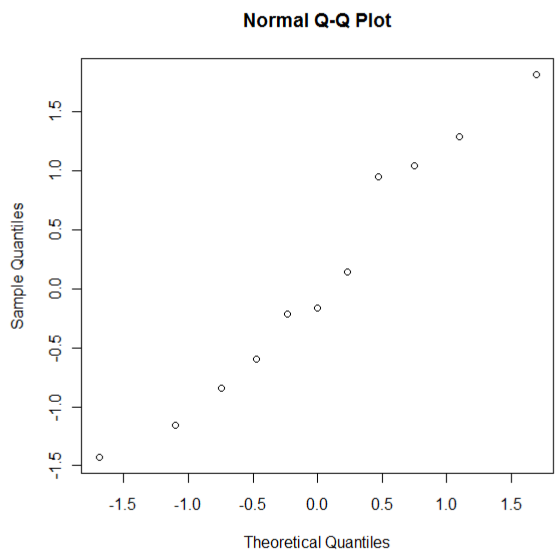
STAT 511 HW#4 SPRING 2009

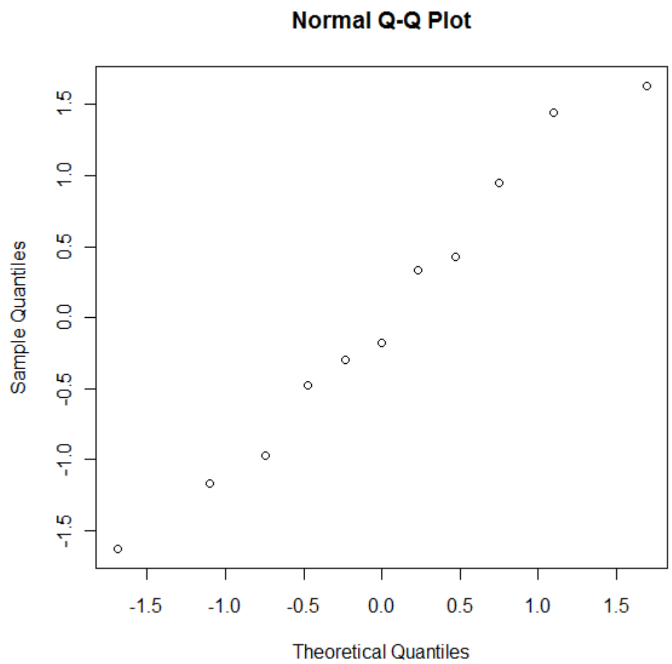
PROBLEM 1.

(a)

```
y<-c(62.9,21.5,19.3,64.1,32.9,30.9,93.4,64.6,58.6,26.6,51.3)
x1<-c(-1,-1,-1,0,0,0,1,1,1,-0.5,0.5)
x2<-c(-1,0,1,-1,0,1,-1,0,1,0.33,-0.33)
library(MASS)
model1<- lm(y~x1+x2)
summary(model1)
                  Estimate Std. Error t value Pr(>|t|)
      (Intercept)   47.827      3.300  14.494 5.03e-07 ***
      x1            18.341      4.299   4.267  0.00274 **
      x2           -18.286      4.395  -4.161  0.00316 **
      Residual standard error: 10.94 on 8 degrees of freedom
      Multiple R-squared: 0.824,      Adjusted R-squared:  0.78
      F-statistic: 18.73 on 2 and 8 DF,  p-value: 0.000959
qqnorm(stdres(model1))
```




```
model2<-lm(y~x1+x2+I(x1^2)+I(x2^2)+I(x1*x2))
summary(model2)
                  Estimate Std. Error t value Pr(>|t|)
      (Intercept)   33.752      2.006  16.826 1.36e-05 ***
      x1            18.341      1.362  13.467 4.04e-05 ***
      x2           -18.286      1.392 -13.132 4.57e-05 ***
      I(x1^2)       10.152      2.325   4.367  0.00725 **
      I(x2^2)       14.399      2.219   6.489  0.00130 **
      I(x1 * x2)     2.082      1.727   1.205  0.28198
      Residual standard error: 3.467 on 5 degrees of freedom
      Multiple R-squared: 0.989,      Adjusted R-squared: 0.9779
      F-statistic: 89.58 on 5 and 5 DF,  p-value: 6.875e-05
qqnorm(stdres(model2))
```



