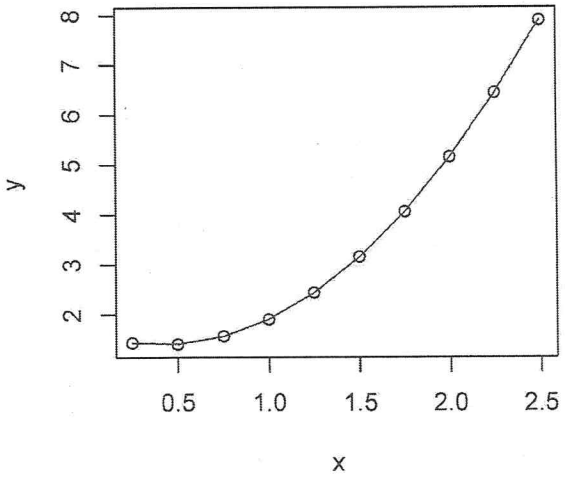
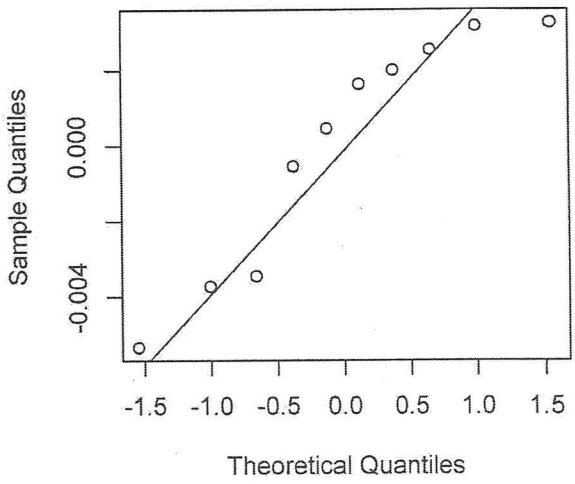
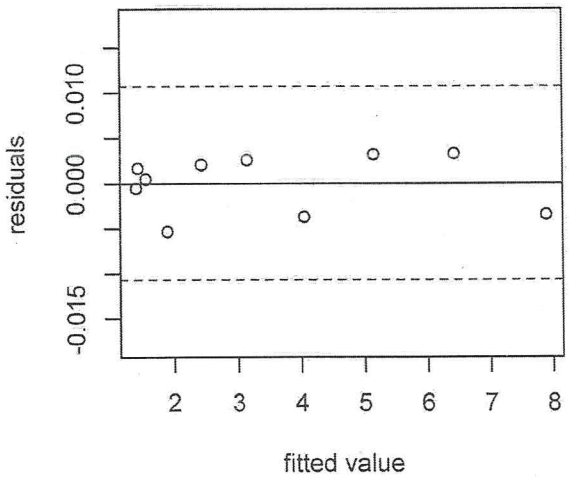
7.2 a. $\hat{y} = 1.633 - 1.232x + 1.495x^2$

b. $F_{obs} = 1.859 \times 10^6$, P-value $< 2.2 \times 10^{-16}$

Since the p-value is practically zero, the regression is significant.







c. Test $H_0: \beta_2 = 0$ vs $H_1: \beta_2 \neq 0$

$t_{obs} = 601.6$, P-value $< 2 \times 10^{-16}$

There is very strong evidence that $\beta_2 \neq 0$. Hence the quadratic term is needed in the model.

d. Yes. This is because the model is a quadratic model and has all hazards associated with extrapolating a polynomial model.

(R code at the back).

