多元数据分析第三次作业

应数 2101　杨嘉昱　2216113458

2024 年 5 月 7 日

第二章作业

2.2

Solution.

1. 模型为

$$Y = X\beta + \varepsilon$$

最小二乘估计为

$$\hat{\beta} = \left(X^{\mathrm{T}}X\right)^{-1} X^{\mathrm{T}}Y = \frac{X^{\mathrm{T}}Y}{X^{\mathrm{T}}X} = \frac{\sum_{j=1}^{n} x_j y_j}{\sum_{j=1}^{n} x_j^2}.$$

由于

$$\hat{\beta} = \frac{X^{\mathrm{T}}Y}{X^{\mathrm{T}}X} = \frac{X^{\mathrm{T}}(X\beta + \varepsilon)}{X^{\mathrm{T}}X} = \beta + \frac{X^{\mathrm{T}}\varepsilon}{X^{\mathrm{T}}X}.$$

从而

$$\mathbb{E}\hat{\beta} = \beta + \mathbb{E}\left(\frac{X^{\mathrm{T}}\varepsilon}{X^{\mathrm{T}}X}\right) = \beta$$

即 $\hat{\beta}$ 为 β 的无偏估计。

2. 令

$$\hat{\sigma}^2 = \frac{1}{n-1} Y^{\mathrm{T}}(I-H)Y, \qquad H = \frac{XX^{\mathrm{T}}}{X^{\mathrm{T}}X}.$$

则

$$\begin{aligned}(n-1)\mathbb{E}\hat{\sigma}^2 &= \mathbb{E}\mathrm{tr}\left(Y^{\mathrm{T}}(I-H)Y\right) = \mathbb{E}\mathrm{tr}\left(\varepsilon^{\mathrm{T}}(I-H)\varepsilon\right) \\ &= \mathbb{E}\mathrm{tr}\left((I-H)\varepsilon\varepsilon^{\mathrm{T}}\right) = \sigma^2 \mathrm{tr}\left(I - \frac{XX^{\mathrm{T}}}{X^{\mathrm{T}}X}\right) = (n-1)\sigma^2\end{aligned}$$

从而

$$\hat{\sigma}^2 = \frac{1}{n-1}\frac{X^{\mathrm{T}}XY^{\mathrm{T}}Y - \left(X^{\mathrm{T}}Y\right)^2}{X^{\mathrm{T}}X} = \frac{1}{n-1} \cdot \frac{\left(\sum_{j=1}^{n} x_j^2\right)\left(\sum_{j=1}^{n} y_j^2\right) - \left(\sum_{j=1}^{n} x_j y_j\right)^2}{\sum_{j=1}^{n} x_j^2}$$

是 σ^2 的无偏估计。

3. 全模型为

$$Y = X\beta + \varepsilon.$$

$$\mathrm{SSE}(R) = Y^{\mathrm{T}}(I-H)Y = \frac{X^{\mathrm{T}}XY^{\mathrm{T}}Y - \left(X^{\mathrm{T}}Y\right)^2}{X^{\mathrm{T}}X}, \qquad f_R = n-1.$$

约简模型为

$$Y = \varepsilon$$

$$\mathrm{SSE}(F) = \sum_{j=1}^{n}\left(y_j - 0\right)^2 = Y^{\mathrm{T}}Y, \qquad f_F = n.$$

从而统计量以及零分布为

$$F=\frac{(\mathrm{SSE}(R)-\mathrm{SSE}(F))/1}{\mathrm{SSE}(F)/n}=n\cdot\frac{\mathrm{SSE}(R)-\mathrm{SSE}(F)}{\mathrm{SSE}(F)}\sim F(1,n).$$

4. 由于

$$\hat\beta\sim\mathcal{N}\left(\beta,\sigma^2\left(X^{\mathrm{T}}X\right)^{-1}\right),\qquad \frac{n-1}{\sigma^2}\hat\sigma^2\sim\chi(n-1).$$

依假设 $\beta=0$。从而 t 统计量集及零分布为：

$$\frac{\hat\beta\sqrt{X^{\mathrm{T}}X}}{\hat\sigma}\sim t(n-1).$$

5. 由于

$$\hat y_0=x_0\hat\beta\sim\mathcal{N}\left(x_0^{\mathrm{T}}\beta,\sigma^2x_0^{\mathrm{T}}\left(X^{\mathrm{T}}X\right)^{-1}x_0\right)\qquad y_0=x_0^{\mathrm{T}}\beta+\varepsilon\sim\mathcal{N}\left(x_0^{\mathrm{T}}\beta,\sigma^2\right)$$

故

$$\hat y_0-y_0\sim\mathcal{N}\left(0,\sigma^2\left(1+x_0^{\mathrm{T}}\left(X^{\mathrm{T}}X\right)^{-1}x_0\right)\right)$$

从而

$$\frac{\hat y_0-y_0}{\sqrt{\mathrm{MSE}\left(1+x_0^{\mathrm{T}}\left(X^{\mathrm{T}}X\right)^{-1}x_0\right)}}\sim t(n-1).$$

从而置信度为 $1-\alpha$ 的置信区间为

$$\begin{aligned}&\hat y_0\pm t_{1-\alpha/2}(n-p)\sqrt{\mathrm{MSE}\left(1+x_0^{\mathrm{T}}\left(X^{\mathrm{T}}X\right)^{-1}x_0\right)}\\=&x_0\frac{X^{\mathrm{T}}Y}{X^{\mathrm{T}}X}\pm t_{1-\alpha/2}(n-p)\sqrt{\frac{1}{n-1}\frac{X^{\mathrm{T}}XY^{\mathrm{T}}Y-\left(X^{\mathrm{T}}Y\right)^2}{X^{\mathrm{T}}X}\cdot\frac{X^{\mathrm{T}}X+x_0^{\mathrm{T}}x_0}{X^{\mathrm{T}}X}}.\end{aligned}$$

□

2.3

Solution. 记

$$X_i=(x_{1i},\cdots,x_{ni})^{\mathrm{T}},(i=1,2),\qquad X_3=(\sqrt{x_{13}},\cdots,\sqrt{x_{n3}})^{\mathrm{T}}$$

那么全模型为

$$Y=\beta_0+\beta_1X_1+\beta_2X_2+\beta_3X_1X_2+\beta_4X_3+\varepsilon.$$

此时自由度以及误差平方和为

$$f_R=4,\qquad \mathrm{SSE}(R)$$

1. 约简模型为

$$Y=\beta_0+\beta_1X_1+\beta_2X_2+\varepsilon$$

自由度以及误差平方和为

$$f_F=2,\qquad \mathrm{SSE}(F)$$

统计量以及零分布为

$$\frac{(\mathrm{SSE}(R)-\mathrm{SSE}(F))/(f_R-f_F)}{\mathrm{SSE}(F)/f_F}=\frac{\mathrm{SSE}(R)-\mathrm{SSE}(F)}{\mathrm{SSE}(F)}\sim F(2,2).$$

2. 约简模型为

$$Y = \beta_0 + \beta_1(X_1 + X_3) + \beta_3 X_1 X_2 + \beta_4 X_3 + \varepsilon.$$

自由度以及误差平方和为

$$f_F = 3, \qquad \mathrm{SSE}(F)$$

统计量以及零分布为

$$\frac{(\mathrm{SSE}(R) - \mathrm{SSE}(F))/(f_R - f_F)}{\mathrm{SSE}(F)/f_F} = 3 \cdot \frac{\mathrm{SSE}(R) - \mathrm{SSE}(F)}{\mathrm{SSE}(F)} \sim F(1,3).$$

3. 约简模型为

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + X_3 + \varepsilon.$$

自由度以及误差平方和为

$$f_F = 3, \qquad \mathrm{SSE}(F)$$

统计量以及零分布为

$$\frac{(\mathrm{SSE}(R) - \mathrm{SSE}(F))/(f_R - f_F)}{\mathrm{SSE}(F)/f_F} = 3 \cdot \frac{\mathrm{SSE}(R) - \mathrm{SSE}(F)}{\mathrm{SSE}(F)} \sim F(1,3).$$

□

2.4

Solution. 记

$$Y = (y_1, \cdots, y_n)^{\mathrm{T}}, \qquad \varepsilon = (\varepsilon_1, \cdots, \varepsilon_n)^{\mathrm{T}}, \qquad X_i = (x_{1i}, \cdots, x_{ni})^{\mathrm{T}}, i = 1, 2$$

其中 $n = 15$。又记

$$\beta = (\beta_0, \beta_1, \beta_2)^{\mathrm{T}}, \qquad X = (\mathbb{1}, X_1, X_2)$$

那么模型为

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon = X\beta + \varepsilon.$$

