TD11_ex4

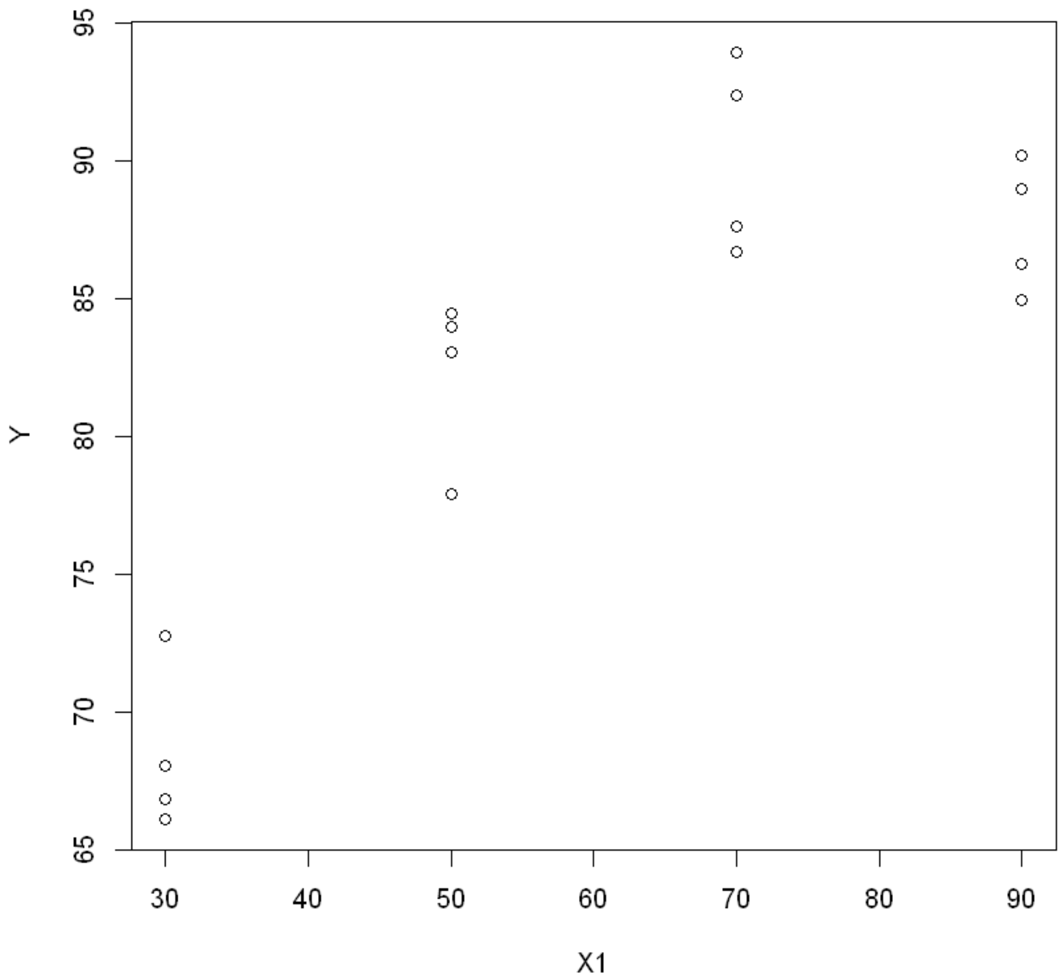
November 20, 2018

1 MTH2302B - H2017

2 TD11 - Exercice 4

2.1 Observations

In [1]:
```
# Données
X1=c(30,30,30,30,50,50,50,50,70,70,70,70,90,90,90,90)
Y=c(66.83,72.77,68.04,66.11,84.45,83.06,77.92,84.00,93.94,86.71,92.38,87.60,86.27,90.1

# Tracer le nuage de points
plot(X1,Y)
```