b)

```
anova(model1,model2)
      Analysis of Variance Table
      Model 1: y ~ x1 + x2
      Model 2: y ~ x1 + x2 + I(x1^2) + I(x2^2) + I(x1 * x2)
        Res.Df    RSS Df Sum of Sq      F   Pr(>F)
      1      8 958.22
      2      5  60.12  3    898.11 24.899 0.001963 **
```

A quadratic curvature in response (as a function of x's) appears to be statistically detectable.

c)

The local minimum is achived at (x1,x2)=(-0.9756637 , 0.7055119)

d)

For model 1 (x1,x2)=(-1,1) and for model 2 (x1,x2)=(-0.9756637 , 0.7055119)

```
new.x1<-c(-1,1)
x01<-c(1,new.x1)
y01<-sum(model1$coef*x01)
X1<-cbind(rep(1,length(x1)),x1,x2)
MSE1<-anova(model1)$Mean[length(anova(model1)$Mean)]
cXXc1<-x01%*%ginv(t(X1)%*%X1)%*%x01
##lower and upper confidence limits
cll1<-y01-qt(0.95,model1$df.residual)*sqrt(MSE1*cXXc1)
cul1<-y01+qt(0.95,model1$df.residual)*sqrt(MSE1*cXXc1)
```

(-1.510036 ,23.91091) however time cannot be negative, therefore we choose (0,23.91091)

```
##lower and upper prediction limits
pll1<-y01-qt(0.95,model1$df.residual)*sqrt(MSE1*(1+cXXc1))
pul1<-y01+qt(0.95,model1$df.residual)*sqrt(MSE1*(1+cXXc1))
```

(-12.79413, 35.195) but time cannot be negative therefore we choose (0, 35.195)

```
new.x2<-c(-0.9756637 , 0.7055119)
```

```
x02<-c(1,new.x2,new.x2^2,prod(new.x2))
y02<-sum(model2$coef*x02)
X2<-cbind(rep(1,length(x1)),x1,x2,x1^2,x2^2,x1*x2)
MSE2<-anova(model2)$Mean[length(anova(model2)$Mean)]
cXXc2<-x02%*%ginv(t(X2)%*%X2)%*%x02
##lower and upper confidence limits
cll2<-y02-qt(0.95,model2$df.residual)*sqrt(MSE2*cXXc2)
cul2<-y02+qt(0.95,model2$df.residual)*sqrt(MSE2*cXXc2)
```
(13.50563 , 23.20289)

```
##lower and upper prediction limits
pll2<-y02-qt(0.95,model2$df.residual)*sqrt(MSE2*(1+cXXc2))
pul2<-y02+qt(0.95,model2$df.residual)*sqrt(MSE2*(1+cXXc2))
```
(9.849634, 26.85888)

PROBLEM2.

```
X1<-c(x1,0,0)
X2<-c(x2,0,0)
Y<-c(y,30,36)
```

- Using functions that are build in R
```
cells<-lm(Y~as.factor(X1)*as.factor(X2))
model2<-lm(Y~X1+X2+I(X1^2)+I(X2^2)+I(X1*X2))
anova(cells, model2)
  Analysis of Variance Table

  Model 1: Y ~ as.factor(X1) * as.factor(X2)
  Model 2: Y ~ X1 + X2 + I(X1^2) + I(X2^2) + I(X1 * X2)
    Res.Df     RSS Df Sum of Sq      F Pr(>F)
  1      2  18.007
  2      7  78.793 -5   -60.787 1.3503 0.4772
```
- Using matrix calculations
```
project<-function(X){X%*%ginv(t(X)%*%X)%*%t(X)}
X<-cbind(rep(1,length(X1)),X1,X2,X1^2,X2^2,X1*X2)
Xstar<-matrix(0,dim(X)[1],11)
Xstar[1:11,1:11]<-diag(rep(1,11))
Xstar[12:13,5]<-1
F.ratio<-((t(Y)%*%(project(Xstar)-project(X))%*%Y)/(11-
6))/(t(Y)%*%(diag(rep(1,dim(Xstar)[1]))-project(Xstar))%*%Y/(13-11))
p.value<-1-pf(F.ratio,11-6,13-11)
```

There is not enough evidence of lack of fit.

