HW4

Zeqiu.Yu

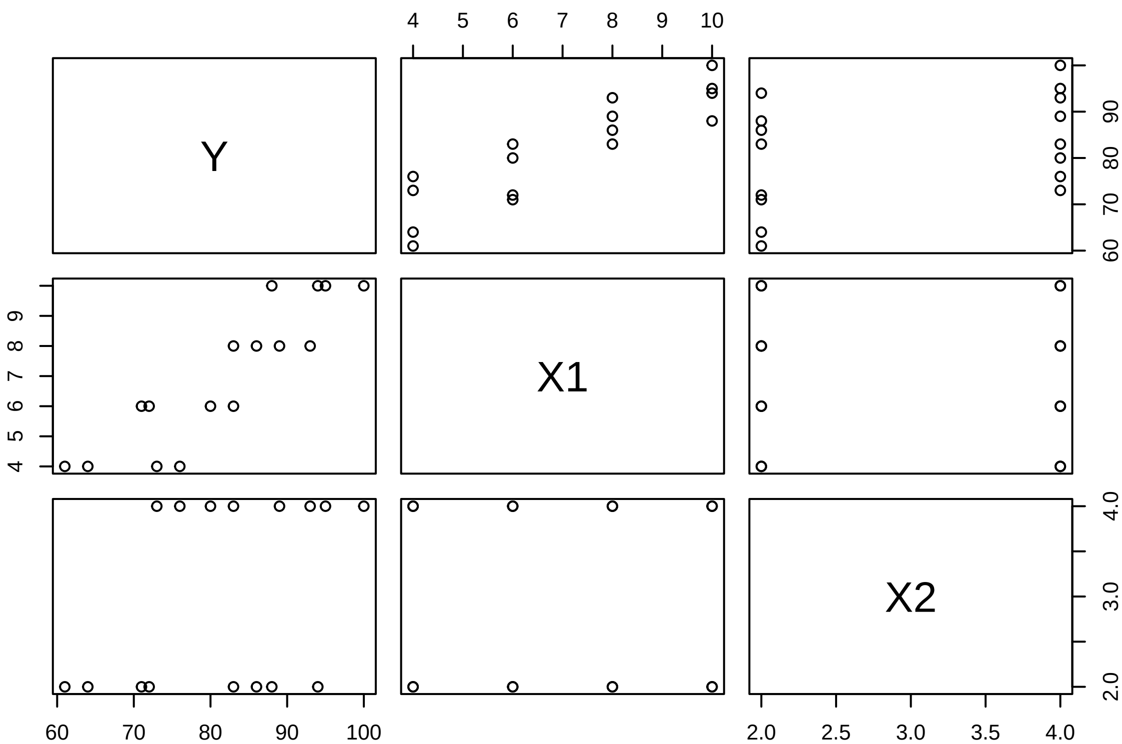
2022-10-26

Homework 4

Question 6.5

(a)

```
input1 <- read.table("./BookDataSets/Chapter  6 Data Sets/CH06PR05.txt")
names(input1) <- c("Y","X1", "X2")
pairs(input1[,c(1,2,3)])
```




```
cor(input1[,c(1,2,3)])
```

```
##             Y        X1        X2
## Y  1.0000000 0.8923929 0.3945807
## X1 0.8923929 1.0000000 0.0000000
## X2 0.3945807 0.0000000 1.0000000
```

