> data=read.table("618.txt")
> attach(data)
```

```
> X1=data[,2]
> X2=data[,3]
> X3=data[,4]
> X4=data[,5]
> Y=data[,1]
> fitA=lm(Y~X4+X1+X2+X3)
> anova(fitA)
Analysis of Variance Table

Response: Y
          Df Sum Sq Mean Sq F value    Pr(>F)    
X4         1 67.775  67.775 52.4369 3.073e-10 ***
X1         1 42.275  42.275 32.7074 2.004e-07 ***
X2         1 27.857  27.857 21.5531 1.412e-05 ***
X3         1  0.420   0.420  0.3248    0.5704    
Residuals 76 98.231   1.293                      
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

SSR(X_4) = 67.775, SSR($X_1|X_4$) = 168.782-126.508 = 42.274,
SSR($X_2|X_1,X_4$) = 126.508-98.650 = 27.858,
SSR($X_3|X_1, X_2,X_4$) = 98.650-98.231 = 0.419

b.

```
> fitB=lm(Y~X4+X1+X2)
> anova(fitB,fitA)
Analysis of Variance Table

Model 1: Y ~ X4 + X1 + X2
Model 2: Y ~ X4 + X1 + X2 + X3
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1     77 98.650                           
2     76 98.231  1   0.41975 0.3248 0.5704
> qf(0.99,1,76)
[1] 6.980578
```

H0: $\beta_3 = 0$ Ha: $\beta_3 \neq 0$

$$F^* = \frac{SSR(X_3 \mid X_1, X_2, X_4)/1}{SSE(X_1, X_2, X_3, X_4)/13} = \frac{0.419/1}{98.231/76} = 0.324$$

$$F^{-1}_{1,76}(0.99) = 6.980578$$

Since $F^* < F^{-1}_{1,76}(0.99)$, we can conclude $H_0 : \beta_3 = 0$.

The P-value is 0.570

7.10

```
> anova(fit1)
Analysis of Variance Table

Response: Y + 0.1 * X1 - 0.4 * X2
          Df  Sum Sq Mean Sq F value    Pr(>F)    
X3         1   9.205   9.205  6.5187   0.01263 *  
X4         1  31.872  31.872 22.5713 9.058e-06 ***
Residuals 78 110.141   1.412                      
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> anova(fit2)
Analysis of Variance Table

Response: Y
          Df Sum Sq Mean Sq F value    Pr(>F)    
X1         1 14.819  14.819 11.4649  0.001125 ** 
X2         1 72.802  72.802 56.3262 9.699e-11 ***
X3         1  8.381   8.381  6.4846  0.012904 *  
X4         1 42.325  42.325 32.7464 1.976e-07 ***
Residuals 76 98.231   1.293                      
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

H_0: $\beta_1=-0.1$, $\beta_2=0.4$; H_a: $\beta_1\neq-0.1$ or $\beta_2\neq 0.4$

Full model: $Y=\beta_0+\beta_1X_1+\beta_2X_2+\beta_3X_3+\beta_4X_4$

Reduced model: $Y=\beta_0-0.1X_1+0.4X_2+\beta_3X_3+\beta_4X_4$

$$F^*=\frac{SSE(R)-SSE(F)}{df_R-df_F}\div\frac{SSE(F)}{df_F}=\frac{110.141-98.231}{2}\div\frac{98.231}{76}=4.6073$$

Since $F^*<F^{-1}_{2,76}(0.99)=4.89584$, we can conclude $H_0:\beta_1=-0.1,\beta_2=0.4$.

7.12

According to executions in R, we can get the following results

```
> cor(X1,Y)^2
[1] 0.796365
> cor(X2,Y)^2
[1] 0.155694
> cor(X1,X2)^2
[1] 0
>
```

```
> summary(fit1)

Call:
lm(formula = Y ~ X1)

Residuals:
   Min     1Q Median     3Q    Max 
-7.475 -4.688 -0.100  4.638  7.525 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   50.775      4.395  11.554 1.52e-08 ***
X1             4.425      0.598   7.399 3.36e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.349 on 14 degrees of freedom
Multiple R-squared:  0.7964,	Adjusted R-squared:  0.7818 
F-statistic: 54.75 on 1 and 14 DF,  p-value: 3.356e-06
```