PROBLEM3

```
y<-c(12,13,14,20,8,10,6,7,8,10,11,13,7)
A<-c(1,1,1,1,2,2,2,2,2,2,3,3,3,3)
B<-c(1,2,2,3,1,2,3,3,3,1,1,2,3)
```

a)

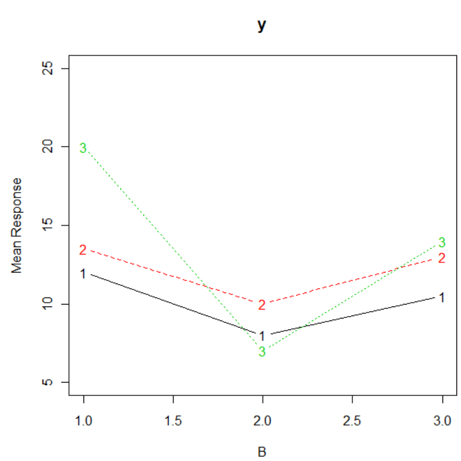
```
d<-data.frame(y,A,B)
d
       y A B
    1  12 1 1
    2  13 1 2
    3  14 1 2
    4  20 1 3
    5   8 2 1
    6  10 2 2
    7   6 2 3
    8   7 2 3
    9   8 2 3
    10 10 3 1
    11 11 3 1
    12 13 3 2
    13  7 3 3
```

b)

```
d$A<-as.factor(d$A)
d$B<-as.factor(d$B)
means<-tapply(d$y,list(d$A,d$B),mean)
means
         1    2  3
    1 12.0 13.5 20
    2  8.0 10.0  7
    3 10.5 13.0  7
```

c)

```
x.axis<-unique(d$B)
matplot(c(1,3),c(5,25),type="n",xlab="B",ylab="Mean Response",main="y")
matlines(x.axis,means,type="b")
```



d)

```
options(contrasts=c("contr.sum","contr.sum"))
lm.out1<-lm(y~A*B,data=d)
summary.aov(lm.out1,ssType=1)
               Df  Sum Sq Mean Sq F value    Pr(>F)
   A            2 108.469  54.235 72.3128 0.0007243 ***
   B            2   2.759   1.380  1.8395 0.2713371
   A:B          4  60.541  15.135 20.1802 0.0064769 **
   Residuals    4   3.000   0.750
lm.out2<-lm(y~B*A,data=d)
summary.aov(lm.out2,ssType=1)
               Df Sum Sq Mean Sq F value   Pr(>F)
   B            2 19.819   9.910  13.213 0.017284 *
   A            2 91.409  45.705  60.940 0.001010 **
   B:A          4 60.541  15.135  20.180 0.006477 **
   Residuals    4  3.000   0.750
summary.aov(lm.out1,ssType=3)
               Df  Sum Sq Mean Sq F value    Pr(>F)
   A            2 108.469  54.235 72.3128 0.0007243 ***
   B            2   2.759   1.380  1.8395 0.2713371
   A:B          4  60.541  15.135 20.1802 0.0064769 **
   Residuals    4   3.000   0.750
summary.aov(lm.out2,ssType=3)
               Df Sum Sq Mean Sq F value   Pr(>F)
   B            2 19.819   9.910  13.213 0.017284 *
   A            2 91.409  45.705  60.940 0.001010 **
   B:A          4 60.541  15.135  20.180 0.006477 **
   Residuals    4  3.000   0.750
```