1. β 的最小二乘估计为

$$\hat{\beta} = \left(X^{\mathrm{T}} X\right)^{-1} X^{\mathrm{T}} Y$$

均方误差的估计为

$$\hat{\sigma}^2 = \frac{1}{14} Y^{\mathrm{T}} (I - H) Y, \qquad H = X \left(X^{\mathrm{T}} X\right)^{-1} X^{\mathrm{T}}.$$

带入数值求解得:

$$\hat{\beta} = (3.45, 4.96, 0.01)^{\mathrm{T}}, \qquad \hat{\sigma}^2 = 4.74.$$

2. 方差分析表为 由于 p 值很小，拒绝假设 $H_0: \beta_1 = \beta_2 = 0$，即认为 Y 与 X_1, X_2 的线性回归关系显

方差来源	自由度	平方和	均方	F 值	p 值
回归 (R)	2	53844.72	26922.36	5679.466	0
误差 (E)	12	56.88	4.74		
总和 (T)	14	53901.6			

著。另一方面，计算得

$$R = 0.9994722.$$

说明 Y 与 X_1, X_2 线性关系显著。

3. 置信区间分别为

$$[4.83, 5.09] \qquad [0.007, 0.01]$$

4. 全模型为

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \varepsilon$$

约简模型为

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

检验假设为

$$H_0 : \beta_3 = 0.$$

统计量观测值与 p 值为

$$F = 0.03156013, \qquad p = 0$$

从而拒绝原假设，认为有必要引入交叉项。

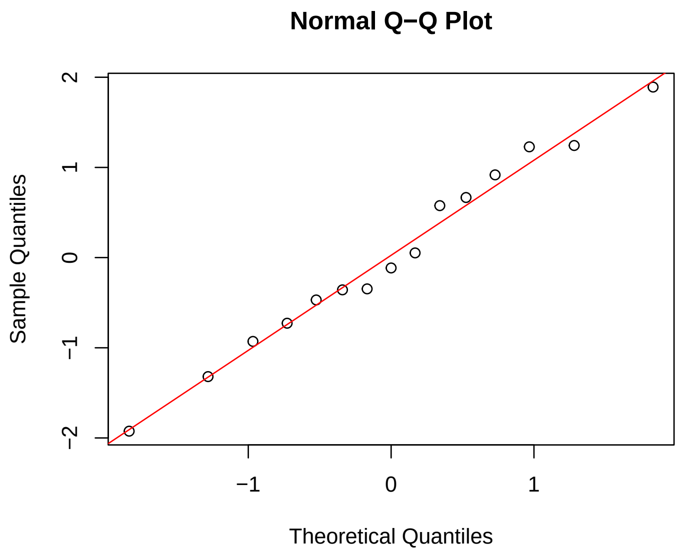
5. 预测值与置信区间分别为

$$1113.568, \qquad [1086.16, 1140.975].$$

6. 拟合值，残差与学生画残差分别为

$$\begin{aligned}
\hat{y} = &(161.90, 122.67, 224.43, 131.24, 67.70, 169.68, 79.73, 189.67,\\
&119.83, 53.29, 253.72, 228.69, 144.98, 100.53, 210.94)^{\mathrm{T}}\\
\hat{\varepsilon} = &(0.10, -2.66, -1.43, -0.24, -0.70, -0.68, 1.27,\\
&2.32, -3.83, 1.71, -1.72, 3.31, -0.98, 2.46, 1.06)^{\mathrm{T}}\\
r = &(0.05, -1.32, -0.73, -0.11, -0.36, -0.35, 0.67,\\
&1.23, -1.92, 0.92, -0.93, 1.89, -0.47, 1.24, 0.58)^{\mathrm{T}}
\end{aligned}$$

QQ 图如下



由于散点图近似线性，从而可以认为误差项是正态的。并且求的相关系数为

$$\hat{\rho} = 0.9933947.$$

本题所用代码如下:

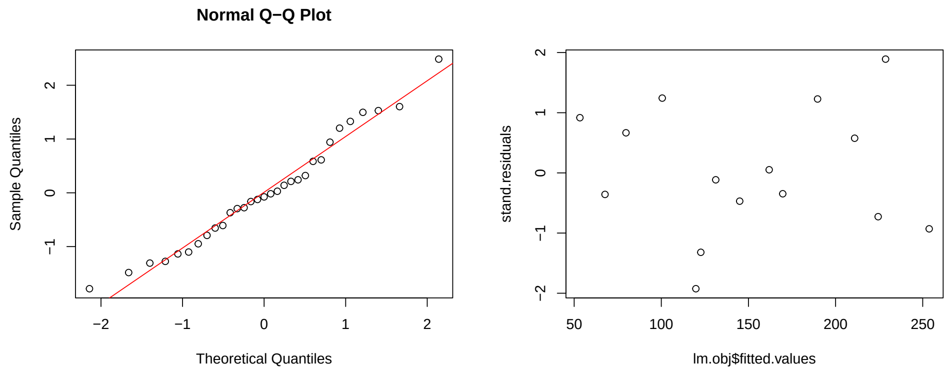
```r
D <- read.table('exercise2_4.txt')
colnames(D) <- c('y', 'X1', 'X2')
lm.obj <- lm(y~., data = D)
summary(lm.obj)

sse <- sum (( fitted (lm.obj) - D$y)^2)
ssr <- sum (( fitted (lm.obj) - mean (D$y))^2)
sst = sse + ssr
msr = ssr / 2
mse = sse / 12
f_value = msr / mse
p_value <- 1 - pf(f_value, 2, 12)
R_value <- sqrt(ssr/sst)
ssr
sse
sst
msr
mse
f_value
p_value
R_value

A <- summary(lm.obj)$coefficients
alpha <- 0.05
df <- lm.obj$df.residual
A[,1]
left <- A[,1] - A[,2] * qt(1 - alpha / 2, df)
right <- A[,1] + A[,2] * qt(1 - alpha / 2, df)
left
right

new.data <- D
new.data[,4] <- D[,2] * D[,3]
fit <- lm(y~., data = new.data)
f_0_num <- (sum(lm.obj$residuals^2) - sum(fit$residuals^2)) / (lm.obj$df.residual - fit$df.
    residual)
f_0_den <- sum(fit$residuals^2) / fit$df.residual
f_0 <- f_0_num / f_0_den
p.value <- 1 - pf(f_0, lm.obj$df.residual - fit$df.residual, fit$df.residual)
f_0
p_value

predict(lm.obj, data.frame(X1 = 220, X2 = 2055), interval = 'prediction', type = 'response')

stand.residuals <- rstandard(lm.obj)
qqnorm(stand.residuals)
x <- seq(from = -2.5, to = 2.5, by = 0.1)
y <- sd(stand.residuals) * x + mean(stand.residuals)
abline(x, y, coef = c(mean(stand.residuals), sd(stand.residuals)), col = 'red')

correlation <- cor(qqnorm(stand.residuals)$x, qqnorm(stand.residuals)$y)
fitted_value <- lm.obj$fitted.values
residual_value <- lm.obj$residuals
fitted_value
residual_value
stand.residuals
correlation
```

□

2.6

Solution.