In [2]:
```
# Nombre de mesures
n = length(X1)

# Moyennes échantillonnales
x_bar = mean(X1)
y_bar = mean(Y)

# Somme des carrés corrigée
S_xx = sum((X1-x_bar)^2)
S_yy = sum((Y-y_bar)^2)
# Somme des produits croisés corrigée
S_xy = sum( (X1-x_bar) * (Y-y_bar))
```

2.2 a) Régression linéaire simple

```
In [3]: # Régression linéaire avec lm
        linReg = lm(Y~X1)
        summary(linReg)

        beta_hat = unname(coefficients(linReg))
        S_beta = unname(summary(linReg)$coefficients[,2])
```

```
Call:
lm(formula = Y ~ X1)

Residuals:
    Min      1Q  Median      3Q     Max
-6.9777 -4.6085 -0.2646  4.4218  8.5395

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 62.55512    3.75179   16.67 1.25e-10 ***
X1           0.32636    0.05859    5.57 6.91e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.241 on 14 degrees of freedom
Multiple R-squared:  0.6891,Adjusted R-squared:  0.6668
F-statistic: 31.02 on 1 and 14 DF,  p-value: 6.905e-05
```

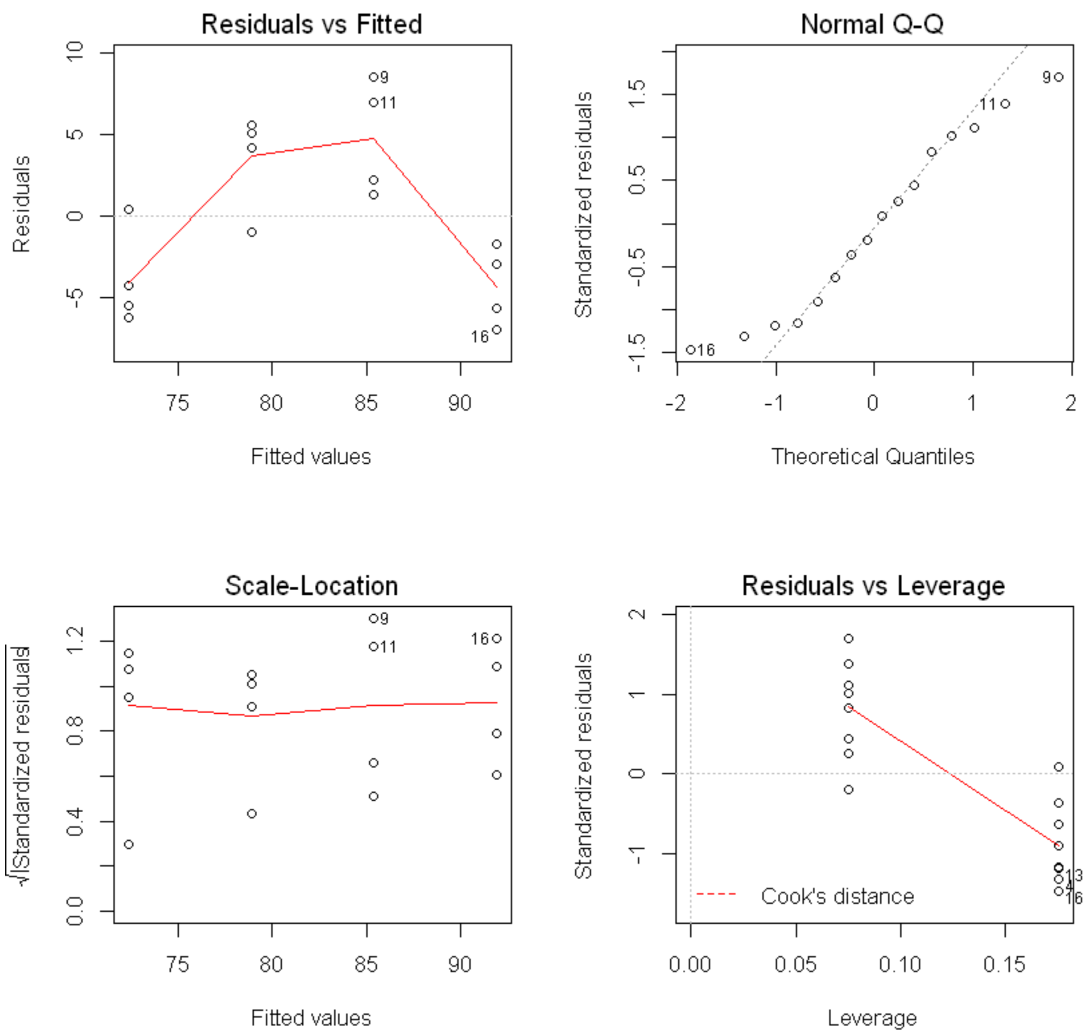
```
In [4]: anova(linReg)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
X1	1	852.0999	852.09985	31.0247	6.905069e-05
Residuals	14	384.5129	27.46521	NA	NA

2.3 b) Analyse des résidus