e)

```
##Using sum restriction
Xeffects <- model.matrix(lm.out1)
# Type I
ssA<-t(d$y)%*%(project(Xeffects[,1:3])-project(Xeffects[,1]))%*%d$y
ssB<-t(d$y)%*%(project(Xeffects[,1:5])-project(Xeffects[,1:3]))%*%d$y
ssAB<-t(d$y)%*%(project(Xeffects)-project(Xeffects[,1:5]))%*%d$y
# Type II
ssA<-t(d$y)%*%(project(Xeffects[,1:5])-project(Xeffects[,c(1,4,5)]))%*%d$y
ssB<-t(d$y)%*%(project(Xeffects[,1:5])-project(Xeffects[,1:3]))%*%d$y
ssAB<-t(d$y)%*%(project(Xeffects)-project(Xeffects[,1:5]))%*%d$y
# Type III
ssA<-t(d$y)%*%(project(Xeffects)-project(Xeffects[,-c(2:3)]))%*%d$y
ssB<-t(d$y)%*%(project(Xeffects)-project(Xeffects[,-c(4:5)]))%*%d$y
ssAB<-t(d$y)%*%(project(Xeffects)-project(Xeffects[,-c(6:9)]))%*%d$y
```

	Type I SS	Type II SS	Type III SS
A	108.4692	91.40926	92.16279
B	2.759259	2.759259	7.269767
A:B	60.54074	60.54074	60.54074

```
##Using SAS restrictions
options(contrasts=c("contr.SAS","contr.SAS"))
Xsas=model.matrix(~A*B,data=d)
##Now use same code as before but replace Xeffects with Xsas and the same SS are obtained.
```

f)

```
Xcells<-matrix(0,13,9)
Xcells[1,1]<-1
Xcells[2:3,2]<-1
Xcells[4,3]<-1
Xcells[5,4]<-1
Xcells[6,5]<-1
Xcells[7:9,6]<-1
Xcells[10:11,7]<-1
Xcells[12,8]<-1
Xcells[13,9]<-1
Xincomp.full<-Xcells[-c(10:11),-7]
Xincomp.red<-Xeffects[-c(10,11),1:5]
Y.incomp<-d$y[-c(10:11)]
full<-lm(Y.incomp~Xincomp.full-1)
reduced<-lm(Y.incomp~Xincomp.red-1)
anova(reduced,full)
Analysis of Variance Table

Model 1: Y.incomp ~ Xincomp.red - 1
Model 2: Y.incomp ~ Xincomp.full - 1
  Res.Df    RSS Df Sum of Sq      F  Pr(>F)
1      6 60.146
2      3  2.500  3    57.646 23.058 0.01421
```

There is enough evidence to show that there is interaction between factors A and B.

g)

This problem ask for $\mu + \alpha_1^* + \beta_3^*$ which is equivalent to estimate $\mu + \alpha_1 - \beta_1 - \beta_2$

In that case

```
C<-c(1,1,0,-1,-1)
CB<-C%*%ginv(t(Xincomp.red)%*%Xincomp.red)%*% t(Xincomp.red)%*%Y.incomp
```

An OLS for this quantity is **14.68293**, however this is not a missing cell.

If you wanted to estimate a mean response for the missing cell you should estimate $\mu + \alpha_3^* + \beta_1^*$ or $\mu - \alpha_1 - \alpha_2 - +\beta_1$

Then

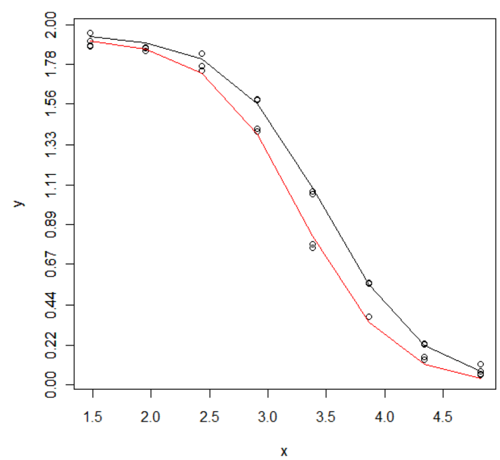
```
C<-c(1,-1,-1,1,0) ## and use the same CB as before
```