1. 正态 QQ 图和学生化残差 - $\hat{y}$ 图如下

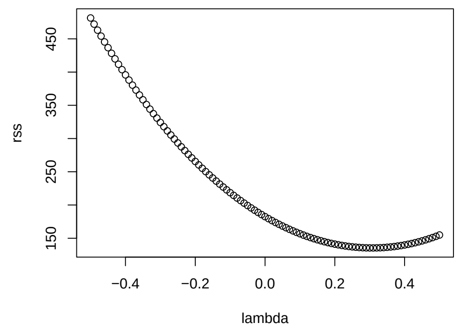


相关系数为

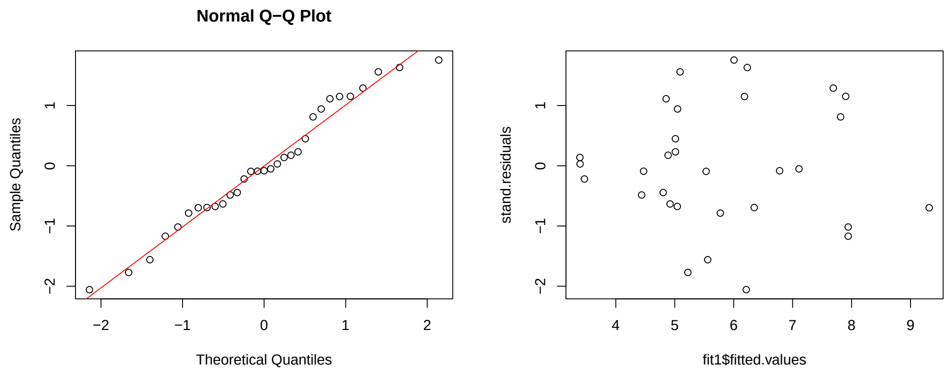
$$\hat{\rho} = 0.9883331.$$

由于散点图近似线形，因此可认为模型是合理的。

2. 计算得变换参数的值为 $\lambda = 0.66$



正态 QQ 图和学生化残差 - $\hat{y}$ 图如下



相关系数为

$$\hat{\rho} = 0.9889423.$$

说明拟合程度较线性更加成功。

本题所用代码如下：

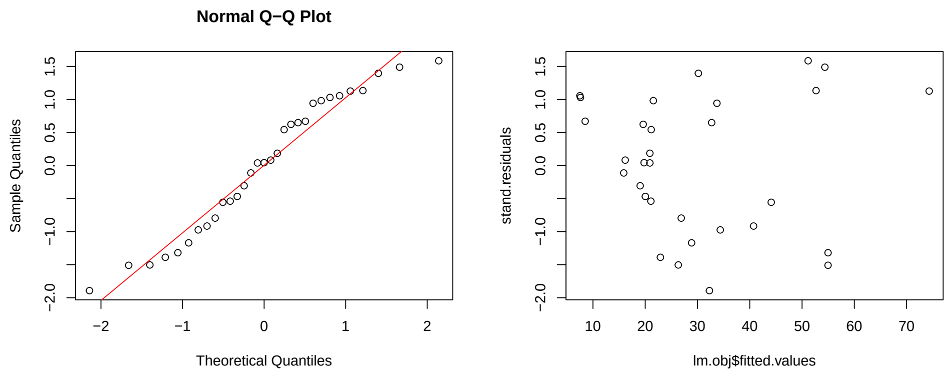
```
D <- read.table('exercise2_6.txt')
colnames(D) <- c('X1', 'X2', 'y')
lm.obj <- lm(y~., data = D)
summary(lm.obj)

stand.residuals <- rstandard(lm.obj)
qqnorm(stand.residuals)
correlation <- cor(qqnorm(stand.residuals)$x, qqnorm(stand.residuals)$y)
x <- seq(from = -2.5, to = 2.5, by = 0.1)
y <- sd(stand.residuals) * x + mean(stand.residuals)
abline(x, y, coef = c(mean(stand.residuals), sd(stand.residuals)), col = 'red')
correlation

plot(lm.obj$fitted.values, stand.residuals)

n <- dim(D)[1]
X <- D[,1:2]
y <- D[,3]
X <- as.matrix(cbind(matrix(1,n,1), X))
H <- X %*% solve(t(X) %*% X) %*% t(X)
lambda <- seq(from = -0.5, to = 0.5, by = 0.01)
K <- length(lambda)
rss <- rep(0, K)
for (k in 1:K) {
    if (lambda[k] != 0) {
    Z <- (1 / lambda[k]) * (y^(lambda[k]) - 1) / (prod(y)) ^ ((lambda[k] - 1) / n)
    } else {
    Z <- log(y) * (prod(y))^(1 / n)
    }
    rss[k] <- t(Z) %*% (diag(n) - H) %*% Z
}
opt.lamb <- lambda[which.min(rss)]
opt.lamb

new.y <- (y ^ opt.lamb - 1) / (opt.lamb)
new.data <- D
new.data[,3] <- new.y
fit1 <- lm(y~., data = new.data)
summary(fit1)

stand.residuals <- rstandard(fit1)
qqnorm(stand.residuals)
correlation <- cor(qqnorm(stand.residuals)$x, qqnorm(stand.residuals)$y)
x <- seq(from = -2.5, to = 2.5, by = 0.1)
y <- sd(stand.residuals) * x + mean(stand.residuals)
abline(x, y, coef = c(mean(stand.residuals), sd(stand.residuals)), col = 'red')
correlation

plot(fit1$fitted.values, stand.residuals)
```

□

2.7

Solution.

1. 正态 QQ 图和学生化残差 - $\hat{y}$ 图如下

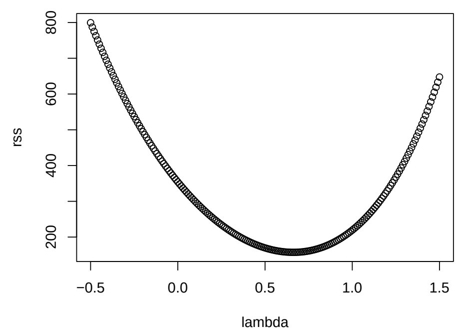


相关系数为

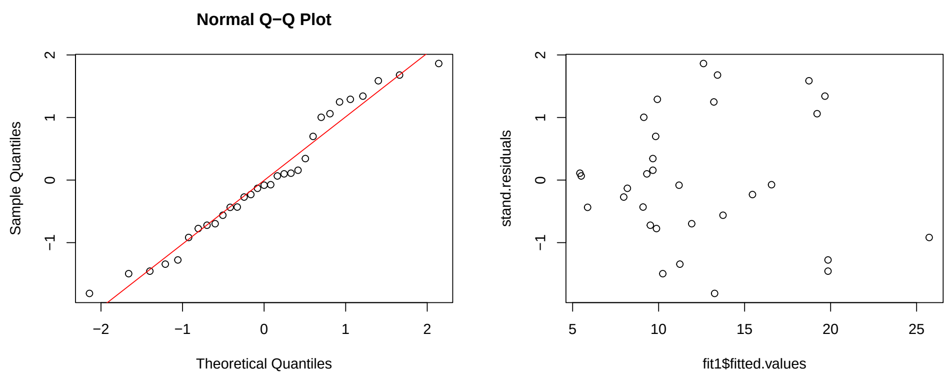
$$\hat{\rho} = 0.9791288.$$

由于散点图不是近似线形，因此可认为模型是不合理的。

2. 计算得变换参数的值为 $\lambda = 0.66$



正态 QQ 图和学生化残差 - $\hat{y}$ 图如下



QQ 图近似线性，相关系数为

$$\hat{\rho} = 0.9889423.$$

说明拟合程度较线性更加成功。

所用代码如下

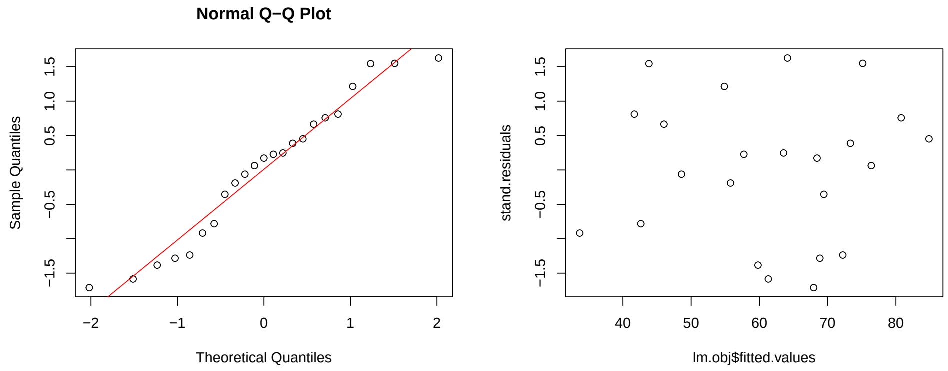
```
D_r <- read.table('exercise2_6.txt')
D <- D_r
D[,1] <- D_r[,1] * D_r[,1]
colnames(D) <- c('X1', 'X2', 'y')
lm.obj <- lm(y~., data = D)
summary(lm.obj)

stand.residuals <- rstandard(lm.obj)
qqnorm(stand.residuals)
correlation <- cor(qqnorm(stand.residuals)$x, qqnorm(stand.residuals)$y)
x <- seq(from = -2.5, to = 2.5, by = 0.1)
y <- sd(stand.residuals) * x + mean(stand.residuals)
abline(x, y, coef = c(mean(stand.residuals), sd(stand.residuals)), col = 'red')
correlation

plot(lm.obj$fitted.values, stand.residuals)

n <- dim(D)[1]
X <- D[,1:2]
y <- D[,3]
X <- as.matrix(cbind(matrix(1,n,1), X))
H <- X %*% solve(t(X) %*% X) %*% t(X)
lambda <- seq(from = -0.5, to = 1.5, by = 0.01)
K <- length(lambda)
rss <- rep(0, K)
for (k in 1:K) {
    if (lambda[k] != 0) {
    Z <- (1 / lambda[k]) * (y^(lambda[k]) - 1) / (prod(y)) ^ ((lambda[k] - 1) / n)
    } else {
    Z <- log(y) * (prod(y))^(1 / n)
    }
    rss[k] <- t(Z) %*% (diag(n) - H) %*% Z
}
plot(lambda,rss)
opt.lamb <- lambda[which.min(rss)]
opt.lamb

new.y <- (y ^ opt.lamb - 1) / (opt.lamb)
new.data <- D
new.data[,3] <- new.y
fit1 <- lm(y~., data = new.data)
summary(fit1)

stand.residuals <- rstandard(fit1)
qqnorm(stand.residuals)
correlation <- cor(qqnorm(stand.residuals)$x, qqnorm(stand.residuals)$y)
x <- seq(from = -2.5, to = 2.5, by = 0.1)
y <- sd(stand.residuals) * x + mean(stand.residuals)
abline(x, y, coef = c(mean(stand.residuals), sd(stand.residuals)), col = 'red')
correlation

plot(fit1$fitted.values, stand.residuals)
```

□

2.9

Solution.