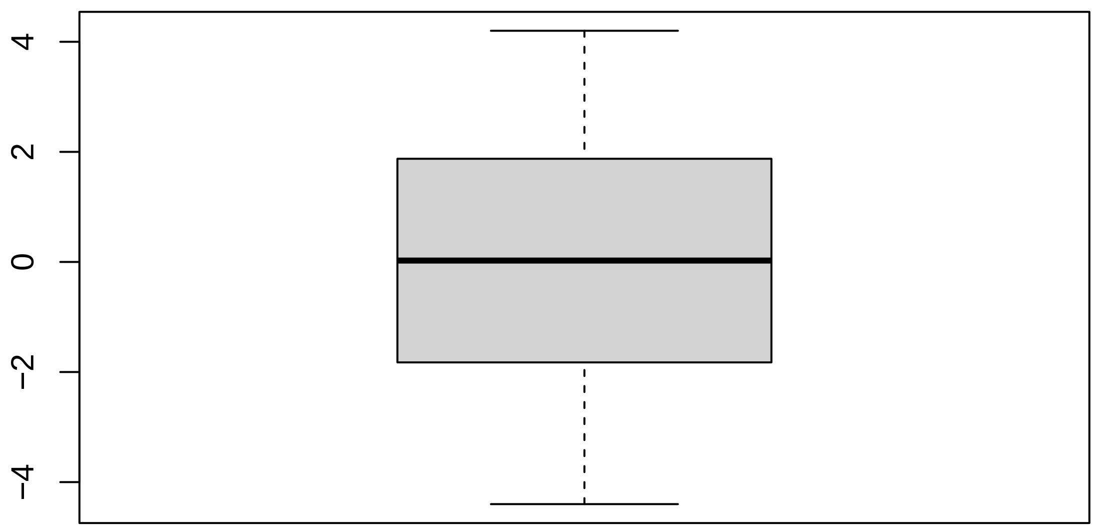
The scatter plot matrix shows Y and X_1 has a relative strong relationship. Y and X_2 are less correlated.
(b)

```r
fit <- lm(Y~X1+X2, data = input1)
summary(fit)
```

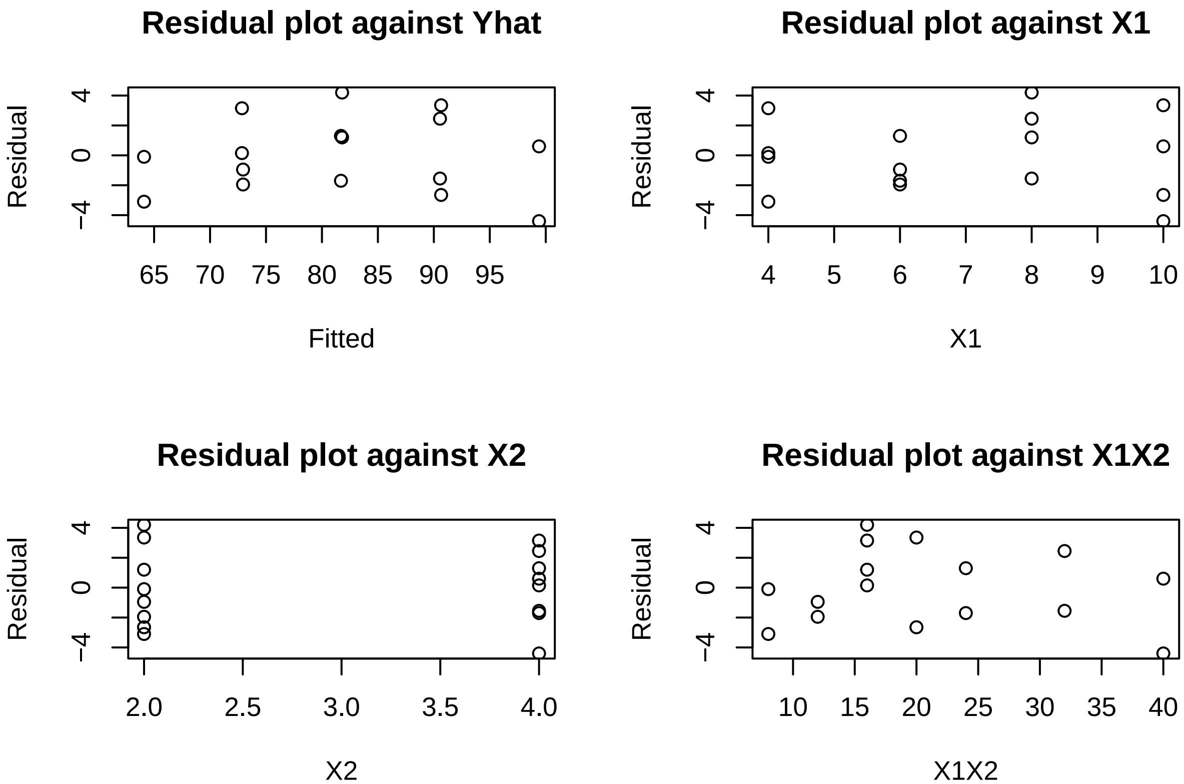
```
##
## Call:
## lm(formula = Y ~ X1 + X2, data = input1)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -4.400 -1.762  0.025  1.587  4.200
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  37.6500     2.9961  12.566 1.20e-08 ***
## X1            4.4250     0.3011  14.695 1.78e-09 ***
## X2            4.3750     0.6733   6.498 2.01e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.693 on 13 degrees of freedom
## Multiple R-squared:  0.9521, Adjusted R-squared:  0.9447
## F-statistic: 129.1 on 2 and 13 DF,  p-value: 2.658e-09
```