7.11 (4th); 7.13 (5th)

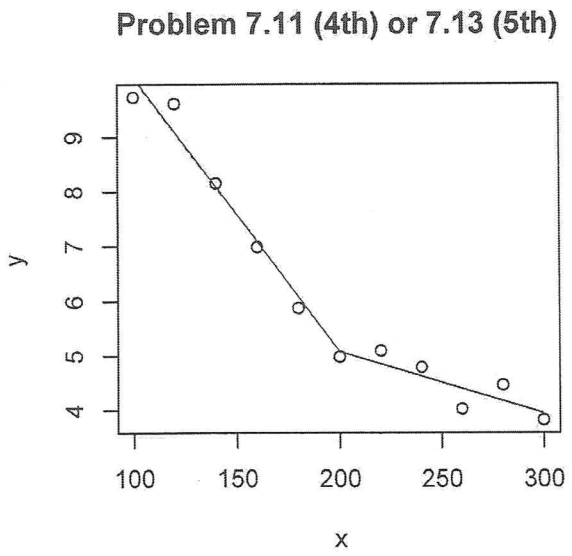
a. $\hat{y} = 15.116 - 0.050x + 0.039(x-200)_+^1$

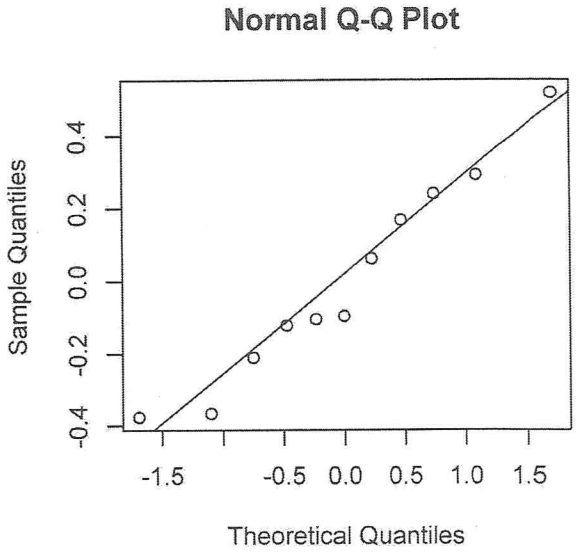
b. Test $H_0: \beta_{11} = 0$ vs $H_1: \beta_{11} \neq 0$

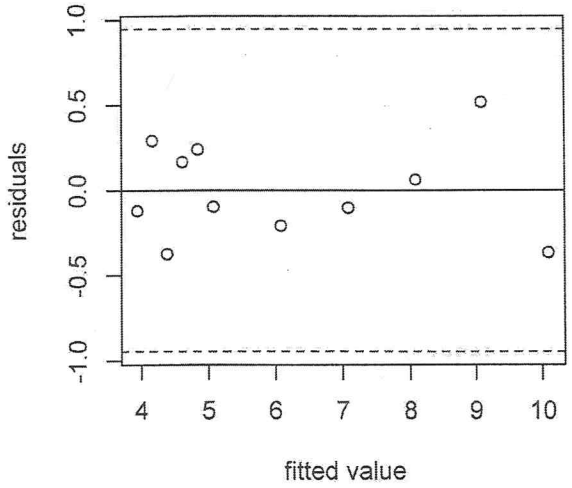
$t_{obs} = 6.534$, P-value $= 0.000$

There is very strong evidence that $\beta_{11} \neq 0$, so the slope changes at $x = 200$.

c.







(R code at the back)

d. there are no problems with the model from the residual analysis. the model in part a is adequate.

Question 3 9.6 (4th) or 10.6 (5th)

a. The model selected by stepwise regression is

$$\hat{Y} = 13.51 + 2.42 x_1 + 8.48 x_2 + 2.00 x_5 - 2.18 x_7$$

b. Backward selection gives the same model as in a.

c. $H_0: \beta_5 = \beta_7 = 0$ vs H_1: At least one of β_5, β_7 is not zero.

