Assignment 2

Submitted By : Vikram Singh Chandel

Roll No: 173040020

Question 1

a)

```
rm(list=ls())
#install.packages("xlsx")
require("xlsx")
#setwd("/home/de/Dropbox/Hydroinformatics/Assignment 2/")
raw_data=read.xlsx("Temp_173040020_26.5_78.5.xls",sheetIndex = 1,header = FALSE)
summary(raw_data)
YearRain=raw_data[ ,c(1,4)]
annav=matrix(c(0),nrow = 34,ncol = 2)

for(i in 1:34){
ind = which(raw_data[ ,1]==1979+i)
annav[i,1]=1979+i
annav[i,2]=mean(YearRain[ind,2])
}
```

Variable '**annav**' 34x2 contains Year and Annual Mean Temperature.

b)

```
dataF=data.frame("Year" = annav[ ,1],"ANMTEMP" = annav[ ,2])
mfit=lm(ANMTEMP~Year,dataF)
```

Intercept ($\beta 1$)	-46.42202
Slope($\beta 1$)	0.03809
SE of $\beta 0$	41.13844
SE of $\beta 1$	0.02061
Sigma Hat:	1.179
R-squared:	0.0965
t-value for slope $\beta 1$	1.849
p value for slope $\beta 1$	0.0738

Hypothesis testing:

Null Hypothesis: $\beta_1 = 0$

Alternate hypothesis: $\beta_1 \neq 0$

$t_{observed}$=1.849 (from R output and can be calculated manually)

$t_{critical}$ for 99% confidence level is 2.738

As $t_{observed} < t_{critical}$ so we have no evidence to reject null hypothesis.And nothing can be really said about linear relation ship between Annual Mean Temperature and Time.

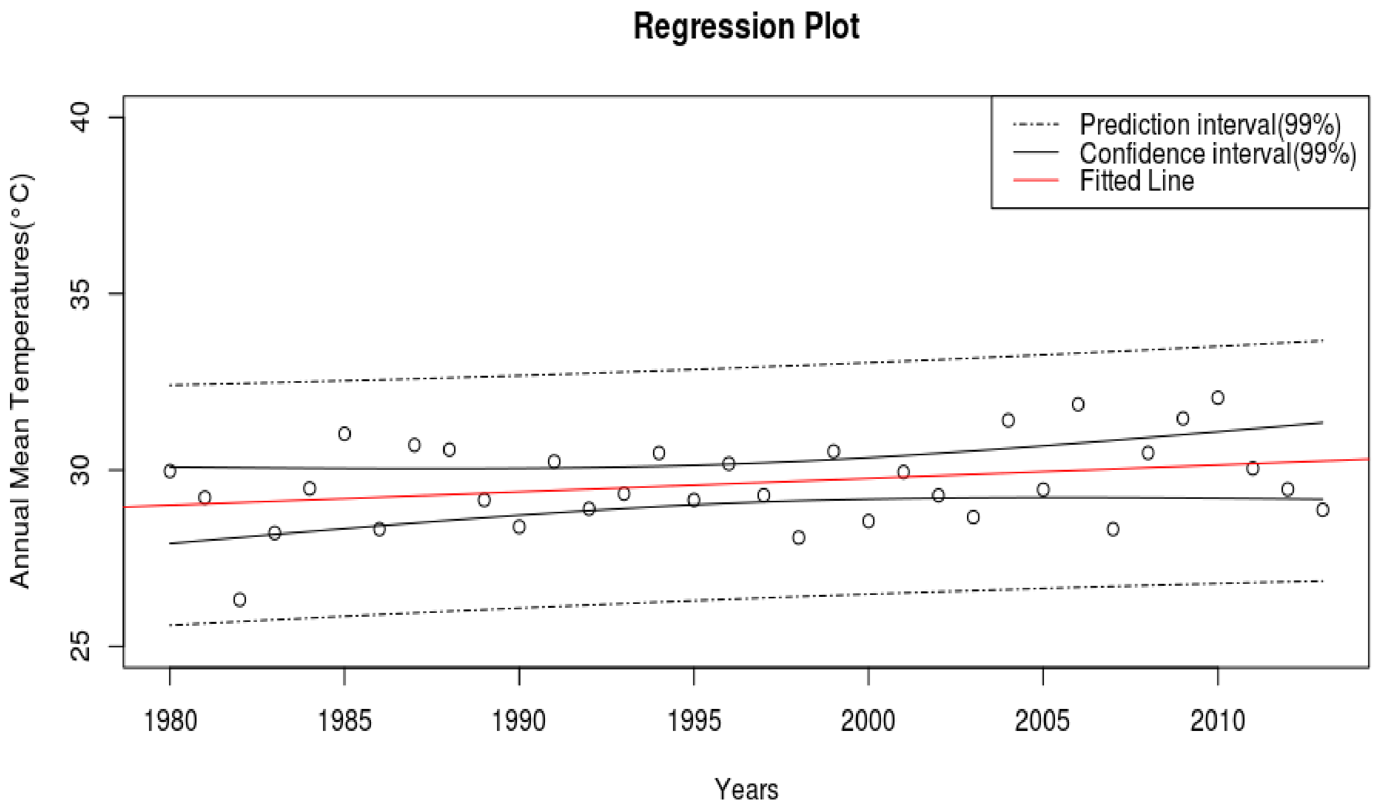
Command to find R^2 from model is

```
summary(mfit)$r.squared
```