```
In [5]: par(mfrow=c(2,2))
        Y_hat= fitted.values(linReg)

        # Calcul des résidus
        res=residuals(linReg)

        plot(linReg)

        # Test de normalité
        shapiro.test(res)

        # Conclusion ?
```

```
Shapiro-Wilk normality test

data:  res
W = 0.94329, p-value = 0.3912
```



2.4 1.c) Test de signification du modèle

On teste $H_0 : \beta_1 = 0$ contre $H_1 : \beta_1 \neq 0$

```
In [6]: summary(linReg)
```

```
Call:
lm(formula = Y ~ X1)

Residuals:
    Min      1Q  Median      3Q     Max
-6.9777 -4.6085 -0.2646  4.4218  8.5395

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 62.55512    3.75179   16.67 1.25e-10 ***
X1           0.32636    0.05859    5.57 6.91e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.241 on 14 degrees of freedom
Multiple R-squared:  0.6891,Adjusted R-squared:  0.6668
F-statistic: 31.02 on 1 and 14 DF,  p-value: 6.905e-05
```

2.5 2.c) Intervalle de confiance pour la pente de la droite de régression

Pour un niveau de confiance $1-\alpha$, l'intervalle de confiance pour β_1 est donné par :

$$\beta_1 \in \hat{\beta_1} \pm t_{\alpha/2,n-2}\sqrt{\frac{MSE}{S_{XX}}}$$