1. 正态 QQ 图和学生化残差 - $\hat{y}$ 图如下



   相关系数为

$$\hat{\rho} = 0.9816644.$$

   由于散点图近似线形，因此可认为模型是合理的。

2. (a) $R_a^2(p)$ 准则：计算得

$$X_1 \leftrightarrow 0.5794702, \qquad X_1, X_2 \leftrightarrow 0.6305423 \qquad X_1, X_2, X_3 \leftrightarrow 0.6209731$$

   故选取 X_1, X_2 作为变量。

   (b) C_p 准则：计算得

$$X_1 \leftrightarrow 4.299472, \qquad X_1, X_2 \leftrightarrow 2.495063 \qquad X_1, X_2, X_3 \leftrightarrow 4.000000$$

   故选取 X_1, X_2 作为变量。

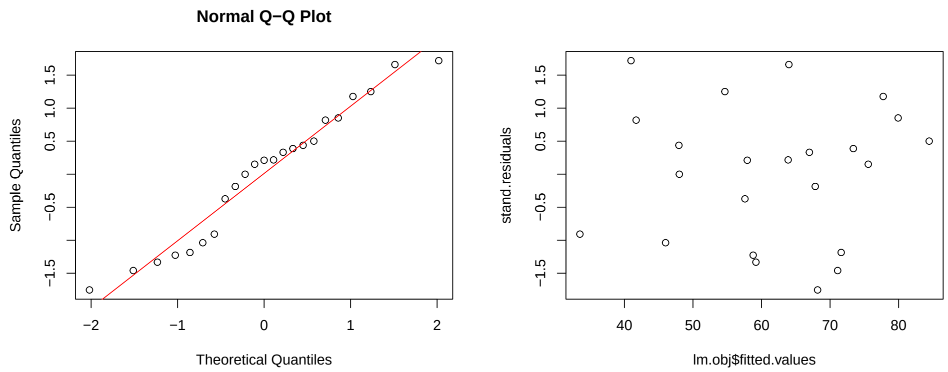
   (c) $PRESS_p$ 准则：计算得

$$X_1 \leftrightarrow 0.5985851, \qquad X_1, X_2 \leftrightarrow 0.6641294 \qquad X_1, X_2, X_3 \leftrightarrow 0.6726586$$

   故选取 X_1 作为变量。

3. 逐步回归法得：选取 X_1, X_2 作为变量。基本与 2 中结果一致。

4. 正态 QQ 图和学生化残差 - $\hat{y}$ 图如下



相关系数为

$$\hat{\rho} = 0.982029.$$

由于散点图近似线形，因此可认为模型是合理的。

所用代码如下

```
D <- read.table('exercise2_9.txt')
colnames(D) <- c('X1', 'X2', 'X3', 'y')
lm.obj <- lm(y~., data = D)
summary(lm.obj)

stand.residuals <- rstandard(lm.obj)
qqnorm(stand.residuals)
correlation <- cor(qqnorm(stand.residuals)$x, qqnorm(stand.residuals)$y)
x <- seq(from = -2.5, to = 2.5, by = 0.1)
y <- sd(stand.residuals) * x + mean(stand.residuals)
abline(x, y, coef = c(mean(stand.residuals), sd(stand.residuals)), col = 'red')
correlation

plot(lm.obj$fitted.values, stand.residuals)

library(leaps)
regfit.full = regsubsets(y~., data = D)
summary(regfit.full)
summary(regfit.full)$adjr2
summary(regfit.full)$cp
summary(regfit.full)$rsq

model <- lm(y ~ ., data = D)
step_model <- step(model, direction = "both")
summary(step_model)

new.data <- D[,1:2]
new.data[,3] <- D[,4]
colnames(new.data) <- c('X1', 'X2', 'y')
new.data
lm.obj <- lm(y~., data = new.data)
summary(lm.obj)

stand.residuals <- rstandard(lm.obj)
qqnorm(stand.residuals)
correlation <- cor(qqnorm(stand.residuals)$x, qqnorm(stand.residuals)$y)
x <- seq(from = -2.5, to = 2.5, by = 0.1)
y <- sd(stand.residuals) * x + mean(stand.residuals)
abline(x, y, coef = c(mean(stand.residuals), sd(stand.residuals)), col = 'red')
correlation

plot(lm.obj$fitted.values, stand.residuals)

```

□

2.10

Solution.

1. 带入数值求解得

$$\beta = (-20.00, 0.61, 0.18, 8.98)^{\mathrm{T}}$$

检验 p 值分别为

$$0.542, \qquad 0.526 \qquad 0.153 \qquad 0.510$$

从而

$$\mathbb{P}(Y=1)=\frac{\exp(-20.00+0.61X_1+0.18X_2+8.98X_3)}{1+\exp(-20.00+0.61X_1+0.18X_2+8.98X_3)}$$

并且只有 X_2 对是否破产有重要影响。

2. 检验假设为

$$H_0:\beta_3=0.$$

计算得 p 值为 $0.510>0.05$，可认为无显著影响。

建立仅含 X_1,X_2 的模型为

$$\mathbb{P}(Y=1)=\frac{\exp(-0.55037+0.15737X_1+0.19475X_2)}{1+\exp(-0.55037+0.15737X_1+0.19475X_2)}$$

p 值分别为

$$0.5628 \qquad 0.0357 \qquad 0.1117$$

可认为 X_1 对 $\mathbb{P}(Y=1)$ 有显著影响。预测值如下:

7.930776e-13 3.290103e-01 2.220446e-16 1.224829e-04 1.665939e-05 6.684758e-10

7.972692e-04 2.220446e-16 8.682391e-01 3.739876e-10 1.118967e-06 1.118674e-13

2.220446e-16 2.148712e-02 6.072085e-12 2.220446e-16 5.694457e-05 2.379878e-02

1.100124e-03 8.784556e-06 2.211979e-05 9.335252e-03 2.013984e-09 5.541148e-11

1.396926e-02 1.247900e-04 4.289639e-06 2.377275e-03 1.987758e-03 7.295078e-05

2.395350e-02 2.019661e-05 1.963000e-01 9.999181e-01 9.999528e-01 4.278330e-01

9.998776e-01 9.999041e-01 9.942752e-01 9.999028e-01 9.962502e-01 9.999976e-01

9.999927e-01 9.999982e-01 9.999931e-01 9.415830e-01 9.999936e-01 9.970490e-01

9.999937e-01 9.928044e-01 9.994180e-01 5.071745e-01 8.717798e-01 9.990596e-01

9.999562e-01 9.995826e-01 9.904502e-01 9.999816e-01 9.999931e-01 9.999643e-01

9.999907e-01 9.998978e-01 9.997753e-01 9.999619e-01 9.974965e-01 7.933926e-01

可见大部分预测比较成功。

3. 计算得概率分别为

$$\mathbb{P}(Y=1)=1, \qquad \mathbb{P}(Y=1)=0.9938452$$

可见概率相差不大

所用代码如下

```r
D <- read.table('exercise2_10.txt')
colnames(D) <- c('X1', 'X2', 'X3', 'y')
log.obj <- glm(y~., family = binomial(link = 'logit'), data = D)
model_summary <- summary(log.obj)
wald_values <- model_summary$coefficients[, "Pr(>|z|)"]
p_values <- model_summary$coefficients[, "Pr(>|z|)"]
wald_values
p_values

new.obj <- glm(y~ X1 + X2, family = binomial(link = 'logit'), data = D)
summary(new.obj)

```

```
predict.glm(new.obj, newdata = D[,c(1,2)], type = 'response')
new_data <- data.frame(X1 = 48.8, X2 = -10.5, X3 = 1.8)
predict(log.obj, newdata = new_data, type = "response")
predict(new.obj, newdata = new_data[,c(1,2)], type = "response")
```

□

第三章作业

3.1

Proof. 由于

$$\alpha_i = \mu_{i\cdot} - \mu = \frac{1}{b}\sum_{j=1}^{b}\mu_{ij} - \frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b}\mu_{ij} = \frac{1}{b}\sum_{j=1}^{b}\left(\mu_{ij} - \frac{1}{a}\sum_{i=1}^{a}\mu_{ij}\right)$$

故

$$\sum_{i=1}^{a}\alpha_i = \sum_{i=1}^{a}\frac{1}{b}\sum_{j=1}^{b}\left(\mu_{ij} - \frac{1}{a}\sum_{i=1}^{a}\mu_{ij}\right) = \frac{1}{b}\sum_{j=1}^{b}\left(\sum_{i=1}^{a}\mu_{ij} - a\cdot\frac{1}{a}\sum_{i=1}^{a}\mu_{ij}\right) = 0.$$

