MA4605 Laboratory F week 7

Using ANOVA for testing the goodness-of-fit of regression

The following results were obtained when each of a series of standard silver solutions was analysed by flame-absorption spectrometry.

Concentration(X)	0	5	10	15	20	25	30
Absorbance(Y)	0.003	0.127	0.251	0.390	0.498	0.625	0.763

```
Conc=c( 0, 5, 10, 15, 20, 25, 30)
Abso=c( 0.003, 0.127, 0.251, 0.390, 0.498, 0.625, 0.763)
```

Determine the slope and the intercept of the calibration plot and their confidence limits.

We determine the slope and intercept of the calibration plot using the linear regression model function **`lm()`**.

```
FitA = lm(Abso~Conc)
summary(FitA)
confint(FitA)
```

<u>Using ANOVA to test the goodness-of-fit of the linear relationship.</u>

We can also test for significance the slope of the simple linear regression model by means of ANOVA. The null hypothesis for the analysis of variance is $H_0 : \beta_1 = 0$ and the alternative hypothesis is $H_1 : \beta_1 \neq 0$, where β_1 is the slope parameter.

For Multiple Linear Regression the ANOVA procedure is joint significance test for the regression coefficients, to complement the individual hypothesis tests and significance values expressed in the `summary()` command output.

The null hypothesis states that all the regression coefficients are zero, with the exception of the constant β_0. The alternative hypothesis simply states that at least one of the parameters is non zero.

The ANOVA table for the SLR regression line is obtained by calculating the variation due to regression (SSR) and the variation around the regression line(SSE).

Source	Sum of Squares (SS)	DF	Mean Squares (MS)	F
Total	TSS= $\sum (y_i - \overline{y})^2$	n-1		
Regression	SSR= $\sum (\hat{y}_i - \overline{y})^2$	1	MSR= $\frac{SSR}{1}$	F= $\frac{MSR}{MSE}$
Error	SSE= $\sum (y_i - \hat{y}_i)^2$	n-2	MSE= $\frac{SSE}{n-2}$	

Correctly, the degrees of freedom for "Regression" and "Error" are ***k*** and ***n-k-1*** respectively, where ***k*** is the number of independent variables.

The fitted values are the values of y predicted using the regression equation.

These fitted values can be extracted from the linear model using the predict() function.

The residuals can also be extracted from the model with the resid()function.

```
preds = predict(FitA)

resids =  resid(FitA)
```

1) Compute the total sum of squares TSS

In your submission sheet , write down the mean value for "Abso" and the value for TSS.

```
Abso

length(Abso)

mean(Abso)

Abso-mean(Abso)

(Abso-mean(Abso))^2

TSS = sum((Abso-mean(Abso))^2)
TSS
```

2) The sum of squares due to regression SSR

In your submission sheet, write down the value for SSR.

```
preds

(preds-mean(Abso))^2

SSR = sum((preds-mean(Abso))^2)
SSR
```

3) The sum of squares around regression SSE

In your submission sheet, write down the value for SSE. Also compute the summation of SSR and SSE. (Is it equal to TSS)

```
SSE = sum((resids)^2)
SSE

SSR + SSE
```

4) The mean squares

The mean sum of squares are :

$$\mathrm{MSR} = \frac{SSR}{1}$$

$$\mathrm{MSE} = \frac{SSE}{n-2}$$

Where the sample size is 'n'.

We can compute the Mean Square Values and the Test Statistic Fts. Write these values down in your submission sheet.

```
MSE = SSE/(length(Abso)-2)
MSR = SSR/1
MSE
MSR

Fts = MSR/MSE
Fts
```

The Critical Value can be computed using the following code, with the degrees of freedom specified as before. (This is for demonstration purposes only)

```
k=1 # number of independent variables

df1=k
df2=n-k-1

qf(0.975,df1,df2,lower.tail=F).
```

The test statistic value is very significant since it is much greater than the critical value F1;5;0:05= 0.001084778.

We reject the null hypothesis that states $\beta_1 = 0$.

The ANOVA table can be obtained for the regression model with the `anova()` command.

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

Response: Abso
          Df  Sum Sq Mean Sq F value    Pr(>F)
Conc       1 0.44327 0.44327  8979.5 2.481e-09 ***
Residuals  5 0.00025 0.00005
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The same conclusion can be drawn from the t-test for the slope:
The p-value = 2.48e-09 is less than 0.05 hence we reject the null hypothesis.