1 Problem 1

1.1 a

We have $F = \frac{MSR}{MSE}$. For the right part of equation in problem, we have

$$\begin{aligned} \frac{n-p-1}{p}\frac{R^2}{1-R^2} &= \frac{n-p-1}{p}\frac{SSR}{SSE} \\ &= \frac{\frac{SSR}{p}}{\frac{SSE}{n-p-1}} \\ &= \frac{MSR}{MSE} \end{aligned} \tag{1}$$

Thus we get the equation.

1.2 b

We calculate the F statistics as 7.128 and its p-value is 0.003 thus we reject the H_0.

2 Problem 2

2.1 a

```
setwd('C:/Users/ycyma/Desktop/Courseworks/2017 Spring/Advanced Data
    Analysis/HW/HW3')
a<-read.csv('CompUSys.csv')
x<-a$x
y<-a$y
comp<-lm(y~x)
summary(comp)
```

```
Call:
lm(formula = y ~ x)

Residuals:
    Min      1Q  Median      3Q     Max
-6.8729 -2.9696 -0.4751  2.8260  7.3315

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  11.4641     3.4390   3.334  0.00875 **
x            24.6022     0.8045  30.580 2.09e-10 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 4.615 on 9 degrees of freedom
Multiple R-squared:  0.9905,    Adjusted R-squared:  0.9894
```

```
F-statistic: 935.1 on 1 and 9 DF,  p-value: 2.094e-10
```

2.2 b

```
> confint(comp)
                2.5 %    97.5 %
(Intercept)  3.684472  19.24371
x           22.782272  26.42215
```

[language=R]

2.3 c

```
> predict(comp,newdata = data.frame(x=6), interval="confidence")
       fit       lwr       upr
1 159.0773 154.1388 164.0159
```

2.4 d

```
> predict(comp,newdata = data.frame(x=6), interval="prediction")
       fit       lwr       upr
1 159.0773 147.5279 170.6268
```

2.5 e

```
> predict(comp,newdata = data.frame(x=6), level=1-0.05/4,interval="
    confidence")
       fit       lwr       upr
1 159.0773 152.2858 165.8689
> predict(comp,newdata = data.frame(x=7), level=1-0.05/4,interval="
    confidence")
       fit       lwr       upr
1 183.6796 174.8148 192.5443
```

2.6 f

```
> full<-lm(y~factor(x))
> anova(comp,full)
Analysis of Variance Table

Model 1: y ~ x
Model 2: y ~ factor(x)
  Res.Df   RSS Df Sum of Sq      F Pr(>F)
1      9 191.7
2      4 100.0  5     91.702 0.7336 0.6353
```

We can't reject H_0

3 Problem 3

3.1 a

```
b<-read.csv('mileage.csv')
y<-b$y
x2<-b$x2
x1<-b$x1
n=length(y)
d1<-rep(0,n)
d2<-rep(0,n)
for (i in 1:n)
{
  if (x1[i]=='B') d1[i]=1
  else d1[i]=0
  if (x1[i]=='C') d2[i]=1
  else d2[i]=0
}
fit1<-lm(y~factor(x1)+x2)
summary(fit1)
fit2<-lm(y~d1+d2+x2)
summary(fit2)
```

```
> fit1<-lm(y~factor(x1)+x2)
> summary(fit1)

Call:
lm(formula = y ~ factor(x1) + x2)

Residuals:
     Min       1Q   Median       3Q      Max
-4.6171  -1.6321   0.5508   1.3756   4.0021

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  32.0171     1.0005  32.002   <2e-16 ***
```

```
factor(x1)B    1.5218      1.2650    1.203     0.245
factor(x1)C    0.5252      1.6194    0.324     0.749
x2            -0.4192      0.6042   -0.694     0.497
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 2.532 on 18 degrees of freedom
Multiple R-squared:  0.09453, Adjusted R-squared:  -0.05638
F-statistic: 0.6264 on 3 and 18 DF,  p-value: 0.6072

> fit2<-lm(y~d1+d2+x2)
> summary(fit2)

Call:
lm(formula = y ~ d1 + d2 + x2)

Residuals:
     Min       1Q   Median       3Q      Max
-4.6171  -1.6321   0.5508   1.3756   4.0021

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  32.0171     1.0005  32.002   <2e-16 ***
d1            1.5218     1.2650   1.203    0.245
d2            0.5252     1.6194   0.324    0.749
x2           -0.4192     0.6042  -0.694    0.497
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 2.532 on 18 degrees of freedom
Multiple R-squared:  0.09453, Adjusted R-squared:  -0.05638
F-statistic: 0.6264 on 3 and 18 DF,  p-value: 0.6072
```

We use two different methods and get the same answer.

3.2 b

```
> fit3<-lm(y~x2)
> anova(fit3,fit1)
Analysis of Variance Table

Model 1: y ~ x2
Model 2: y ~ factor(x1) + x2
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1     20 125.14
2     18 115.42  2    9.7138 0.7574 0.4832
```

we can't reject H_0