```
In [7]: # Intervalle de confiance à 95% pour beta1
        confint(linReg,parm='X1',level = 0.95)
```

	2.5 %	97.5 %
X1	0.2006928	0.4520322

2.6 d) Intervalle de prévision

Pour un niveau de confiance $1-\alpha$, l'intervalle de prévision pour $Y|x = x_0$ est donné par :

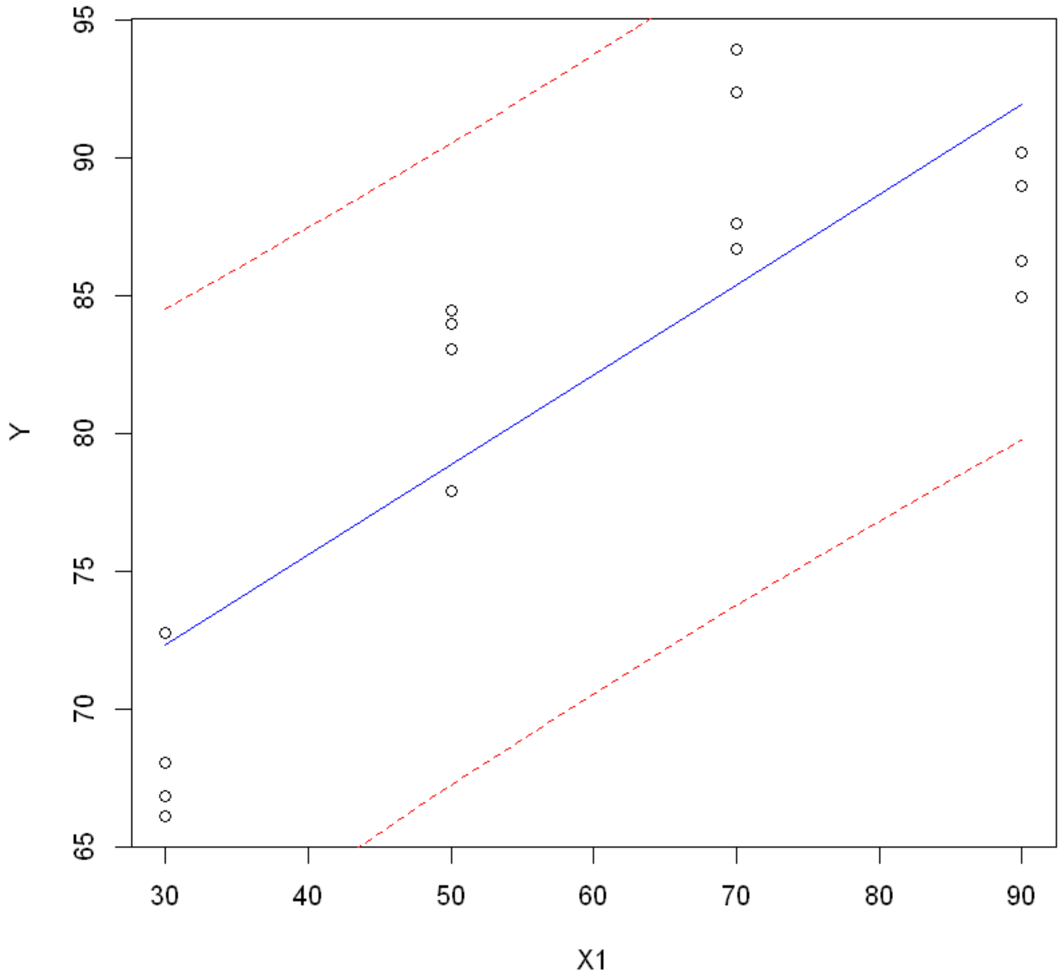
$$Y_0 \in \hat{y_0} \pm t_{\alpha/2,n-2}\sqrt{MS_E\left(1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{S_{XX}}\right)}$$

```
In [8]: # Intervalle de prévision à 95% pour y en x=60
        alpha= 0.05
        x0 = 60
        y0_hat = beta_hat[1] + beta_hat[2] * x0

        MSE = sum(residuals(linReg)^2) / (n-2)
        l =y0_hat - qt(alpha/2,n-2,lower.tail=FALSE) * sqrt(MSE * (1+1/n + (x0 - x_bar)^2 / S_x
```

```
u =y0_hat + qt(alpha/2,n-2,lower.tail=FALSE) * sqrt(MSE * (1+1/n + (x0 - x_bar)^2 / S_x
cat(l, u)
```

```
70.5507 93.72305
```

```
In [9]: # Tracer les intervalles de prévision à 95%
        alpha = 0.05

        x_prev = seq(min(X1),max(X1),by=(max(X1)-min(X1))/20)
        y_prev = beta_hat[1] + beta_hat[2] * x_prev

        MSE = sum(residuals(linReg)^2) / (n-2)
        l_prev =y_prev - qt(alpha/2,n-2,lower.tail=FALSE) * sqrt(MSE * (1+1/n + (x_prev - x_bar
        u_prev =y_prev + qt(alpha/2,n-2,lower.tail=FALSE) * sqrt(MSE * (1+1/n + (x_prev - x_bar

        plot(X1,Y)
        lines(X1,fitted.values(linReg),col='blue')
        lines(x_prev,l_prev,lty=2,col='red')
        lines(x_prev,u_prev,lty=2,col='red')

        remove(x_prev,y_prev,l_prev,u_prev)
```



2.7 e) Régression polynomiale du second degré

2.7.1 1. Mise en forme des données

In [10]:
```
X0 = rep(1,n)
X2 = X1^2
k=2