同理，

$$\sum_{j=1}^{n}\beta_j = 0.$$

另一方面，由于

$$\gamma_{ij} = (\mu_{ij} - \mu) - (\alpha_i + \beta_j), \qquad \beta_j = \mu_{\cdot j} - \mu, \qquad \mu_{\cdot j} = \frac{1}{a}\sum_{i=1}^{a}\mu_{ij},$$

故

$$\sum_{i=1}^{a}\gamma_{ij} = \sum_{i=1}^{a}\mu_{ij} - a\mu - \sum_{i=1}^{a}\alpha_i - a\beta_j = \sum_{i=1}^{a}\mu_{ij} - a\mu - a(\mu_{\cdot j} - \mu) = \sum_{i=1}^{a}\mu_{ij} - a\mu_{\cdot j} = 0.$$

同理

$$\sum_{j=1}^{b}\gamma_{ij} = 0$$

□

3.2

Proof. 由于

$$\begin{aligned} SS_A &= bc\sum_i(\alpha_i + \bar{\varepsilon}_{i\cdot\cdot} - \bar{\varepsilon})^2 = bc\sum_i\alpha_i^2 + 2bc\sum_i\alpha_i(\bar{\varepsilon}_{i\cdot\cdot} - \bar{\varepsilon}) + bc\sum_i(\bar{\varepsilon}_{i\cdot\cdot} - \bar{\varepsilon})^2 \\ &= bc\sum_i\alpha_i^2 + 2bc\sum_i\alpha_i(\bar{\varepsilon}_{i\cdot\cdot} - \bar{\varepsilon}) + bc\sum_i\bar{\varepsilon}_{i\cdot\cdot}^2 - 2bc\sum_i\bar{\varepsilon}_{i\cdot\cdot}\bar{\varepsilon} + bc\sum_i\bar{\varepsilon}^2 \\ &= bc\sum_i\alpha_i^2 + 2bc\sum_i\alpha_i(\bar{\varepsilon}_{i\cdot\cdot} - \bar{\varepsilon}) + bc\sum_i\bar{\varepsilon}_{i\cdot\cdot}^2 - abc\bar{\varepsilon}^2 \end{aligned}$$

其中用到了

$$\sum_i\bar{\varepsilon}_{i\cdot\cdot} = \frac{1}{bc}\sum_{i,j,k}\varepsilon_{ijk} = \frac{1}{a}\bar{\varepsilon}.$$

而

$$\bar{\varepsilon}_{i\cdot\cdot} \sim \mathcal{N}\left(0, \frac{\sigma^2}{bc}\right), \qquad \bar{\varepsilon} \sim \mathcal{N}\left(0, \frac{\sigma^2}{abc}\right)$$

故

$$\begin{aligned}\mathbb{E}SS_A &= bc\sum_i \alpha_i^2 + bc\sum_i \mathbb{E}\bar{\varepsilon}_{i\cdot\cdot}^2 - abc\mathbb{E}\bar{\varepsilon}^2 = bc\sum_i \alpha_i^2 + abc\mathbb{V}\mathrm{ar}\,\bar{\varepsilon}_{i\cdot\cdot} - abc\mathbb{V}\mathrm{ar}\,\bar{\varepsilon} \\ &= bc\sum_i \alpha_i^2 + (a-1)\sigma^2.\end{aligned}$$

同理有

$$\mathbb{E}SS_B = (b-1)\sigma^2 + ac\sum_{j=1}^{b}\beta_j^2.$$

$$\begin{aligned}SS_{AB} &= c\sum_{i,j}\left(\gamma_{ij} + \bar{\varepsilon}_{ij\cdot} - \bar{\varepsilon}_{i\cdot\cdot} - \bar{\varepsilon}_{\cdot j\cdot} + \bar{\varepsilon}\right)^2 \\ &= c\sum_{i,j}\gamma_{ij}^2 + c\sum_{i,j}\gamma_{ij}\left(\bar{\varepsilon}_{ij\cdot} - \bar{\varepsilon}_{i\cdot\cdot} - \bar{\varepsilon}_{\cdot j\cdot} + \bar{\varepsilon}\right) + c\sum_{i,j}\left(\bar{\varepsilon}_{ij\cdot} - \bar{\varepsilon}_{i\cdot\cdot} - \bar{\varepsilon}_{\cdot j\cdot} + \bar{\varepsilon}\right)^2\end{aligned}$$

而

$$\gamma_{ij} \sim \mathcal{N}(0,\sigma^2)\ \text{i.i.d} \qquad \Longrightarrow \qquad \mathbb{E}\left(c\sum_{i,j}\gamma_{ij}^2 + c\sum_{i,j}\gamma_{ij}\left(\bar{\varepsilon}_{ij\cdot} - \bar{\varepsilon}_{i\cdot\cdot} - \bar{\varepsilon}_{\cdot j\cdot} + \bar{\varepsilon}\right)\right) = c\sum_{i,j}\gamma_{ij}^2$$

故只需要计算 SS_{AB} 的最后一项即可。事实上

$$\sum_{i,j}\left(\bar{\varepsilon}_{ij\cdot} - \bar{\varepsilon}_{i\cdot\cdot} - \bar{\varepsilon}_{\cdot j\cdot} + \bar{\varepsilon}\right)^2 = \sum_{i,j}\left(\bar{\varepsilon}_{ij\cdot}^2 - \bar{\varepsilon}_{i\cdot\cdot}^2 - \bar{\varepsilon}_{\cdot j\cdot}^2 + \bar{\varepsilon}^2\right) + 2\sum_{i,j}\left(-\bar{\varepsilon}_{ij\cdot}\bar{\varepsilon}_{i\cdot\cdot} - \bar{\varepsilon}_{ij\cdot}\bar{\varepsilon}_{\cdot j\cdot} + \bar{\varepsilon}_{ij\cdot}\bar{\varepsilon} + \bar{\varepsilon}_{i\cdot\cdot}\bar{\varepsilon}_{\cdot j\cdot} - \bar{\varepsilon}_{i\cdot\cdot}\bar{\varepsilon} - \bar{\varepsilon}_{\cdot j\cdot}\bar{\varepsilon}\right)$$

而

$$\sum_j \bar{\varepsilon}_{ij\cdot} = \frac{1}{c}\sum_{j,k}\varepsilon_{ijk} = b\bar{\varepsilon}_{i\cdot\cdot}, \qquad \sum_i \bar{\varepsilon}_{ij\cdot} = \frac{1}{c}\sum_{i,k}\varepsilon_{ijk} = a\bar{\varepsilon}_{\cdot j\cdot} \qquad \sum_{i,j}\bar{\varepsilon}_{ij\cdot} = \frac{1}{c}\sum_{i,j,k}\varepsilon_{ijk} = ab\bar{\varepsilon}$$

$$\sum_i \bar{\varepsilon}_{i\cdot\cdot} = \frac{1}{bc}\sum_{i,j,k}\varepsilon_{ijk} = a\bar{\varepsilon} \qquad \sum_j \bar{\varepsilon}_{\cdot j\cdot} = \frac{1}{ac}\sum_{i,j,j}\varepsilon_{ijk} = b\bar{\varepsilon}$$

故有

$$\begin{aligned}&\sum_{i,j}\left(-\bar{\varepsilon}_{ij\cdot}\bar{\varepsilon}_{i\cdot\cdot} - \bar{\varepsilon}_{ij\cdot}\bar{\varepsilon}_{\cdot j\cdot} + \bar{\varepsilon}_{ij\cdot}\bar{\varepsilon} + \bar{\varepsilon}_{i\cdot\cdot}\bar{\varepsilon}_{\cdot j\cdot} - \bar{\varepsilon}_{i\cdot\cdot}\bar{\varepsilon} - \bar{\varepsilon}_{\cdot j\cdot}\bar{\varepsilon}\right) \\ &= -\sum_i \bar{\varepsilon}_{i\cdot\cdot} - \sum_j \bar{\varepsilon}_{\cdot j\cdot} + ab\bar{\varepsilon}^2 + ab\bar{\varepsilon}^2 - ab\bar{\varepsilon}^2 - ab\bar{\varepsilon}^2 = -\sum_i \bar{\varepsilon}_{i\cdot\cdot} - \sum_j \bar{\varepsilon}_{\cdot j\cdot}\end{aligned}$$