```
> summary(fit2)

Call:
lm(formula = Y ~ X2)

Residuals:
    Min      1Q  Median      3Q     Max 
-16.375  -7.312  -0.125   8.688  16.625 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   68.625      8.610   7.970 1.43e-06 ***
X2             4.375      2.723   1.607     0.13    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10.89 on 14 degrees of freedom
Multiple R-squared:  0.1557,	Adjusted R-squared:  0.09539 
F-statistic: 2.582 on 1 and 14 DF,  p-value: 0.1304
```

```
> summary(fit3)

Call:
lm(formula = Y ~ X1 + X2)

Residuals:
   Min     1Q Median     3Q    Max 
-4.400 -1.762  0.025  1.587  4.200 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  37.6500     2.9961  12.566 1.20e-08 ***
X1            4.4250     0.3011  14.695 1.78e-09 ***
X2            4.3750     0.6733   6.498 2.01e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.693 on 13 degrees of freedom
Multiple R-squared:  0.9521,	Adjusted R-squared:  0.9447 
F-statistic: 129.1 on 2 and 13 DF,  p-value: 2.658e-09

>
```

```
> anova(fit1)
Analysis of Variance Table

Response: Y
          Df  Sum Sq Mean Sq F value    Pr(>F)    
X1         1 1566.45 1566.45  54.751 3.356e-06 ***
Residuals 14  400.55   28.61                      
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> anova(fit2)
Analysis of Variance Table

Response: Y
          Df  Sum Sq Mean Sq F value Pr(>F)
X2         1  306.25  306.25  2.5817 0.1304
Residuals 14 1660.75  118.62               
> anova(fit3)
Analysis of Variance Table

Response: Y
          Df  Sum Sq Mean Sq F value    Pr(>F)    
X1         1 1566.45 1566.45 215.947 1.778e-09 ***
X2         1  306.25  306.25  42.219 2.011e-05 ***
Residuals 13   94.30    7.25                      
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> 
```

$R^2_{Y1} = 0.7964$

Interpretation: It means the proportion of variance of Y explained by X1, when X1 is the only covariate in the linear regression model. It also equals to the square of cor(X1,Y). X2 explains a high variation of Y.

$R^2_{Y2} = 0.1557$

Interpretation: It means the proportion of variance of Y explained by X2, when X1 is the only covariate in the linear regression model. It also equals to the square of cor(X2,Y). Compared to X1, X2 explains a significant less variation of Y.

$R^2_{12} = 0$

Interpretation : It equals to the square of cor(X2,Y) = 0, which indicates that X1 and X2 are not correlated with each other.

$$R^2_{Y1|2} = \frac{SSR(X_1 | X_2)}{SSE(X_2)} = 0.9432$$

Interpretation: It measures the proportion reduction in the variation in Y remaining after X2 is included in the model that is gained by also including X1 in the model. When X1 is added into the regression model containing X2, thee rror sum of squares SSE(X2) is reduced by 94.32%.

$$R^2_{Y2|1} = \frac{SSR(X_2 | X_1)}{SSE(X_1)} = 0.7645$$

Interpretation: It measures the proportion reduction in the variation in Y remaining after X1 is included in the model that is gained by also including X2 in the model. When X2 is added into the regression model containing X1, the error sum of squares SSE(X2) is reduced by 94.32%.

$$R^2 = 0.9521$$

Interpretation: It measures the proportion reduction in the variation in Y remaining when X1 and X2 are included in the model. X1 and X2 explain 95.21% of the variation of Y.

7.24

a.

```
> summary(fit1)

Call:
lm(formula = Y ~ X1)

Residuals:
   Min     1Q Median     3Q    Max 
-7.475 -4.688 -0.100  4.638  7.525 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   50.775      4.395  11.554 1.52e-08 ***
X1             4.425      0.598   7.399 3.36e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.349 on 14 degrees of freedom
Multiple R-squared:  0.7964,	Adjusted R-squared:  0.7818 
F-statistic: 54.75 on 1 and 14 DF,  p-value: 3.356e-06
```

Therefore, the fitted regression function is $Y = 4.425X_1 + 50.775$.

b.