# Formation de la matrice X
X = matrix(c(X0,X1,X2), nrow = n,ncol = 3)
C = solve( t(X) %*% X)
```

2.7.2 2. Ajustement du modèle de régression

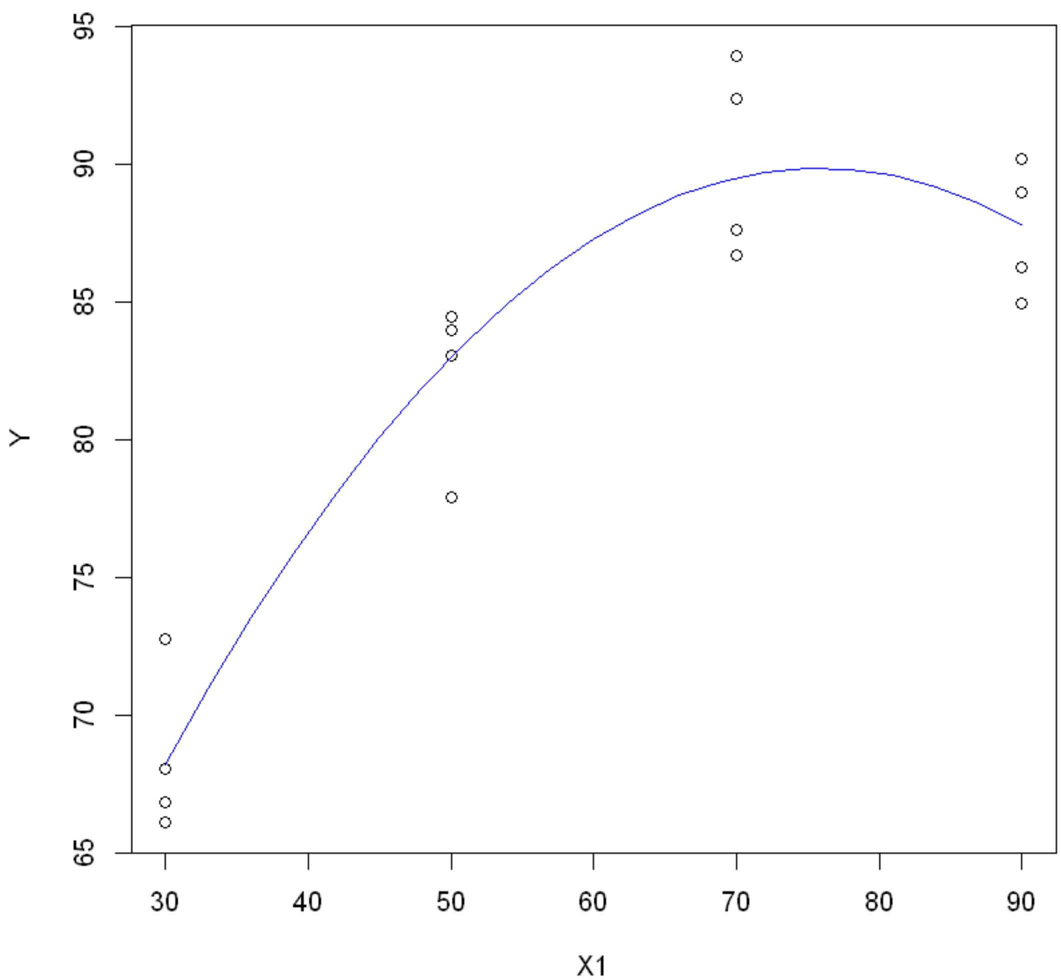
In [11]:
```
# Régression avec lm
qReg = lm(Y ~ X1+X2)
summary(qReg)