故

$$\begin{aligned}\mathbb{E}SS_{AB} &= c\sum_{i,j}\gamma_{ij}^2 + c\sum_{i,j}\left(\mathbb{E}\bar{\varepsilon}_{ij\cdot}^2 - \mathbb{E}\bar{\varepsilon}_{i\cdot\cdot}^2 - \mathbb{E}\bar{\varepsilon}_{\cdot j\cdot}^2 + \mathbb{E}\bar{\varepsilon}^2\right) - 2c\left(\sum_i \mathbb{E}\bar{\varepsilon}_{i\cdot\cdot} + \sum_j \mathbb{E}\bar{\varepsilon}_{\cdot j\cdot}\right) \\ &= c\sum_{i,j}\gamma_{ij}^2 + abc\sigma^2\left(\frac{1}{c} - \frac{1}{bc} - \frac{1}{ac} + \frac{1}{abc}\right) + 0 \\ &= c\sum_{i,j}\gamma_{ij}^2 + \sigma^2(a-1)(b-1).\end{aligned}$$

最后

$$\begin{aligned}SS_E &= \sum_{i,j,k}\left(\varepsilon_{ijk} - \bar{\varepsilon}_{ij\cdot}\right)^2 = \sum_{ijk}\varepsilon_{ijk}^2 - 2\sum_{i,j,k}\varepsilon_{ijk}\bar{\varepsilon}_{ij\cdot} + \sum_{i,j,k}\bar{\varepsilon}_{ij\cdot}^2 = \sum_{ijk}\varepsilon_{ijk}^2 + \sum_{i,j,k}\bar{\varepsilon}_{ij\cdot}^2 - 2\sum_{i,j}\bar{\varepsilon}_{ij\cdot}\sum_k \varepsilon_{ijk} \\ &= \sum_{ijk}\varepsilon_{ijk}^2 + \sum_{i,j,k}\bar{\varepsilon}_{ij\cdot}^2 - 2c\sum_{i,j}\bar{\varepsilon}_{ij\cdot}^2.\end{aligned}$$

从而

$$\mathbb{E}SS_E = abc\frac{\sigma^2}{1} + abc\frac{\sigma^2}{c} - 2abc\cdot\frac{\sigma^2}{c} = ab(c-1)\sigma^2.$$

$\square$

3.5

Solution.

1. 带入数据得检验的 p 值为 4.33×10^{-5} 远远小于 $\alpha=0.05$。从而拒绝原假设认为有显著影响。
2. 置信区间分别为

$$[6.25239, 7.503166], \qquad [7.652239, 8.614427] \qquad [8.289951, 10.11005]$$

同时置信区间分别为

$$[-3.793829, 1.282718] \qquad [-4.431541, -0.2129032], \qquad [-3.031692, 0.8983584]$$

可以看出投入经费越高，生产能力改善月显著。

所用代码如下

```
group1 <- c(7.6, 8.2, 6.8, 5.8, 6.9, 6.6, 6.3, 7.7, 6.0)
group2 <- c(6.7, 8.1, 9.4, 8.6, 7.8, 7.7, 8.9, 7.9, 8.3, 8.7, 7.1, 8.4)
group3 <- c(8.5, 9.7, 10.1, 7.8, 9.6, 9.5)

# 进行方差分析
model <- aov(c(group1, group2, group3) ~ rep(c("L", "M", "H"), c(length(group1), length(group2
    ), length(group3))))

# 查看方差分析结果
summary(model)

# 设定置信水平为95%
alpha <- 0.05

# 计算每个组均值的置信区间
mean_group1 <- mean(group1)
mean_group2 <- mean(group2)
mean_group3 <- mean(group3)

se_group1 <- sd(group1) / sqrt(length(group1))
se_group2 <- sd(group2) / sqrt(length(group2))
se_group3 <- sd(group3) / sqrt(length(group3))

ci_group1 <- c(mean_group1 - qt(1 - alpha / 2, df = length(group1) - 1) * se_group1,
               mean_group1 + qt(1 - alpha / 2, df = length(group1) - 1) * se_group1)
ci_group2 <- c(mean_group2 - qt(1 - alpha / 2, df = length(group2) - 1) * se_group2,
               mean_group2 + qt(1 - alpha / 2, df = length(group2) - 1) * se_group2)
ci_group3 <- c(mean_group3 - qt(1 - alpha / 2, df = length(group3) - 1) * se_group3,
               mean_group3 + qt(1 - alpha / 2, df = length(group3) - 1) * se_group3)

# 计算差值的Bonferroni校正置信区间
diff_ci_group1_group2 <- ci_group1 - ci_group2
diff_ci_group1_group3 <- ci_group1 - ci_group3
diff_ci_group2_group3 <- ci_group2 - ci_group3

# 进行Bonferroni校正
alpha_bonferroni <- alpha / 3  # 对每个置信区间进行调整

ci_diff_group1_group2_bonferroni <- diff_ci_group1_group2 + c(-1, 1) * qt(1 - alpha_bonferroni
    / 2, df = Inf)
ci_diff_group1_group3_bonferroni <- diff_ci_group1_group3 + c(-1, 1) * qt(1 - alpha_bonferroni
    / 2, df = Inf)
```

```
ci_diff_group2_group3_bonferroni <- diff_ci_group2_group3 + c(-1, 1) * qt(1 - alpha_bonferroni
    / 2, df = Inf)

# 输出结果
cat("Group 1 mean CI (95%):", ci_group1, "\n")
cat("Group 2 mean CI (95%):", ci_group2, "\n")
cat("Group 3 mean CI (95%):", ci_group3, "\n")
cat("Group 1 - Group 2 difference CI (Bonferroni adjusted, 95%):",
    ci_diff_group1_group2_bonferroni, "\n")
cat("Group 1 - Group 3 difference CI (Bonferroni adjusted, 95%):",
    ci_diff_group1_group3_bonferroni, "\n")
cat("Group 2 - Group 3 difference CI (Bonferroni adjusted, 95%):",
    ci_diff_group2_group3_bonferroni, "\n")

```

□

3.6

Solution.

1. 样本均值和标准差如下:

	Fe^{2+}	Fe^{2+}
高剂量	3.698889	5.936667
中剂量	8.203889	9.632222
低剂量	11.750000	12.639444

	Fe^{2+}	Fe^{2+}
高剂量	2.030870	2.806778
中剂量	5.447386	6.691215
低剂量	7.028150	6.082089

表 1: 样本均值与标准差

标准差较为明显，可见假定误差的等发差性不合理

2. 变换后的样本均值和标准差如下:

	Fe^{2+}	Fe^{2+}
高剂量	1.160924	1.680129
中剂量	1.901225	2.090045
低剂量	2.279981	2.403389

	Fe^{2+}	Fe^{2+}
高剂量	0.5854773	0.4645464
中剂量	0.6585116	0.5736511
低剂量	0.6563113	0.5693701

表 2: 变换后的样本均值与标准差

此时标准差趋于一致。

3. 方差分析表如下

	Sum Sq	Mean Sq	F value	Pr(>F)
Fe^{2+}	15.59	7.794	22.524	7.91e-09 ***
Fe^{3+}	2.07	2.074	5.993	0.0161 *
$Fe^{2+} : Fe^{3+}$	0.81	0.405	1.171	0.3143

表 3: 方差分析表

由于 $0.3143 > \alpha$，可见交互效应不显著，而各自影响较为显著。

4. 关于剂量的置信区间与同时置信区间分别为：（顺序为：高 + 中，高 + 低，中 + 低）

$$[-0.7996574, -0.3505595] \qquad [0.4978426, 0.9469405] \qquad [1.0729510, 1.5220490]$$

$$[-0.8506702, -0.2995467] \qquad [0.4468298, 0.9979533] \qquad [1.0219382, 1.5730618]$$

关于铁离子种类的置信区间与同时置信区间分别为

$$[-0.5521593, -0.002128917], \qquad [-0.592573, 0.03828476]$$

可见剂量对存留量的影响差异比较显著，而铁离子种类对存留量的影响差异不大。

所用代码如下

```
Pre_data <- data.frame(X=c
    (0.71,1.66,2.01,2.16,2.42,2.42,2.56,2.60,3.31,3.64,3.74,3.74,4.39,4.50,5.07,5.26,8.15,
    8.24,2.20,2.93,3.08,3.49,4.11,4.95,5.16,5.54,5.68,6.25,7.25,7.90,8.85,11.96,15.54,15.89,
    18.30,18.59,2.25,3.93,5.08,5.82,5.84,6.89,8.50,8.56,9.44,10.52,13.46,13.57,14.76,16.41,
    16.96,17.56,22.82,29.13,2.20,2.69,3.54,3.75,3.83,4.08,4.27,4.53,5.32,6.18,6.22,6.33,6.97,
    6.97,7.52,8.36,11.65,12.45,4.04,4.16,4.42,4.93,5.49,5.77,5.86,6.28,6.97,7.06,7.78,
    9.23,9.34,9.91,13.46,18.40,23.89,26.39,2.71,5.43,6.38,6.38,8.32,9.04,9.56,10.01,10.08,
    10.62,13.80,15.99,17.90,18.25,19.32,19.87,21.60,22.25),
A = gl(3,18,108),
B = gl(2,54,108))

model <- aov(X ~ A + B + A:B, data = Pre_data)
summary(model)
means <- aggregate(X ~ A + B, data = Pre_data, FUN = mean)
std_devs <- aggregate(X ~ A + B, data = Pre_data, FUN = sd)

print(means)
print(std_devs)

Pre_data$X <- log(Pre_data$X)
means <- aggregate(X ~ A + B, data = Pre_data, FUN = mean)
std_devs <- aggregate(X ~ A + B, data = Pre_data, FUN = sd)

print(means)
print(std_devs)

model <- aov(X ~ A + B + A:B, data = Pre_data)
summary(model)

attach(Pre_data)
mu.A = c(mean(X[A == 1]), mean(X[A == 2]), mean(X[A = 3]))
mu.B = c(mean(X[B == 1]), mean(X[B == 2]))
a <- 3
b <- 2
c <- 18
alpha = 0.05
mse <- sum(model$residuals^2) / (a * b * (c - 1))

index.martix <- matrix(c(1,1,2,2,3,3), nrow = 3, ncol = 2)
mean.diff.cfi <- matrix(0,3,2)
mean.diff.cfi
adj.alpha <- alpha / 3
index.martix[1,]
# A 置信区间
t_quan <- qt(1 - 0.5 * alpha, a * b * (c - 1))
for (i in 1:3) {
```

