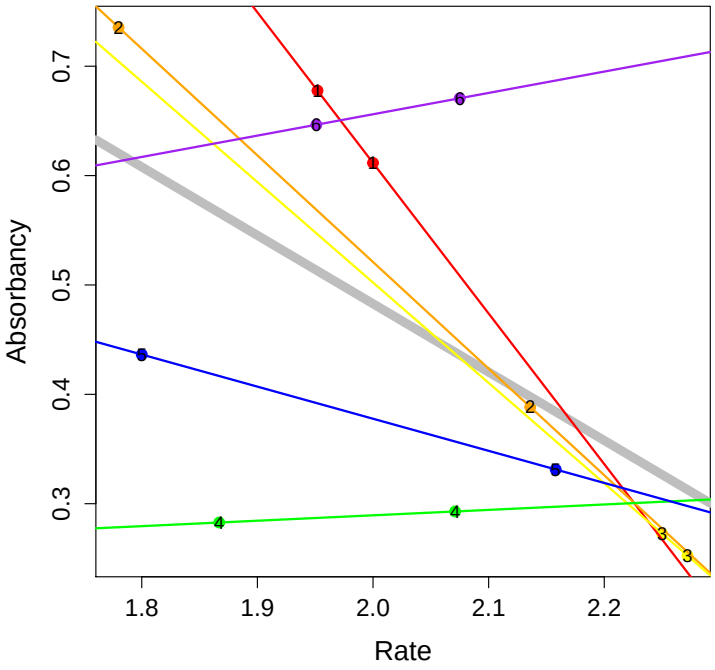
Chapter 9 Question 4.

Solution 1. *for part (a):*

```
#read the data
setwd("C:\\Akira\\R")
towel = read.csv("towel.csv", header = TRUE)
#plot absorbancy versus rate, using treatment level as the plotting symbol
#see whether between the treatments the absorbancy(y) vs rate(x) is different
cols = factor(towel$Treatment)
levels(cols) = c("red", "orange", "yellow", "green", "blue", "purple")
par(mar = c(5,5,4,2))
fit = lm(towel$Absorbancy~towel$Rate)
fit1 = lm(towel$Absorbancy~towel$Rate, subset = towel$Treatment ==1)
fit2 = lm(towel$Absorbancy~towel$Rate, subset = towel$Treatment ==2)
fit3 = lm(towel$Absorbancy~towel$Rate, subset = towel$Treatment ==3)
fit4 = lm(towel$Absorbancy~towel$Rate, subset = towel$Treatment ==4)
fit5 = lm(towel$Absorbancy~towel$Rate, subset = towel$Treatment ==5)
fit6 = lm(towel$Absorbancy~towel$Rate, subset = towel$Treatment ==6)
plot(towel$Rate, towel$Absorbancy, xlab = "Rate", ylab = "Absorbancy",
   cex.lab = 1.5, cex.axis = 1.3, cex = 1.3, pch = 19,
   col = as.character(cols))
abline(a = coef(fit)[[1]], b = coef(fit)[[2]], lwd = 8, col = "grey")
abline(a = coef(fit1)[[1]], b = coef(fit1)[[2]], lwd = 2, col = "red")
abline(a = coef(fit2)[[1]], b = coef(fit2)[[2]], lwd = 2, col = "orange")
abline(a = coef(fit3)[[1]], b = coef(fit3)[[2]], lwd = 2, col = "yellow")
abline(a = coef(fit4)[[1]], b = coef(fit4)[[2]], lwd = 2, col = "green")
abline(a = coef(fit5)[[1]], b = coef(fit5)[[2]], lwd = 2, col = "blue")
abline(a = coef(fit6)[[1]], b = coef(fit6)[[2]], lwd = 2, col = "purple")
text(towel$Rate, towel$Absorbancy, labels= towel$Treatment)
```