```
> summary(fit2)

Call:
lm(formula = Y ~ X2)

Residuals:
    Min      1Q  Median      3Q     Max 
-16.375  -7.312  -0.125   8.688  16.625 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   68.625      8.610   7.970 1.43e-06 ***
X2             4.375      2.723   1.607     0.13    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 10.89 on 14 degrees of freedom
Multiple R-squared:  0.1557,	Adjusted R-squared:  0.09539 
F-statistic: 2.582 on 1 and 14 DF,  p-value: 0.1304
```

When relating Y to X_1 and X_2, the regression function is

$Y = 37.65 + 4.425X_1 + 4.375X_2$. The coefficient of X_1 here is equivalent to that in

part (a). And it indicates that X_1 and X_2 is uncorrelated.

c.

```
> fit4<-lm(Y~X2+X1)
> anova(fit4)
Analysis of Variance Table

Response: Y
          Df  Sum Sq Mean Sq F value    Pr(>F)    
X2         1  306.25  306.25  42.219 2.011e-05 ***
X1         1 1566.45 1566.45 215.947 1.778e-09 ***
Residuals 13   94.30    7.25                      
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>
>
```

Yes, it is equal. $SSR(X_1 | X_2) = SSR(X_1) = 1566.45$.

d.

```
> data<-read.table("6.txt")
> attach(data)
> cor(data)
          V1        V2        V3
V1 1.0000000 0.8923929 0.3945807
V2 0.8923929 1.0000000 0.0000000
V3 0.3945807 0.0000000 1.0000000
```

According to the output of R, X_1 and X_2 are uncorrelated.

7.33

Proof:

$$r_{Y1}^2 = 1 - \frac{SSE(X_1)}{SSTO} \Rightarrow SSE(X_1) = (1 - r_{Y1}^2)SSTO \qquad (1)$$

$$SSR(X_1) = r_{Y1}^2 SSTO \qquad (2)$$

Since $SSR(X_1, X_2) = b_1 \sum (X_{i1} - \bar{X}_1)(Y_i - \bar{Y}) + b_2 \sum (X_{i2} - \bar{X}_2)(Y_i - \bar{Y})$

While $b_1 = \dfrac{\dfrac{\sum (X_{i1} - \bar{X}_1)(Y_i - \bar{Y})}{\sum (X_{i1} - \bar{X}_1)^2} - \left[\dfrac{\sum (Y_i - \bar{Y})^2}{\sum (X_{i1} - \bar{X}_1)^2}\right]^{1/2} r_{Y2} r_{12}}{1 - r_{12}^2}$,

$$b_2 = \frac{\dfrac{\sum (X_{i2} - \bar{X}_2)(Y_i - \bar{Y})}{\sum (X_{i2} - \bar{X}_2)^2} - \left[\dfrac{\sum (Y_i - \bar{Y})^2}{\sum (X_{i2} - \bar{X}_2)^2}\right]^{1/2} r_{Y1} r_{12}}{1 - r_{12}^2}$$

Substituting b_1, b_2 into the above expression of $SSR(X_1, X_2)$ yields

$$SSR(X_1, X_2) = \frac{1}{1 - r_{12}^2}\left[\sum (Y_i - \bar{Y})^2 r_{Y1}^2 + \sum (Y_i - \bar{Y})^2 r_{Y2}^2 - 2\sum (Y_i - \bar{Y})^2 r_{Y1} r_{Y2} r_{12}\right] \qquad (3)$$

Since $r_{Y2|1}^2 = \dfrac{SSR(X_2 \mid X_1)}{SSE(X_1)} = \dfrac{SSR(X_1, X_2) - SSR(X_1)}{SSE(X_1)}$ (4)

Substituting (1), (2), (3) into (4) and simplifying yields:

$$r_{Y2|1}^2 = \frac{(r_{Y2} - r_{12} r_{Y1})^2}{(1 - r_{Y1}^2)(1 - r_{12}^2)}$$