Partial F-statistic: $F_{obs} = 1.88$

P-value $= 0.1792$

There is no evidence against H_0

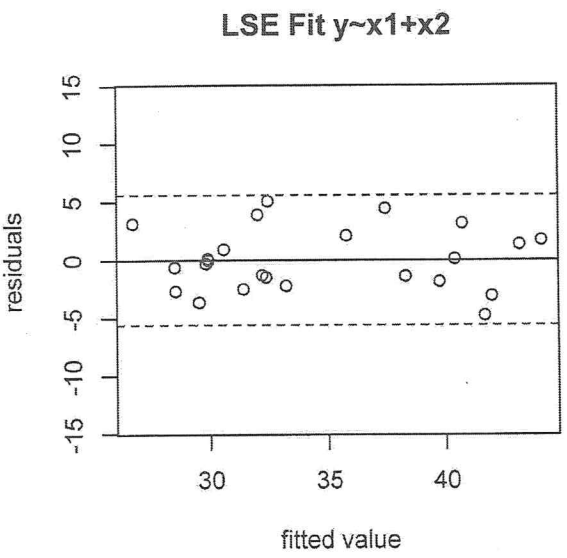
d. Residual analysis for model $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$

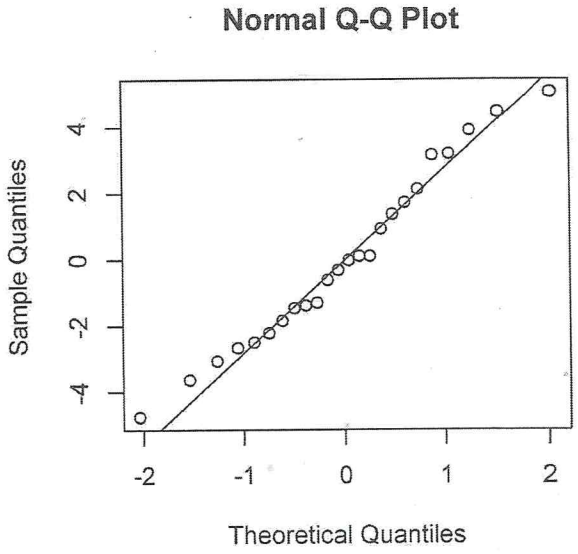
Estimated model: $\hat{Y} = 10.04 + 2.71 x_1 + 6.16 x_2$

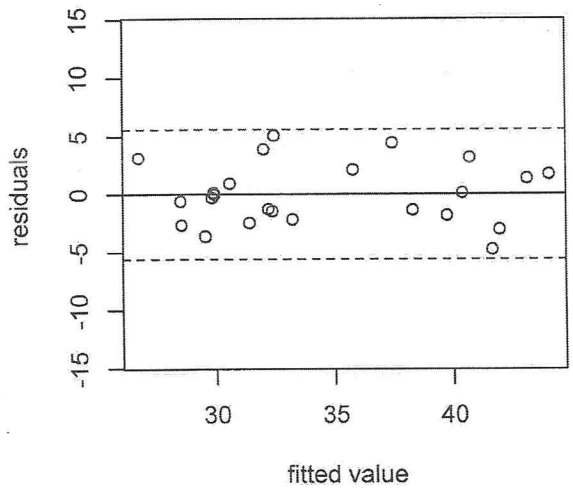
Residual plots find no outliers and no violations of normality assumption. The model is adequate.

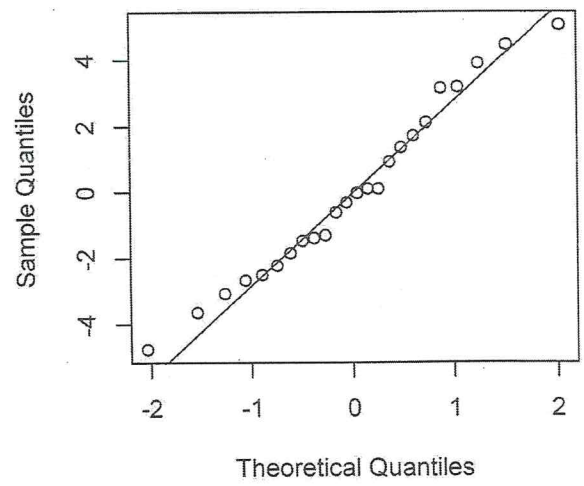
(R code at the back)