# Coefficients de régression et leur écart-type echantillonal
beta_hat = unname(coefficients(qReg))
S_beta = unname(summary(qReg)$coefficients[,2])

cat('beta_hat =', beta_hat, '\n')
cat('S_beta =', S_beta)
```

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

Residuals:
    Min      1Q  Median      3Q     Max
-5.0739 -1.9697 -0.0596  1.6878  4.5446

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 30.620281   6.085428   5.032  0.00023 ***
X1           1.562550   0.223406   6.994 9.42e-06 ***
X2          -0.010302   0.001841  -5.595 8.71e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.946 on 13 degrees of freedom
Multiple R-squared:  0.9088,Adjusted R-squared:  0.8947
F-statistic: 64.73 on 2 and 13 DF,  p-value: 1.744e-07



beta_hat = 30.62028 1.56255 -0.01030156
S_beta = 6.085428 0.2234055 0.001841366
```

2.7.3 3. Courbe de régression

In [12]:
```
# valeurs de la régression
Y_hat = fitted.values(qReg)

# Courbe de regression
plot(X1,Y)

x_reg = seq(min(X1),max(X1),by = 3)
y_reg = beta_hat[1] + beta_hat[2] * x_reg + beta_hat[3] * x_reg^2
lines(x_reg,y_reg,col='blue')
```

```
remove(x_reg)
remove(y_reg)
```



2.7.4 4. Table d'analyse de la variance

In [13]: anova(qReg)

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
X1	1	852.0999	852.099851	98.16811	2.002748e-07
X2	1	271.6728	271.672806	31.29869	8.705056e-05
Residuals	13	112.8401	8.680007	NA	NA

2.8 f) Test de signification global

On teste $H_0 : \beta_1 = \beta_2 = 0$ contre H_1 : au moins un des $\beta_j \neq 0$

```
In [14]: summary(qReg)
```

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

Residuals:
    Min      1Q  Median      3Q     Max
-5.0739 -1.9697 -0.0596  1.6878  4.5446

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 30.620281   6.085428   5.032  0.00023 ***
X1           1.562550   0.223406   6.994 9.42e-06 ***
X2          -0.010302   0.001841  -5.595 8.71e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.946 on 13 degrees of freedom
Multiple R-squared:  0.9088,Adjusted R-squared:  0.8947
F-statistic: 64.73 on 2 and 13 DF,  p-value: 1.744e-07
```

2.9 g) Comparaison des modèles

```
In [15]: summary(linReg)
         summary(qReg)
```

```
Call:
lm(formula = Y ~ X1)

Residuals:
    Min      1Q  Median      3Q     Max
-6.9777 -4.6085 -0.2646  4.4218  8.5395

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 62.55512    3.75179   16.67 1.25e-10 ***
X1           0.32636    0.05859    5.57 6.91e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.241 on 14 degrees of freedom
```

```
Multiple R-squared:  0.6891,Adjusted R-squared:  0.6668
F-statistic: 31.02 on 1 and 14 DF,  p-value: 6.905e-05
```

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

Residuals:
    Min      1Q  Median      3Q     Max
-5.0739 -1.9697 -0.0596  1.6878  4.5446

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 30.620281   6.085428   5.032  0.00023 ***
X1           1.562550   0.223406   6.994 9.42e-06 ***
X2          -0.010302   0.001841  -5.595 8.71e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.946 on 13 degrees of freedom
Multiple R-squared:  0.9088,Adjusted R-squared:  0.8947
F-statistic: 64.73 on 2 and 13 DF,  p-value: 1.744e-07
```

2.10 h) Rendement optimal

In [16]:
```
beta = unname(coefficients(qReg))

x_opt = -beta[2] / (2*beta[3])
x_opt
```

75.8404368269376