c)

```
conf=predict( mfit, data.frame( Year= 1980:2013 ) ,interval ="confidence" , level = 0.99 )
pred=predict( mfit, data.frame( Year= 1980:2013 ) ,interval ="prediction" , level = 0.99 )
plot(annav[ ,1],annav[ ,2],xlim = c(1980,2013),ylim = c(25,40),xlab="Years",ylab = "Annual Mean
Temperatures(°C)",main="Regression Plot")

#plot  fit line
abline (mfit , lwd =1,col="red")

#plot confidence interval
lines(1980:2013,conf[ ,2],col="black",lwd=1)
lines(1980:2013,conf[ ,3],col="black",lwd=1)

#plot prediction interval
lines(1980:2013,pred[ ,2],col="black",lwd=1,lty=4)
lines(1980:2013,pred[ ,3],col="black",lwd=1,lty=4)

legend("topright",c("Prediction interval(99%)","Confidence interval(99%)","Fitted
Line"),lty=c(4,1,1),lwd=c(1,1,1),col=c("black","black","red"))

#optional
x=seq(-4,4,0.1)
```

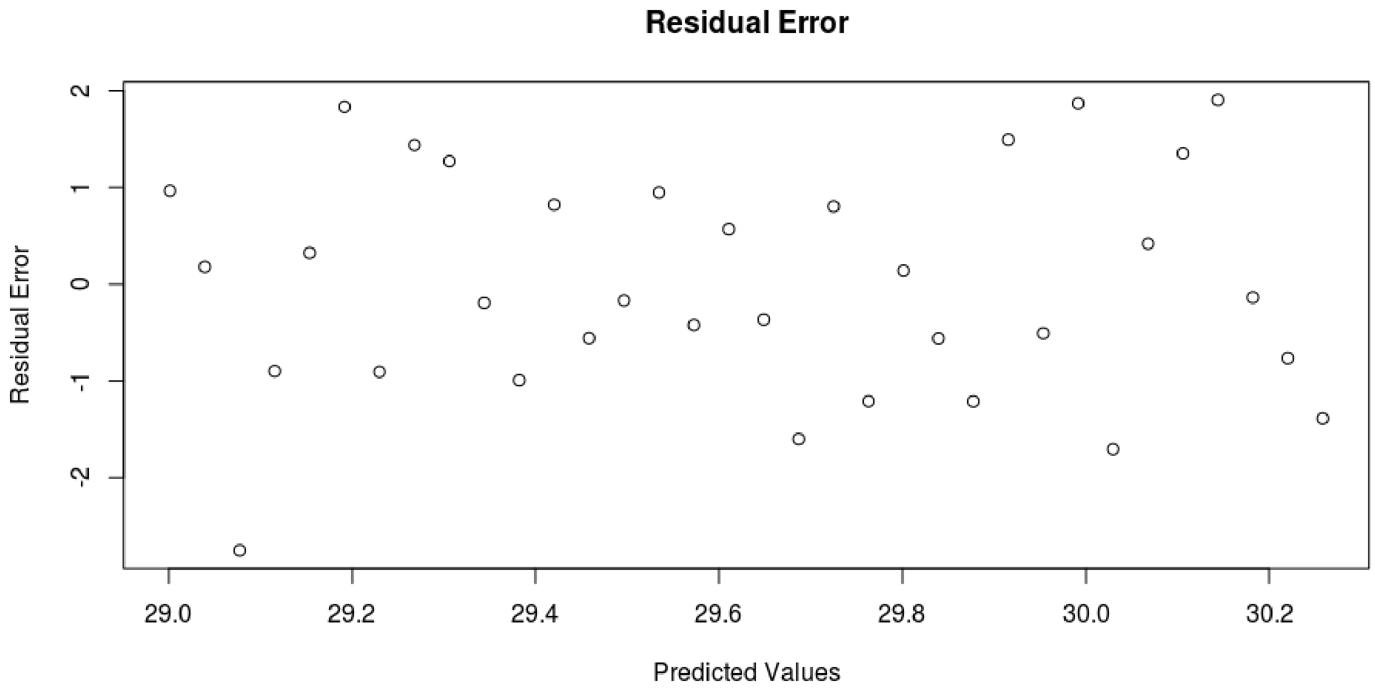
```
plot(annav[ ,1],annav[ ,2],xlim = c(1980,2013),ylim = c(25,35),xlab="Years",ylab = "Annual Mean Temperatures(°C)",main="Regression Plot")
#plot(x,dnorm(x,0,summary(mfit)$sigma),type = 'l',xlim=c(-4,4),ylim=c(0,0.4))
polygon(1982+5*c(0,dnorm(x,0,summary(mfit)$sigma),0),pred[2,1]+c(-4, x, 4), col='gray95')
polygon(1990+5*c(0,dnorm(x,0,summary(mfit)$sigma),0),pred[10,1]+c(-4, x, 4), col='gray95')
polygon(2000+5*c(0,dnorm(x,0,summary(mfit)$sigma),0),pred[20,1]+c(-4, x, 4), col='gray95')
abline (mfit , lwd =1,col="red")
```



Both the intervals are narrower near the mean value of X and wider at the end of training data set which means model performs better near the mean values.