From the plot we find that the response of group 3 *and* 4 *stay much lower than that of group* 1 *and* 6*, which indicates possible treatment effects. Also the slope(rate of change) of response versus the covariate are very different between treatment groups (for example, group* 4 *and* 6 *have positive slopes and the others have negative slopes, and they do not seem to be parallel either). It is possible that the ancova model is not appropriate, but consider we have really small sample here to give enough power, we hold this judgement conservatively.*

For part (b):

Our one way ancova model here is:

$$Y_{it} = \mu + \tau_i + \beta(x_{it} - \bar{x}_{..}) + \epsilon_{it}$$

where $1 \leq i \leq 6$ *and* $1 \leq t \leq 2$*.* τ_i *represents the treatment effect of ith treatment.*

```
  fit <- lm(towel$Absorbancy ~ factor(towel$Treatment)+ towel$Rate)
  anova(fit)

## Analysis of Variance Table
##
## Response: towel$Absorbancy
##                           Df   Sum Sq  Mean Sq F value   Pr(>F)
## factor(towel$Treatment)  5 0.316145 0.063229 11.3128 0.009337 **
## towel$Rate               1 0.040549 0.040549  7.2549 0.043118 *
## Residuals                5 0.027946 0.005589
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

  summary(fit)

##
## Call:
## lm(formula = towel$Absorbancy ~ factor(towel$Treatment) + towel$Rate)
##
## Residuals:
##         1         2         3         4         5         6         7
##  0.083187 -0.056831 -0.083187  0.020866 -0.038321 -0.020866  0.056831
##         8         9        10        11        12
## -0.043575 -0.004516  0.004516  0.043575  0.038321
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 1.64778    0.37616   4.381  0.00715 **
## factor(towel$Treatment)2 -0.09184    0.07484  -1.227  0.27439
## factor(towel$Treatment)3 -0.23737    0.09206  -2.578  0.04952 *
## factor(towel$Treatment)4 -0.36035    0.07477  -4.819  0.00480 **
## factor(towel$Treatment)5 -0.25928    0.07476  -3.468  0.01789 *
## factor(towel$Treatment)6  0.03273    0.07509   0.436  0.68105
## towel$Rate                -0.50766    0.18847  -2.693  0.04312 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.07476 on 5 degrees of freedom
## Multiple R-squared:  0.9273,Adjusted R-squared:  0.8402
## F-statistic: 10.64 on 6 and 5 DF,  p-value: 0.01007
```

The code above revelas information with the following translation:

$$\mu = 1.64778, \beta = -0.50766, \tau_1 = 0, \tau_2 = -0.09184$$
$$\tau_3 = -0.23737, \tau_4 = -0.36035, \tau_5 = -0.25928, \tau_6 = 0.03273$$