From the summary, we get $\hat{Y} = 37.6500 + 4.4250X_1 + 4.3750X_2$. b_1 can be interpreted as holding the sweetness X_2, the mean change in degree of brand liking Y is 4.4250 as moisture content X_1 increase per unit.
(c)&(d)

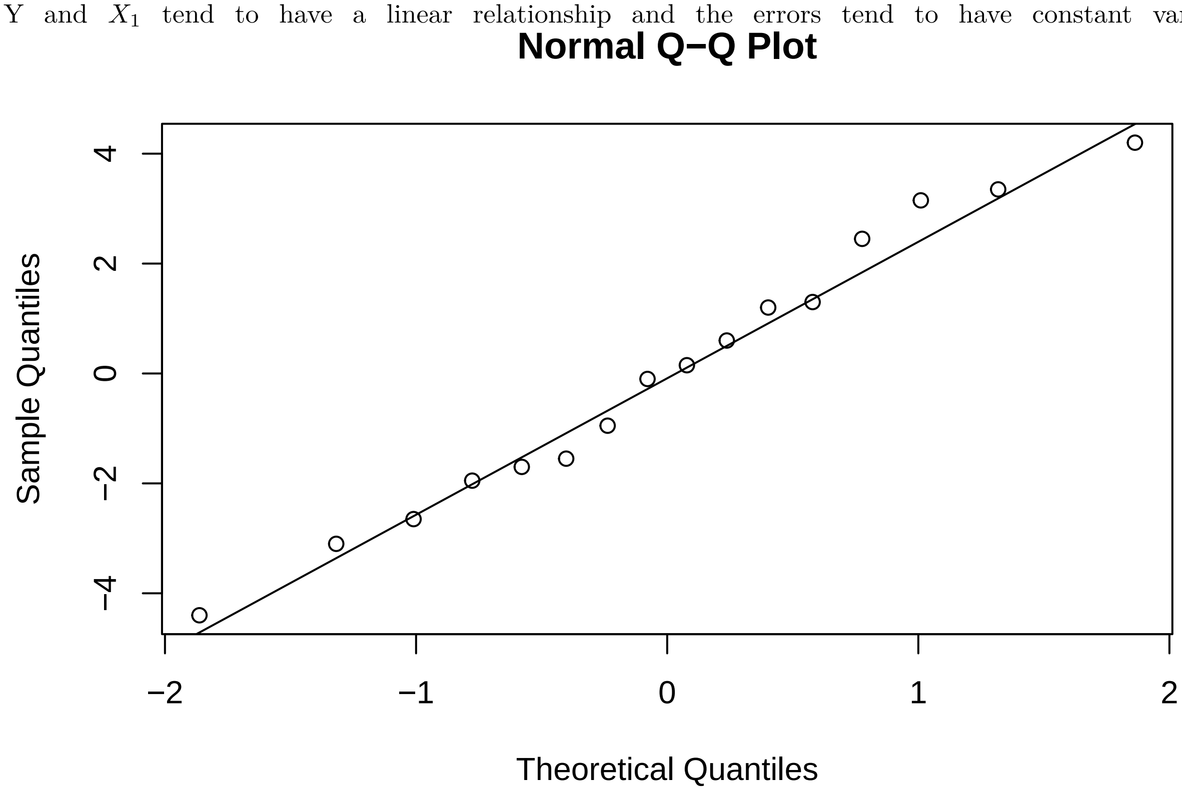
```r
Yhat <- fit$fitted.values
e <- fit$residuals
```



The distuion of the residuals is centered at 0, and it is symmetric.