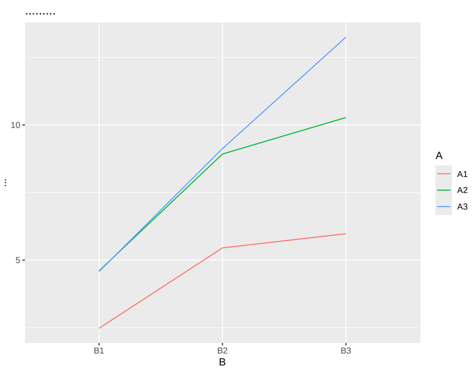
```r
ind <- index.martix[i,]
mean.diff <- mu.A[ind[1]] - mu.A[ind[2]]
fac <- t_quan * sqrt(2 * mse / a / c)
mean.diff.cfi[i,1] <- mean.diff - fac
mean.diff.cfi[i,2] <- mean.diff + fac
}
mean.diff.cfi
# A 同时置信区间
t_quan <- qt(1 - 0.5 * adj.alpha, a * b * (c - 1))
for (i in 1:3) {
ind <- index.martix[i,]
mean.diff <- mu.A[ind[1]] - mu.A[ind[2]]
fac <- t_quan * sqrt(2 * mse / a / c)
mean.diff.cfi[i,1] <- mean.diff - fac
mean.diff.cfi[i,2] <- mean.diff + fac
}
mean.diff.cfi

# B 置信区间
t_quan <- qt(1 - 0.5 * alpha, a * b * (c - 1))
mu.B[1] - mu.B[2] - t_quan * sqrt(2 * mse / b / c)
mu.B[1] - mu.B[2] + t_quan * sqrt(2 * mse / b / c)
# B 同时置信区间
t_quan <- qt(1 - 0.5 * alpha / 2, a * b * (c - 1))
mu.B[1] - mu.B[2] - t_quan * sqrt(2 * mse / b / c)
mu.B[1] - mu.B[2] + t_quan * sqrt(2 * mse / b / c)
```

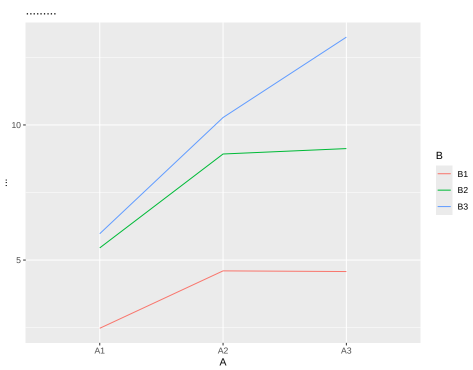
□

3.7

Solution.

1. 作出图像如下





2. 方差分析表如下

	Sum Sq	Mean Sq	F value	Pr(>F)
A	220.02	110.01	1827.9	<2e-16 ***
B	123.66	61.83	1027.3	<2e-16 ***
A:B	29.43	7.36	122.2	<2e-16 ***

可见交互效应显著。

3. B 各区间如下

```
          [,1]       [,2]
[1,] -3.2806901 -2.6693099
[2,] -3.8056901 -3.1943099
[3,] -0.8306901 -0.2193099
```

```
         [,1]     [,2]
[1,] -4.63069 -4.01931
[2,] -5.98069 -5.36931
[3,] -1.65569 -1.04431
```

```
         [,1]     [,2]
[1,] -4.85569 -4.24431
[2,] -8.98069 -8.36931
[3,] -4.43069 -3.81931
```

图 1: B 各水平置信区间

```
          [,1]       [,2]
[1,] -3.2806901 -2.6693099
[2,] -3.8056901 -3.1943099
[3,] -0.8306901 -0.2193099
```

```
         [,1]     [,2]
[1,] -4.63069 -4.01931
[2,] -5.98069 -5.36931
[3,] -1.65569 -1.04431
```

```
         [,1]     [,2]
[1,] -4.85569 -4.24431
[2,] -8.98069 -8.36931
[3,] -4.43069 -3.81931
```

图 2: B 各水平同时置信区间

A 各区间如下

```
          [,1]       [,2]
[1,] -2.4306901 -1.8193099
[2,] -2.4056901 -1.7943099
[3,] -0.2806901  0.3306901
```

```
          [,1]       [,2]
[1,] -3.7806901 -3.1693099
[2,] -3.9806901 -3.3693099
[3,] -0.5056901  0.1056901
```

```
         [,1]     [,2]
[1,] -4.60569 -3.99431
[2,] -7.58069 -6.96931
[3,] -3.28069 -2.66931
```

图 3: A 各水平置信区间

```
          [,1]       [,2]
[1,] -2.4306901 -1.8193099
[2,] -2.4056901 -1.7943099
[3,] -0.2806901  0.3306901
```

```
          [,1]       [,2]
[1,] -3.7806901 -3.1693099
[2,] -3.9806901 -3.3693099
[3,] -0.5056901  0.1056901
```

```
         [,1]     [,2]
[1,] -4.60569 -3.99431
[2,] -7.58069 -6.96931
[3,] -3.28069 -2.66931
```