Question 3 9.6 (4th) ; 10.6 (5th).









```
# Assignment 4 Question 1 (Problem 7.2)

rm(x,y)
data1=read.table(file="data-prob-7-2.prn", header=TRUE)
attach(data1)
print(data1)

x2=x^2

full=lm(y~x+x2)
summary(full)

resid=full$residuals
fitted=y-resid

par(mfrow=c(2,2))
plot(x,y, main="Problem 7.2")
lines(x,fitted, col="blue")

qqnorm(resid)
qqline(resid)
shapiro.test(resid)

s=summary(full)$sigma
plot(fitted,resid,xlab="fitted value", ylab="residuals",
ylim=c(-5*s,5*s))
abline(h=0)
abline(h=-3*s,lty=2,col="blue")
abline(h=3*s,lty=2,col="blue")

#
# Assignment 4 Question 2 (Problem 7.11 (4th) 7.13 (5th))
#

rm(x,y)
data1=read.table(file="data-prob-7-11.prn", header=TRUE)
attach(data1)
print(data1)

x1=pmax(x-200,0)   #positive function at knot x=200

fit=lm(y~x+x1)
summary(fit)

resid=fit$residuals
fitted=y-resid

par(mfrow=c(2,2))
plot(x,y, main="Problem 7.11 (4th) or 7.13 (5th)")
lines(x,fitted, col="blue")

qqnorm(resid)
qqline(resid)
shapiro.test(resid)

s=summary(fit)$sigma
plot(fitted,resid,xlab="fitted value", ylab="residuals",
```

```
ylim=c(-3*s,3*s))
abline(h=0)
abline(h=-3*s,lty=2,col="blue")
abline(h=3*s,lty=2,col="blue")

#
# Assignment 4 Question 3 Problem 9.6 (4th) 10.6 (5th)
#

rm(y,x1,x2,x3,x4,x5,x6,x7,x8,x9)
data1=read.table(file="data-table-B4.prn", header=TRUE)
attach(data1)
print(data1)


### [a] Stepwise regressions, forward and backward selection

model1=lm(y~., data=data1)
step1=step(model1, scope=list(upper=~x1+x2+x3+x4x5+x6+x7+x8
+x9, lower=~1))
summary(step1)


### [b] Backward selection

step2=step(model1, direction="backward")
summary(step2)


### [c] Final model y~x1+x2+x5+x7

fit0=lm(y~x1+x2+x5+x7, data=data1)  # selected full model
fit1=lm(y~x1+x2, data=data1)
anova(fit1,fit0)


### [d] Check model adequacy for y!x1+x2

resid1=fit1$residuals
fitted1=y-resid1
par(mfrow=c(2,2))
s1=summary(fit1)$sigma
plot(fitted1,resid1,main="LSE Fit for y~x1+x2",xlab="fitted
value",
ylab="residuals",ylim=c(-5*s1,5*s1))
abline(h=0)
abline(h=2*s1,lty=2,col="blue")
abline(h=-2*s1,lty=2,col="blue")
qqnorm(resid1)
qqline(resid1)
shapiro.test(resid1)
summary(fit1)

fit2=lm(y~x1+x2, data=data1)
resid2=fit2$residuals
fitted2=y-resid2
s2=summary(fit2)$sigma
```

```
plot(fitted2,resid2,main="Robust Fit for y~x1+x2",xlab="fitted
value",
ylab="residuals",ylim=c(-5*s1,5*s1))
abline(h=0)
abline(h=2*s2,lty=2,col="blue")
abline(h=-2*s2,lty=2,col="blue")
qqnorm(resid2)
qqline(resid2)
shapiro.test(resid2)
summary(fit2)
```