The estimated mean respose for the missing cell is **8.317073.**

PROBLEM 4

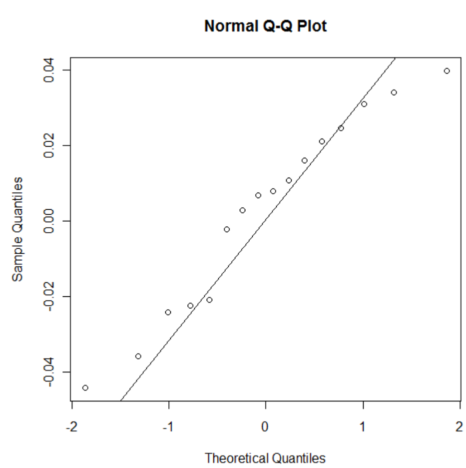
```
d<-rep(c(1/30,1/90,1/270,1/810,1/2430,1/7290,1/21870,1/65610),2)
y1<-
c(1.909,1.856,1.838,1.579,1.057,0.566,0.225,0.072,1.956,1.876,1.841,1.584,1.072,0.56
1,0.229,0.114)
y2<-
c(1.886,1.853,1.747,1.424,0.781,0.377,0.153,0.053,1.880,1.870,1.772,1.406,0.759,0.37
6,0.138,0.058)
x<- -log10(d)
```

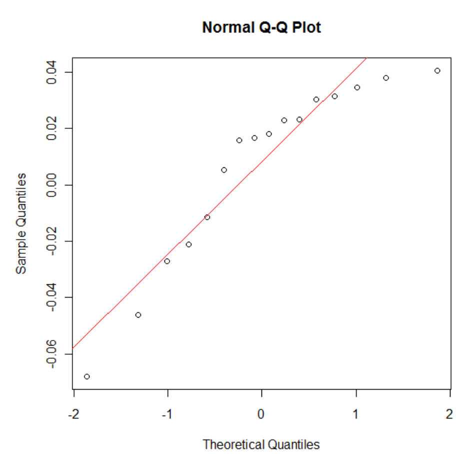
```
yrange<-range(y1,y2)
plot(range(x),yrange,type="n",xlab="x",ylab="y",yaxt="n")
axis(2,at=round(seq(0,2,length=10),2))
points(x,y1)
points(x,y2)
##I used the following eye estimates for both months
inicial<-c(theta1=1.9,theta2=2.4,theta3=3.5)
may<-nls(y1~theta1/(1+exp(theta2*(x-theta3))),start=inicial)
june<- nls(y2~theta1/(1+exp(theta2*(x-theta3))),start=inicial)
new<-data.frame(x,y1,y2,predict(may),predict(june))
fitted<-new[do.call(order,new),]
lines(fitted[,1],fitted[,4],col=1)
lines(fitted[,1],fitted[,5],col=2)
```



It seems that may mean function is right shifted for that of june.

```
qqnorm(fitted[,2]-fitted[,4])
qqline(fitted[,2]-fitted[,4],col=1)
qqnorm(fitted[,3]-fitted[,5])
qqline(fitted[,3]-fitted[,5],col=2)
```





May residuals look fine. June residuals have a little departure from a straight line.

