MA4605 Lab E (Week 6)

Part 1 - Weighted Linear Regression

Part I - Weighted regression. The fluorescence of each of a series of acidic solutions of quinine was determined five times. The results are given below.

Concentration, ng/ml	0	10	20	30	40	50
Fluorescence intensity	4	22	44	60	75	104
(arbitrary units)	3	20	46	63	81	109
	4	21	45	60	79	107
	5	22	44	63	78	101
	4	21	44	63	77	105

Compute the average value and standard deviation for Fluorescence at each level of concentration.

```
Fluo.Matrix = matrix(
c(4,22,44,60,75,104,
3,20,46,63,81,109,
4,21,45,60,79,107,
5,22,44,63,78,101,
4,21,44,63,77,105),
byrow=T,ncol=6)

apply(Fluo.Matrix,2,mean)
apply(Fluo.Matrix,2,sd)
```

Comment on the standard deviations. Are the variance values uniform (roughly the same level) ? Write your answer in the submission sheet.

Using the following R code, fit a linear model for this data (i.e. mean values of “Fluo” v “Conc”). Write out the regression equation in your submission sheet. Comment on the significance of each estimate (mentioning the numbers of asterisks beside each coefficient will suffice).

```
Fluo.Mean = apply(Fluo.Matrix,2,mean)
Conc=c(0,10,20,30,40,50)
Fit1 = lm(Fluo.Mean ~ Conc)
summary(Fit1)
```

Repeat the regression model fitting procedure using each individual observation of Fluoresence. Write down the regression equation. Comment on the significance of each regression coefficient. Also sketch the scatter-plot.

```
Conc=c(0,10,20,30,40,50)
Conc.M=rep(Conc,5)

Fluo.M=c(4,22,44,60,75,104,
3,20,46,63,81,109,
4,21,45,60,79,107,
5,22,44,63,78,101,
4,21,44,63,77,105)
```

```
Fit2 = lm(Fluo.M ~ Conc.M)
summary(Fit2)


plot(Conc.M,Fluo.M,pch=18,col="red")
abline(coef(Fit2))
```

Perform a weighting linear regression model using the following code.

```
Fluo.mean =apply(Fluo.Matrix,2,mean)
Fluo.sd = apply(Fluo.Matrix,2,sd)

weights=Fluo.sd^(-2)/mean(Fluo.sd^(-2))

FitC = lm(Fluo.mean~Conc ,  weights=weights)
```

Write down the regression equation. Comment on the significance of each regression coefficient.

Part 2- Quadratic and Cubic Relationships

In an experiment to determine hydrolysable tannins in plants by absorption spectroscopy the following results were obtained:

```
Abso= c(0.084, 0.183, 0.326, 0.464, 0.643, 0.707, 0.717, 0.734 ,0.749)
Conc= c(0.123, 0.288, 0.562, 0.921, 1.420, 1.717, 1.921, 2.137 ,2.321)

plot(Conc,Abso,pch=18,col="red")

# Generate powers of independent variable - Conc

Conc.squared = Conc^2
Conc.squared

Conc.cubed = Conc^3
Conc.cubed
```

Polynomial regression is similar to Multiple Linear Regression - the various independent variables are simply powers of an underlying variable. Polynomial regression is useful for curvilinear relationships between variables.

On your submission sheet, draw a sketch of the scatter-plot. Comment on the shape of the scatter-plot? Is the relationship linear? Is there curvature present?

For the quadratic and cubic models, the regression equations have the following form.

y dependent variable
x underlying independent variable
b_i regression coefficients

Linear : $y = b_0 + b_1 x$
Quadratic : $y = b_0 + b_1 x + b_2 x^2$
Cubic: $y = b_0 + b_1 x + b_2 x^2 + b_3 x^3$

```
FitA = lm(Abso~Conc)
FitB = lm(Abso~Conc + Conc.squared)
FitC = lm(Abso~Conc + Conc.squared + Conc.cubed)
```

Using the summary command, write out the regression equation for each of these three models.

The number of asterisks beside the p-value indicates the level of significance of the estimates. How many asterisks beside each estimate?

We can remove the intercept term from the model by additionally specifying "-1" in the R code, which specifies to fit a model without an intercept.

(We shall apply it to quadratic and cubic model only)

$y=b_1 x + b_2 x^2$
$y=b_1 x + b_2 x^2 + b_3 x^3$

```
FitD = lm(Abso~Conc + Conc.squared-1)
FitE = lm(Abso~Conc + Conc.squared + Conc.cubed-1)
```

Does removing the intercept improve the model fit? Examine the R-squared values in the summary fit? Discuss?