图 4: A 各水平同时置信区间

所用代码如下

```r
data1 <- data.frame(X_data = c(2.4,2.7,2.3,2.5,  4.6,4.2,4.9,4.7, 4.8,4.5,4.4,4.6,
    5.8,5.2,5.5,5.3,  8.9,9.1,8.7,9.0, 9.1,9.3,8.7,9.4,
    6.1,5.7,5.9,6.2,  9.9,10.5,10.6,10.1, 13.5,13.0,13.3,13.2),
  A_con = gl(3,12,36),
  B_con = gl(3,4,36))

attach(data1)
mu = matrix(0,3,3)
for (i in 1:3) {
for (j in 1:3){
mu[i,j] = mean(X_data[A_con == i & B_con == j])
}
}
mu

library(ggplot2)

# 创建示例数据
data <- data.frame(
B = c("B1", "B2", "B3"),
A1 = c(2.475, 5.450, 5.975),
```

```
A2 = c(4.600, 8.925, 10.275),
A3 = c(4.575, 9.125, 13.250)
)

# 将数据从宽格式转换为长格式
data_long <- tidyr::gather(data, key = "A", value = "value", -B)

# 绘制折线图
ggplot(data_long, aes(x = B, y = value, color = A, group = A)) +
geom_line() +
labs(title = "折线图", x = "B", y = "值")


data <- data.frame(
A = c("A1", "A2", "A3"),
B1 = c(2.475, 4.600, 4.575),
B2 = c(5.450, 8.925, 9.125),
B3 = c(5.975, 10.275, 13.250)
)

# 将数据从宽格式转换为长格式
data_long <- tidyr::gather(data, key = "B", value = "value", -A)

# 绘制折线图
ggplot(data_long, aes(x = A, y = value, color = B, group = B)) +
geom_line() +
labs(title = "折线图", x = "A", y = "值")

model <- aov(X_data ~ A_con + B_con + A_con:B_con, data = data1)
summary(model)

# adj-------------
a <- 3
b <- 3
c <- 5
mse <- sum(model$residuals ^ 2) / a / b / (c - 1)
m <- 3
adj.alpha = 0.05 / m
t_tuan <- qt (1 - 0.5 * adj.alpha, a * b * (c - 1))
index.martix <- matrix(c(1, 1, 2, 2, 3, 3), nrow = 3, ncol = 2)
confi.int.B1 <- matrix(0, 3, 2)
mu.B <- mu[,1]

for (i in 1:3) {
ind <- index.martix[i,]
mean.diff <- mu.B[ind[1]] - mu.B[ind[2]]
fac <- t_quan * sqrt(2 * mse / c)
confi.int.B1[i,1] <- mean.diff - fac
confi.int.B1[i,2] <- mean.diff + fac

}
confi.int.B1

a <- 3
b <- 3
c <- 5
mse <- sum(model$residuals ^ 2) / a / b / (c - 1)
m <- 3
adj.alpha = 0.05 / m
```

```
t_tuan <- qt (1 - 0.5 * adj.alpha, a * b * (c - 1))
index.martix <- matrix(c(1, 1, 2, 2, 3, 3), nrow = 3, ncol = 2)
confi.int.B1 <- matrix(0, 3, 2)
mu.B <- mu[,2]

for (i in 1:3) {
ind <- index.martix[i,]
mean.diff <- mu.B[ind[1]] - mu.B[ind[2]]
fac <- t_quan * sqrt(2 * mse / c)
confi.int.B1[i,1] <- mean.diff - fac
confi.int.B1[i,2] <- mean.diff + fac

}
confi.int.B1

a <- 3
b <- 3
c <- 5
mse <- sum(model$residuals ^ 2) / a / b / (c - 1)
m <- 3
adj.alpha = 0.05 / m
t_tuan <- qt (1 - 0.5 * adj.alpha, a * b * (c - 1))
index.martix <- matrix(c(1, 1, 2, 2, 3, 3), nrow = 3, ncol = 2)
confi.int.B1 <- matrix(0, 3, 2)
mu.B <- mu[,3]

for (i in 1:3) {
ind <- index.martix[i,]
mean.diff <- mu.B[ind[1]] - mu.B[ind[2]]
fac <- t_quan * sqrt(2 * mse / c)
confi.int.B1[i,1] <- mean.diff - fac
confi.int.B1[i,2] <- mean.diff + fac

}
confi.int.B1


# inter-------------
a <- 3
b <- 3
c <- 5
mse <- sum(model$residuals ^ 2) / a / b / (c - 1)

m <- 3
adj.alpha = 0.05
t_tuan <- qt (1 - 0.5 * adj.alpha, a * b * (c - 1))
index.martix <- matrix(c(1, 1, 2, 2, 3, 3), nrow = 3, ncol = 2)
confi.int.B1 <- matrix(0, 3, 2)
mu.B <- mu[,1]

for (i in 1:3) {
ind <- index.martix[i,]
mean.diff <- mu.B[ind[1]] - mu.B[ind[2]]
fac <- t_quan * sqrt(2 * mse / c)
confi.int.B1[i,1] <- mean.diff - fac
confi.int.B1[i,2] <- mean.diff + fac

}
confi.int.B1
```

```
a <- 3
b <- 3
c <- 5
mse <- sum(model$residuals ^ 2) / a / b / (c - 1)
m <- 3
adj.alpha = 0.05
t_tuan <- qt (1 - 0.5 * adj.alpha, a * b * (c - 1))
index.martix <- matrix(c(1, 1, 2, 2, 3, 3), nrow = 3, ncol = 2)
confi.int.B1 <- matrix(0, 3, 2)
mu.B <- mu[,2]

for (i in 1:3) {
ind <- index.martix[i,]
mean.diff <- mu.B[ind[1]] - mu.B[ind[2]]
fac <- t_quan * sqrt(2 * mse / c)
confi.int.B1[i,1] <- mean.diff - fac
confi.int.B1[i,2] <- mean.diff + fac

}
confi.int.B1

a <- 3
b <- 3
c <- 5
mse <- sum(model$residuals ^ 2) / a / b / (c - 1)
m <- 3
adj.alpha = 0.05
t_tuan <- qt (1 - 0.5 * adj.alpha, a * b * (c - 1))
index.martix <- matrix(c(1, 1, 2, 2, 3, 3), nrow = 3, ncol = 2)
confi.int.B1 <- matrix(0, 3, 2)
mu.B <- mu[,3]

for (i in 1:3) {
ind <- index.martix[i,]
mean.diff <- mu.B[ind[1]] - mu.B[ind[2]]
fac <- t_quan * sqrt(2 * mse / c)
confi.int.B1[i,1] <- mean.diff - fac
confi.int.B1[i,2] <- mean.diff + fac

}
confi.int.B1


mu <- t(mu)
# adj-------------
a <- 3
b <- 3
c <- 5
mse <- sum(model$residuals ^ 2) / a / b / (c - 1)
m <- 3
adj.alpha = 0.05 / m
t_tuan <- qt (1 - 0.5 * adj.alpha, a * b * (c - 1))
index.martix <- matrix(c(1, 1, 2, 2, 3, 3), nrow = 3, ncol = 2)
confi.int.B1 <- matrix(0, 3, 2)
mu.B <- mu[,1]

for (i in 1:3) {
ind <- index.martix[i,]
```

```
mean.diff <- mu.B[ind[1]] - mu.B[ind[2]]
fac <- t_quan * sqrt(2 * mse / c)
confi.int.B1[i,1] <- mean.diff - fac
confi.int.B1[i,2] <- mean.diff + fac

}
confi.int.B1

a <- 3
b <- 3
c <- 5
mse <- sum(model$residuals ^ 2) / a / b / (c - 1)
m <- 3
adj.alpha = 0.05 / m
t_tuan <- qt (1 - 0.5 * adj.alpha, a * b * (c - 1))
index.martix <- matrix(c(1, 1, 2, 2, 3, 3), nrow = 3, ncol = 2)
confi.int.B1 <- matrix(0, 3, 2)
mu.B <- mu[,2]

for (i in 1:3) {
ind <- index.martix[i,]
mean.diff <- mu.B[ind[1]] - mu.B[ind[2]]
fac <- t_quan * sqrt(2 * mse / c)
confi.int.B1[i,1] <- mean.diff - fac
confi.int.B1[i,2] <- mean.diff + fac

}
confi.int.B1

a <- 3
b <- 3
c <- 5
mse <- sum(model$residuals ^ 2) / a / b / (c - 1)
m <- 3
adj.alpha = 0.05 / m
t_tuan <- qt (1 - 0.5 * adj.alpha, a * b * (c - 1))
index.martix <- matrix(c(1, 1, 2, 2, 3, 3), nrow = 3, ncol = 2)
confi.int.B1 <- matrix(0, 3, 2)
mu.B <- mu[,3]

for (i in 1:3) {
ind <- index.martix[i,]
mean.diff <- mu.B[ind[1]] - mu.B[ind[2]]
fac <- t_quan * sqrt(2 * mse / c)
confi.int.B1[i,1] <- mean.diff - fac
confi.int.B1[i,2] <- mean.diff + fac

}
confi.int.B1


# inter-------------
a <- 3
b <- 3
c <- 5
mse <- sum(model$residuals ^ 2) / a / b / (c - 1)
m <- 3
adj.alpha = 0.05
t_tuan <- qt (1 - 0.5 * adj.alpha, a * b * (c - 1))
```

```
index.martix <- matrix(c(1, 1, 2, 2, 3, 3), nrow = 3, ncol = 2)
confi.int.B1 <- matrix(0, 3, 2)
mu.B <- mu[,1]

for (i in 1:3) {
ind <- index.martix[i,]
mean.diff <- mu.B[ind[1]] - mu.B[ind[2]]
fac <- t_quan * sqrt(2 * mse / c)
confi.int.B1[i,1] <- mean.diff - fac
confi.int.B1[i,2] <- mean.diff + fac

}
confi.int.B1

a <- 3
b <- 3
c <- 5
mse <- sum(model$residuals ^ 2) / a / b / (c - 1)
m <- 3
adj.alpha = 0.05
t_tuan <- qt (1 - 0.5 * adj.alpha, a * b * (c - 1))
index.martix <- matrix(c(1, 1, 2, 2, 3, 3), nrow = 3, ncol = 2)
confi.int.B1 <- matrix(0, 3, 2)
mu.B <- mu[,2]

for (i in 1:3) {
ind <- index.martix[i,]
mean.diff <- mu.B[ind[1]] - mu.B[ind[2]]
fac <- t_quan * sqrt(2 * mse / c)
confi.int.B1[i,1] <- mean.diff - fac
confi.int.B1[i,2] <- mean.diff + fac

}
confi.int.B1

a <- 3
b <- 3
c <- 5
mse <- sum(model$residuals ^ 2) / a / b / (c - 1)
m <- 3
adj.alpha = 0.05
t_tuan <- qt (1 - 0.5 * adj.alpha, a * b * (c - 1))
index.martix <- matrix(c(1, 1, 2, 2, 3, 3), nrow = 3, ncol = 2)
confi.int.B1 <- matrix(0, 3, 2)
mu.B <- mu[,3]

for (i in 1:3) {
ind <- index.martix[i,]
mean.diff <- mu.B[ind[1]] - mu.B[ind[2]]
fac <- t_quan * sqrt(2 * mse / c)
confi.int.B1[i,1] <- mean.diff - fac
confi.int.B1[i,2] <- mean.diff + fac

}
confi.int.B1
```

□