b)

```
summary(may)
 Parameters:
        Estimate Std. Error t value Pr(>|t|)
 theta1  1.94616    0.01518  128.22  < 2e-16 ***
 theta2  2.42775    0.07133   34.04 4.29e-14 ***
 theta3  3.48749    0.01452  240.25  < 2e-16 ***
 Residual standard error: 0.02773 on 13 degrees of freedom
round(vcov(may),5)
          theta1   theta2   theta3
 theta1  0.00023 -0.00064 -0.00014
 theta2 -0.00064  0.00509  0.00040
 theta3 -0.00014  0.00040  0.00021
theta1<-coef(may)[1]
theta2<-coef(may)[2]
theta3<-coef(may)[3]
Dhat<-matrix(cbind(1/(1+exp(theta2*(x-theta3))),-theta1*(x-theta3)*exp(theta2*(x-
theta3))/((1+exp(theta2*(x-theta3)))^2),theta2* theta1 *exp(theta2*(x-theta3))
/((1+exp(theta2*(x-theta3)))^2)),nrow=16)
res<-fitted[,2]-fitted[,4]
MSE<-(t(res)%*%res)/(length(y1)-length(coef(may)))
covtheta<-round(c(MSE)*solve(t(Dhat)%*%Dhat),5)
covtheta
          [,1]     [,2]     [,3]
 [1,]  0.00023 -0.00064 -0.00014
 [2,] -0.00064  0.00509  0.00040
 [3,] -0.00014  0.00040  0.00021
Same as variance covariance matrix obtained using vcov function.
stderr<-round(sqrt(diag(covtheta)),5)
stderr
   [1] 0.01517 0.07134 0.01449
##Confidence limits
x.new<- -log10(1/270)
yhat<-theta1/(1+exp(theta2*(x.new-theta3)))
Ghat<- cbind(1/(1+exp(theta2*(x.new-theta3))),-theta1*(x.new-
theta3)*exp(theta2*(x.new-theta3))/((1+exp(theta2*(x.new-theta3)))^2),theta2* theta1
*exp(theta2*(x.new-theta3)) /((1+exp(theta2*(x.new-theta3)))^2))
ll<-yhat-qt(0.975,16-3)* sqrt(c(MSE)%*%Ghat%*%solve(t(Dhat)%*%Dhat)%*%t(Ghat))
ul<-yhat+qt(0.975,16-3)* sqrt(c(MSE)%*%Ghat%*%solve(t(Dhat)%*%Dhat)%*%t(Ghat))
   (1.785306 , 1.828742)
##Prediction limits
ll<-yhat-qt(0.975,16-3)* sqrt(c(MSE)%*%(1+Ghat%*%solve(t(Dhat)%*%Dhat)%*%t(Ghat)))
ul<-yhat+qt(0.975,16-3)* sqrt(c(MSE)%*%(1+Ghat%*%solve(t(Dhat)%*%Dhat)%*%t(Ghat)))
   (1.743293 , 1.870756)
```

c)

We know that $\theta_1$is the limiting mean for small x, then the dilution at which may mean optical density reaches 25% of its limiting value is $\theta_3 + \frac{\log(3)}{\theta_2}$.

```
point.estim<-theta3+log(3)/theta2
Ghat<-matrix(c(0,-log(3)/theta2^2,1),nrow=1)
estdev<- sqrt(Ghat%*%covtheta%*%t(Ghat))
```

A point estimate for this quantity is **3.940012** and an estimated standard deviation is **0.01541843.**

d)

```
##95% CI for May parameters
se<-sqrt(diag(covtheta))
theta1.CI<-c(theta1-qt(.975,13)*se[1], theta1+qt(.975,13)*se[1])
theta2.CI<-c(theta2-qt(.975,13)*se[2], theta2+qt(.975,13)*se[2])
theta3.CI<-c(theta3-qt(.975,13)*se[3], theta3+qt(.975,13)*se[3])
   theta1 1.913394 1.978921
   theta2 2.273617 2.581877
   theta3 3.456182 3.518795

##Using confint
confint(may,level=.95)
              2.5%    97.5%
   theta1 1.914375 1.979179
   theta2 2.280051 2.586947
   theta3 3.456286 3.518130
```

Both methods produce intervals that are close.

e)

Using the linear model result a $(1-\alpha)\%$ confidence interval for σ is exercise

$$\left(\sqrt{\frac{SSE}{\chi^2_{n-k,1-\alpha/2}}},\sqrt{\frac{SSE}{\chi^2_{n-k,\alpha/2}}}\right)$$

```
sigma.CI<-c(sqrt(13*MSE/qchisq(.975,13)), sqrt(13*MSE/qchisq(.025,13)))
```