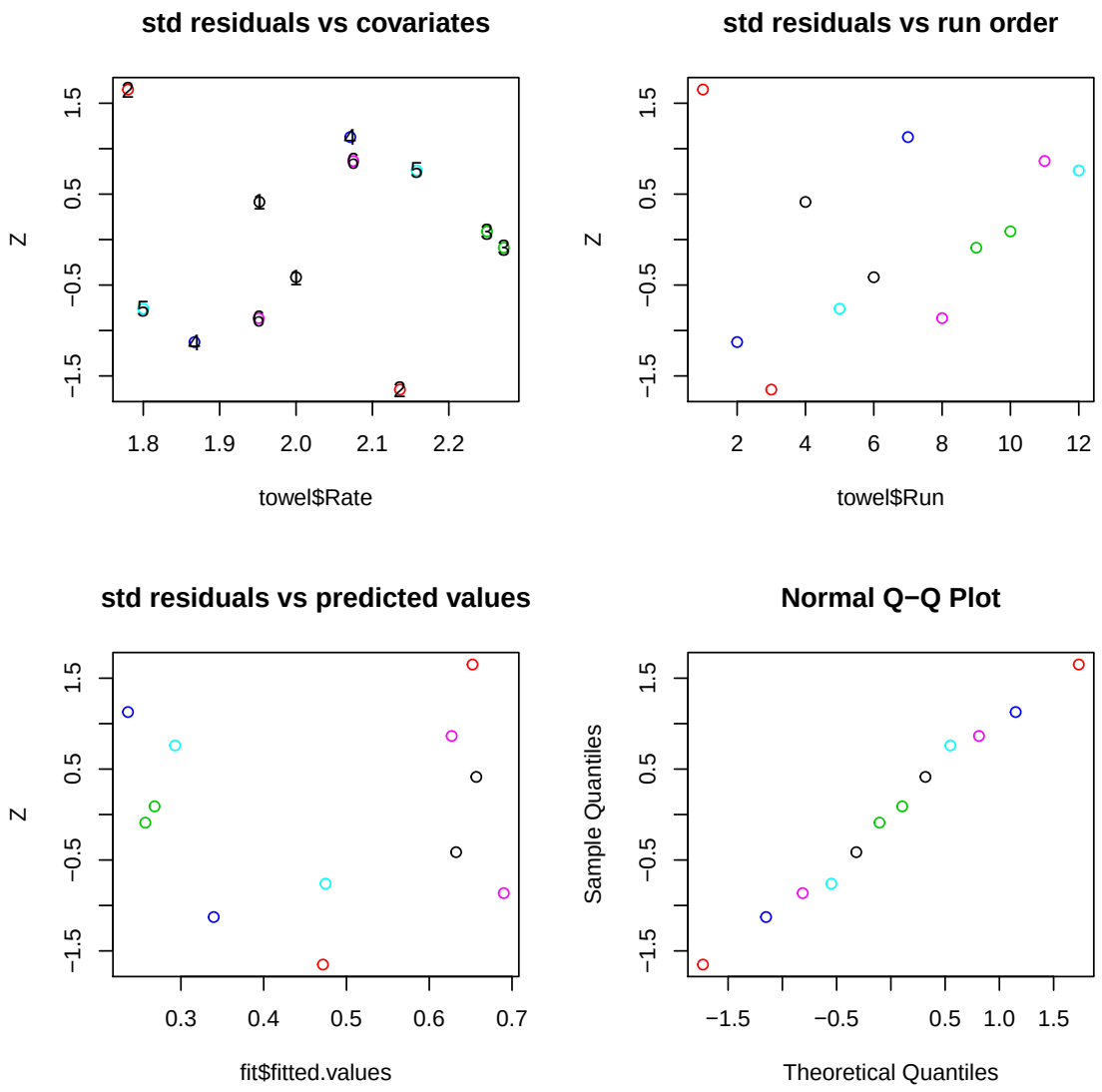
and the ANCOVA table is also given above.

For part (c):

We are plotting (standardized)residuals against the covariate, run order, predicted values, and normal scores to diagnose model assumptions.

```
#compute SSE
SSE <- sum((towel$Absorbancy - fit$fitted.values)^2)
#compute standardized residuals
```

```
Z = fit$residuals/sqrt(SSE/11)
cols = factor(towel$Treatment)
par(mfrow = c(2, 2))
plot(towel$Rate, Z, col = as.character(cols))
text(towel$Rate, Z, labels=towel$Treatment)
title(main = "std residuals vs covariates")
plot(towel$Run, Z, col=as.character(cols))
title(main = "std residuals vs run order")
plot(fit$fitted.values, Z, col=as.character(cols))
title(main = "std residuals vs predicted values")
qqnorm(Z, col=as.character(cols))
```



Let's comment on these graphs:

1. *for residuals against covariate, we actually need to look at the graph under each treatment level(so I differentiated the levels with different colors and labels) to decide if the linear relationship between response and covariate is really appropirate. However we only have two samples here within each treatment level, which has to be linear. But this may not necessarily true when we have more sample points within each treatment group. For now, let's consider it as appropriate.*

