Chapter 10 Question 10.

Solution 1. *we copy the data into a .csv file and read it:*

```
#read the data
setwd("C:\\Akira\\R")
devia = read.csv("devia.csv", header = TRUE)
```

For part (a):

We are fitting a block-treatment model, so we do not consider interaction. Also since each treatment is only applied once in each block, so our model is a randomized complete block design.

$$Y_{hi} = \mu + \theta_h + \tau_i + \epsilon_{hi}$$

with $1 \leq h \leq 14, 1 \leq i \leq 6$, and $\epsilon_{hi} \sim N(0, \sigma^2)$ *i.i.d.*

we fit the data and check error assumption as follows:

```
fit <- lm(response ~ factor(subject) + factor(treatment), data = devia)
anova(fit)

## Analysis of Variance Table
##
## Response: response
##                   Df  Sum Sq Mean Sq F value    Pr(>F)
## factor(subject)   13 29.2687 2.25144 12.4936 4.261e-13 ***
## factor(treatment)  5  3.5482 0.70964  3.9379  0.003515 **
## Residuals         65 11.7135 0.18021
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

To check the error assumption:

```
SSE <- sum((devia$response - fit$fitted.values)^2)
Z = fit$residuals/sqrt(SSE/83)
par(mfrow = c(2, 1))
cols = factor(devia$treatment)
plot(fit$fitted.values, Z, col=as.character(cols))
title(main = "std residuals vs predicted values")
qqnorm(Z, col=as.character(cols))
```

std residuals vs predicted values
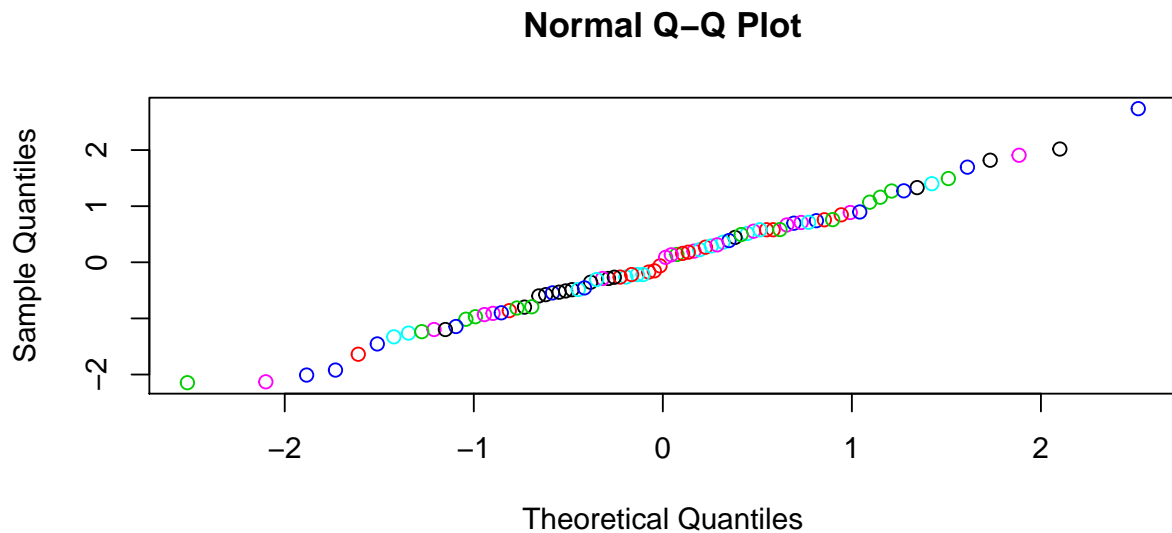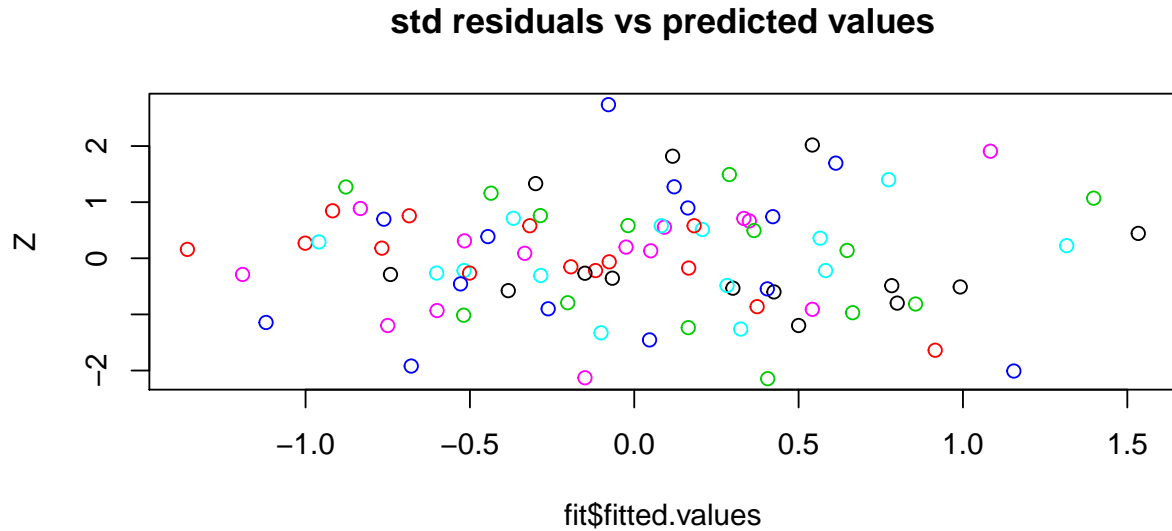
Normal Q–Q Plot

We can not check independence since there is no infor for run order. We plotted residuals against predicted value and do not see much pattern, so the equal variance assumption holds.

We plotted normal qq plot and the straight line pattern shows support for normality.

So in general we think the error assumptions are satisfied here.

For part (b):

We can check interaction between block and treatment factor here:

```r
library(ggplot2)
ggplot(data=devia, aes(x=factor(treatment), y = response, color = subject))+
  geom_line(aes(group=subject), linetype = "dotted") +geom_point()
```
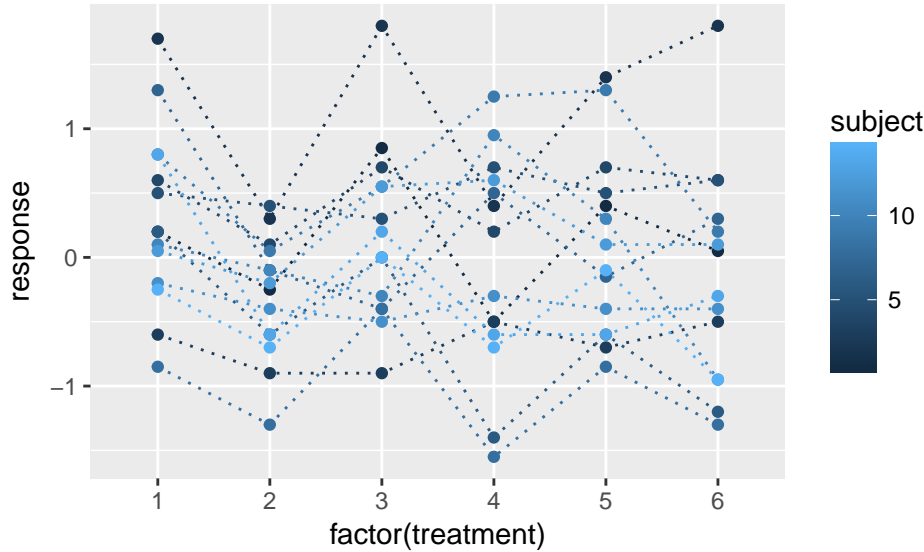

For prat (c):

I coded the treatment combinations as $1, 2, 3, 4, 5, 6$ while 1 for 11, 2 for 12, 3 for 21, 4 for 22, 5 for 31, 6 for 32. So in order to make it clear which treatment combination we are looking at, instead of saying τ_1, I would temporarily use double index as τ_{11}.

we look at the following contrasts:

1. differences in the effects of area for each shape separately. So the contrast is:

$$\tau_{11} - \tau_{12}, \ i.e. \ (1, -1, 0, 0, 0, 0)$$
$$\tau_{21} - \tau_{22}, \ i.e. \ (0, 0, 1, -1, 0, 0)$$
$$\tau_{31} - \tau_{32}, \ i.e. \ (0, 0, 0, 0, 1, -1)$$

2. average difference in the effects of area, so the contrast is:

$$\frac{1}{3}(\tau_{11} + \tau_{21} + \tau_{31}) - \frac{1}{3}(\tau_{12} + \tau_{22} + \tau_{32})$$
$$i.e. \ (\frac{1}{3}, -\frac{1}{3}, \frac{1}{3}, -\frac{1}{3}, \frac{1}{3}, -\frac{1}{3})$$

3. average difference in the effects of shape, so the contrast is:

$$\frac{1}{2}(\tau_{11} + \tau_{12}) - \frac{1}{2}(\tau_{21} + \tau_{22}) \ i.e. \ (\frac{1}{2}, \frac{1}{2}, -\frac{1}{2}, -\frac{1}{2}, 0, 0)$$
$$\frac{1}{2}(\tau_{11} + \tau_{12}) - \frac{1}{2}(\tau_{31} + \tau_{32}) \ i.e. \ (\frac{1}{2}, \frac{1}{2}, 0, 0, -\frac{1}{2}, -\frac{1}{2})$$
$$\frac{1}{2}(\tau_{21} + \tau_{22}) - \frac{1}{2}(\tau_{31} + \tau_{32}) \ i.e. \ (0, 0, \frac{1}{2}, \frac{1}{2}, -\frac{1}{2}, -\frac{1}{2})$$

For part (d):

   We use Scheffe's method to build our confidence intervals here:

```r
a<-c(1, -1, 0, 0, 0, 0)
b<-c(0, 0, 1, -1, 0, 0)
c<-c(0, 0, 0, 0, 1, -1)
contrast <- list(a, b, c)
n <- length(contrast)
msE <- anova(fit)[[3]][[3]]
msE

## [1] 0.180207

#compute \bar{Y}_{dot i}
y <- c()
for (i in 1:6){
y[i] <- mean(devia[devia$treatment==i,]$response)
}
#compute Scheffe critical coefficient
w_S <- sqrt(5*qf(0.01, 5,65, lower.tail = FALSE))
w_S

## [1] 4.069912

#compute 99% confidence interval
for (i in 1:n){
#compute least square estimate
lse <- sum(contrast[[i]]*y)
#compute msd
msd <- w_S*sqrt(msE*sum(contrast[[i]]^2)/14)
#compute lower and upper bound for CI
ci_lower<- lse-msd
ci_upper<- lse+msd
#print the result of CI
cat("The CI for (", contrast[[i]], ") is (",
    ci_lower<- lse-msd, ", ", ci_upper<- lse+msd, ")", "\n")
}

## The CI for ( 1 -1 0 0 0 0 ) is ( -0.03515575 ,  1.27087 )
## The CI for ( 0 0 1 -1 0 0 ) is ( -0.4101558 ,  0.89587 )
## The CI for ( 0 0 0 0 1 -1 ) is ( -0.42087 ,  0.8851558 )
```

   all three intervals include 0, which means we fail to reject the corresponding hypothesis tests. This is interpreted as, with 99% confidence, there is NO statistical significance for the claim that the effects of area is different for any/each separate shape.

For part (e):

(c)(ii) and (c)(iii) are considering the main effect of one treatment averaged over the other, so it is only of interest if the shape and area do not have interaction.

To test interaction, our full model is:

$$Y_{hij} = \mu + \theta_h + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{hij}$$

and reduced model is our block-treatment model.

```
fit <- lm(response ~ factor(subject) + factor(shape)*factor(area), data = devia)
anova(fit)

## Analysis of Variance Table
##
## Response: response
##                           Df  Sum Sq Mean Sq F value     Pr(>F)
## factor(subject)           13 29.2687 2.25144 12.4936 4.261e-13 ***
## factor(shape)              2  0.0859 0.04295  0.2383 0.7886381
## factor(area)               1  2.7868 2.78679 15.4644 0.0002068 ***
## factor(shape):factor(area) 2  0.6755 0.33777  1.8743 0.1616572
## Residuals                 65 11.7135 0.18021
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The p value for interaction test is $0.1616572 > 0.05 > 0.01$, so we retain the null hypothesis that there is NO interaction, so the above contrasts are of our interest here.