Y and X_1 tend to have a linear relationship and the errors tend to have constant variance.



It suggests the error terms tend to be normally distributed.

(e)

```
## Loading required package: zoo
```

```
##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

Let H_0: $\gamma_1 = \gamma_2 = 0$, and $H_a : \gamma_1$ *and* γ_2 are not all 0.

```
bptest(fit,studentize = FALSE)
```

```
##
##  Breusch-Pagan test
##
## data:  fit
## BP = 1.0422, df = 2, p-value = 0.5939
```

p-value larger than α conclude error variance constant, otherwise error variance not constant.
Because p-value is larger than $\alpha = 0.01$, we conclude error variance constant.

(f) Let $H_0 : E[Y] = \beta_0 + \beta_1 X_1 + \beta_2 X_2$ and $H_a : E[Y] \neq \beta_0 + \beta_1 X_1 + \beta_2 X_2$. We have seen replications in the scatter plot, then we can conduct lack of fit test.

```
Full <- lm(Y~as.factor(X1)*as.factor(X2), data = input1)
Reduced <- lm(Y~X1+X2, data = input1)
anova(Full)
```

```
## Analysis of Variance Table
##
## Response: Y
##                             Df  Sum Sq Mean Sq F value    Pr(>F)
## as.factor(X1)                3 1581.50  527.17 73.9883 3.554e-06 ***
## as.factor(X2)                1  306.25  306.25 42.9825 0.0001773 ***
## as.factor(X1):as.factor(X2)  3   22.25    7.42  1.0409 0.4253674
## Residuals                    8   57.00    7.13
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(Reduced)
```

```
## Analysis of Variance Table
##
## Response: Y
##           Df  Sum Sq Mean Sq F value    Pr(>F)
## X1         1 1566.45 1566.45 215.947 1.778e-09 ***
## X2         1  306.25  306.25  42.219 2.011e-05 ***
## Residuals 13   94.30    7.25
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Then MSPE = 7.13, MSLF = (SSTO - SSE)/df = (94.3 - 57)/5 = 7.46. $F^* = \frac{7.46}{7.125} = 1.047 \leq F(0.99; 5, 8) = 6.63$. If $F^* < 6.63$, we will accept H_0, else reject. Hence, we accept H_0

6.6

(a) $H_0 : \beta_1 = \beta_2 = 0$, $H_a :$ *not both* β_1 *and* β_2 *equal* 0. If F^* is smaller than F(0.99; 2,13) = 6.70, we accept H_0, else reject H_0 and accept H_a. According to the summary above, $F^* = 129.1 > 6.70$. Hence, we accept H_a.

(b) 2.658e-09

(c) $\frac{\alpha}{2\times2} = 0.0025$. Then according to the summary table, s$\{b_1\} = 0.3011$, s$\{b_2\} = 0.6733$, and t(0.9975; 13) $= 3.372$. We have $4.4250 \pm 3.372 \times 0.3011$ for β_1, and $4.3750 \pm 3.372 \times 0.6733$ for β_2. Hence, with 99% confidence, β_1 will between 3.410 and 5.440, and β_2 between 2.106 and 6.644 simulatneously.

6.7

(a) $R^2 = \frac{SSR}{SSTO} = \frac{1872.7}{1967.0}$According to the summary table, $R^2 = 0.9521$
When X_1 and X_2 are considered, the variation in Y is reduced by 95.21%.

(b) 0.9521. Yes.

6.8

(a)

```
New1 <- data.frame(X1 = 5, X2 = 4)
predict(fit, New1, interval = "confidence", level = 0.99)
```

```
##      fit      lwr      upr
## 1 77.275 73.88111 80.66889
```

With 99% confidence, the mean predicted value will between 73.88111 and 80.66889 with repect to $X_{h1} = 5$ and $X_{h2} = 4$
(b)

```
New1 <- data.frame(X1 = 5, X2 = 4)
predict(fit, New1, interval = "prediction", level = 0.995)
```

```
##      fit     lwr     upr
## 1 77.275 67.4292 87.1208
```

According to Bonferroni, with 99% confidence, the new observation value will between 67.4292 and 87.1208 with repect to $X_{h1} = 5$ and $X_{h2} = 4$.