That is **(0.02010631 , 0.04468163)**

Based on the profile likelihood a $(1-\alpha)\%$ confidence interval for σ is

$$\left\{\sigma^2 : \log\sigma^2 + \frac{SSE}{n\sigma^2} - \left(\log\frac{SSE}{n} + 1\right) < \frac{1}{n}\chi^2_{1,0.095}\right\}$$

```
sigma2<-seq(0.000001,10,0.000001)
l<-sigma2[log(sigma2)+(13*MSE)/(16*sigma2)-(log(13*MSE/16)+1)<=1/16*qchisq(.95,1)]
sigma.CI<- sqrt (c (min(l),max(l)))
```

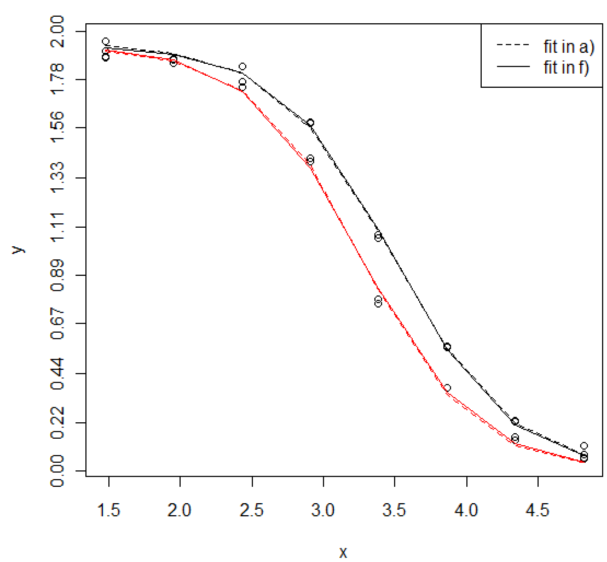
A confidence interval for σ is **(0.01833030 , 0.03697296)**

f)

```
Y<-c(y1,y2)
X<-c(x,x)
J<-c(rep(0,16),rep(1,16))
inicial<-c(theta1=1.9,theta2=2.4,theta3=3.5,theta4=0)
```

```
new.model<-nls(Y~theta1/(1+exp(theta2*(X-(theta3+theta4*J)))),start=inicial)
theta1<-coef(new.model)[1]
theta2<-coef(new.model)[2]
theta3<-coef(new.model)[3]
theta4<-coef(new.model)[4]

plot(range(X),range(Y),type="n",xlab="x",ylab="y",yaxt="n")
axis(2,at=round(seq(0,2,length=10),2))
points(X,Y)
mayo<-theta1/(1+exp(theta2*(x-theta3)))
junio<- theta1/(1+exp(theta2*(x-theta3-theta4)))
new<-data.frame(x,y1,y2,mayo,junio)
fitted<-new[do.call(order,new),]
lines(fitted[,1],fitted[,4],col=1)
lines(fitted[,1],fitted[,5],col=2)
new<-data.frame(x,y1,y2,predict(may),predict(june))
fitted<-new[do.call(order,new),]
lines(fitted[,1],fitted[,4],col=1,lty=2)
lines(fitted[,1],fitted[,5],col=2,lty=2)
legend("topright",legend=c("fit in a)","fit in f)"),lty=c(2,1))
```



The new model fit as well as the previous one.

```
point.estim<-10^theta4
covtheta<- vcov(new.model),
Ghat<-matrix(c(0,0,0,10^theta4*log(10)),nrow=1)
estdev<- sqrt(Ghat%*%covtheta%*%t(Ghat))
CI<-c(point.estim-qt(0.975,32-4)*estdev, point.estim+qt(0.975,32-4)*estdev)
```

A point estimate for the relative potency is **0.598607** and an estimated standard deviation is **0.02408688.** Therefore a 95% confidence interval for this quantity is **(0.5492673 , 0.6479467).**