d)

```
plot(pred[ ,1],summary(mfit)$residuals)
```

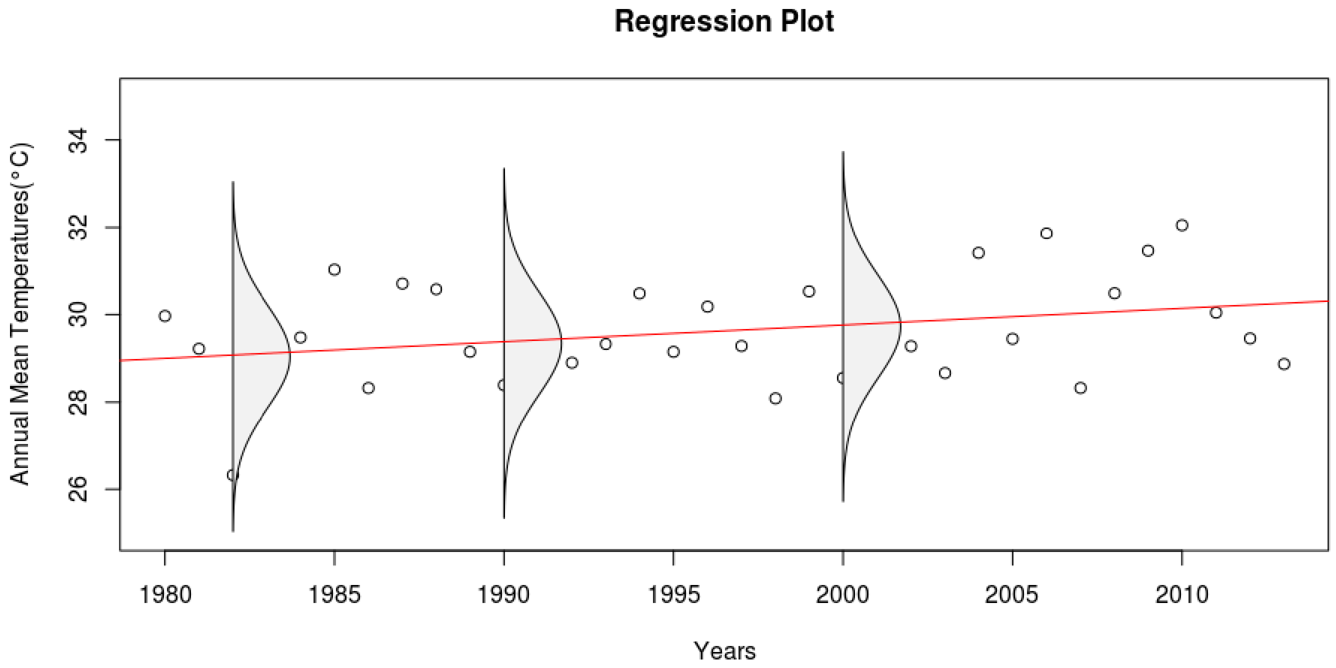


Variation of residual errors with Y is random hence model fitted is appropriate.

e)

The output of fitted glm is similar to linear regression because linear regression is a special case of Generalised linear model with Normal Distribution and link function identity.

(optional)



Question 2

a)

```
rm(list=ls())
#install.packages("xlsx")
require("xlsx")
setwd("/home/de/Dropbox/Hydroinformatics/Assignment 2/")
raw_data=read.xlsx("Temp_173040020_26.5_78.5.xls",sheetIndex = 1,header = FALSE)
summary(raw_data)
YearRain=raw_data[ ,c(1,4)]
annav=matrix(c(0),nrow = 34,ncol = 2)

for(i in 1:34){
ind = which((raw_data[ ,1]==1979+i))
ind2= which(raw_data[ind,4]>33)
annav[i,1]=1979+i
annav[i,2]=length(ind2)
}
```

Variable '**annav**' 34x2 contains Year and frequency of threshold-exceedance .

b)

GLM0 :: $Y=\beta_0$

GLM1 :: $Y=\beta_0 + \beta_1 * X$

*	GLM0(gfit1)	GLM1(gfit2)
$\beta0$	2.72939	-25.798957
$\beta1$	-	0.014284
SE of $\beta0$	0.04381	8.974419
SE of $\beta1$	-	0.004492
t-value for $\beta0$	62.3	-2.875
P value $\beta0$	<2e-16	0.00404
t-value $\beta1$	-	3.180
P value $\beta1$	-	0.00147
AIC	238.56	230.39
BIC:	240.0894	233.4439
Log likelihood	-118.2815	-113.1956

Hypothesis testing:

Null Hypothesis: β_1=0

Alternate hypothesis: β_1 ! =0

t value observed for β_1 is 3.18 and for 95% confidence level t value is 2.036933 (which is less than 3.18) observed and p value is 0.001631429 hence we can reject null hypothesis

c)

Deviace Dobs = 10.177

D01=2*(logLik(gfit2)-logLik(gfit1))

P value for deviance is 1- pchisq(D01, df = 1) = 0.001426013

Moreover Dcirtical=6.634897 for 99% confidence level

As D_{obs}>D_{crit} , we can reject GLM0 in favour of GLM1

Both AIC and BIC for GLM1 are less than that for GLM0,Hence GLM1 is better than GLM0 which is similar to likelihood ratio test.