2. *for residuals against run order, there does not seem to be any trend, so the independence assumption holds*

3. *for residuals against predicted values, it seems the variance is more spread out to the sides and more condensed on the lower half in the middle, which suggests possibility of non-equal variance. But again, we have really small sample sizes here, so we hold this opinion conservatively.*

4. *Finally, the qq plot gives a very well approximated straightline, shows support for the most important normal assumption.*

For part (d):

To formally test the equality of slopes:

Our hypothesis is:

$$H_0 : \beta_i = \beta_j \text{ for all } i, j$$
$$H_a : \beta_i \neq \beta_j \text{ for at least some} i, j$$

Then our full model is:

$$y_{it} = \mu + \tau_i + \beta_i(x_{it} - \bar{x}_{..}) + \epsilon_{it}$$

and the reduced model is:

$$y_{it} = \mu + \tau_i + \beta(x_{it} - \bar{x}_{..}) + \epsilon_{it}$$

To run the formal test, we should compute:

$$ss\beta = ssE_{reduced} - ssE_{full}$$

and find our F statisitc as:

$$F = \frac{ms\beta}{msE_{full}}$$

But if we look at the full model, the degree of freedom for ssE is $12 - 6 - 6 = 0$. *This happens because within each group we have only two samples, and the linear model is gonna be a perfect fit and creates* 0 *residuals.*

So unfortunately we could not formally test this hypothesis.

But back to the graph of (a) *or the first graph of* (c)*, it is clear that the slope between treatment groups are different, so we tend to reject* H_0*. However we should still keep in mind this is a small sample case and the opinion should be held conservatively.*

For part (e)*:*

If we look at the ANCOVA table in part (b)*, given the treatment factors(reduced model is ANOVA model), we see that the covariate does take effect significantly (p value* $0.043118 < 0.05$*). Now when we consider the treatment effects, we should condition on being given covariate. So the code will be different(just like the full model* 1 *and full model* 2 *as in the balloon example of lecture* 21*.)*

So we have:

```
fit <- lm(towel$Absorbancy ~ towel$Rate + factor(towel$Treatment))
anova(fit)

## Analysis of Variance Table
##
## Response: towel$Absorbancy
##                          Df   Sum Sq  Mean Sq F value   Pr(>F)
## towel$Rate                1 0.114637 0.114637 20.5106 0.006232 **
## factor(towel$Treatment)  5 0.242057 0.048411  8.6617 0.016689 *
## Residuals                 5 0.027946 0.005589
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

the p value for testing $H_0 : \tau_1 = \ldots = \tau_6$ *is* $0.016689 < 0.05$*, so we reject* H_0 *and consider at least two treatment groups have different effects, given the existence of covariates.*

For part (f)*:*

Our model here is:

$$Y_{ijt} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \theta(x_{ijt} - \bar{x}_{...}) + \epsilon_{ijt}$$

We run the two way ANCOVA model as follows:

```
  fit <- lm(Absorbancy~Rate + factor(A)*factor(B), data = towel)
  anova(fit)

## Analysis of Variance Table
##
## Response: Absorbancy
##                     Df   Sum Sq  Mean Sq F value   Pr(>F)
## Rate                 1 0.114637 0.114637 20.5106 0.006232 **
## factor(A)            2 0.142468 0.071234 12.7451 0.010890 *
## factor(B)            1 0.006374 0.006374  1.1404 0.334393
## factor(A):factor(B)  2 0.093215 0.046607  8.3389 0.025550 *
```

```
## Residuals            5 0.027946 0.005589
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

For the interaction part we got p value $0.0255 < 0.05$ *so we fail to reject there is no interaction. Since there is interaction, tests for main effects would be not much of interest. However if we do want to get a formal result, from the ANCOVA table, both p values for the main effects* < 0.05 *so we consider both brand and printing do have separate treatment effects on the absorbancy.*