556 Homework 8

Name: Xiaoke Qin

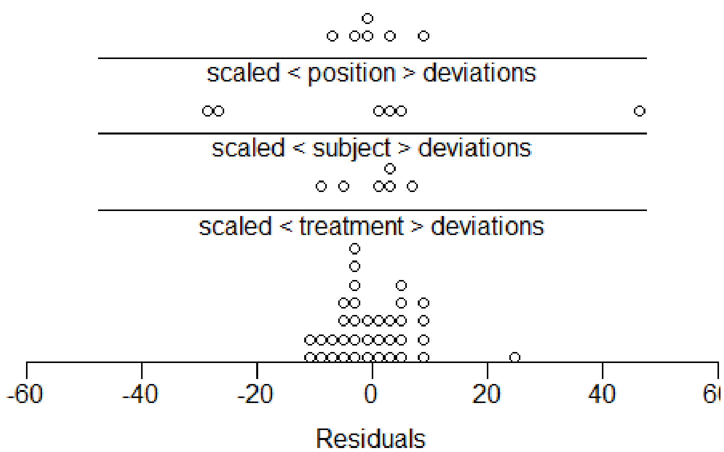
1. (a) Latin square. Each treatment appears once in every row and column to eliminate the effect of interaction of factors.

   (b) The randomization can be applied when we decide the symbols of positions, subjects and treatment.

   (c) Under contrast.treatment, we have

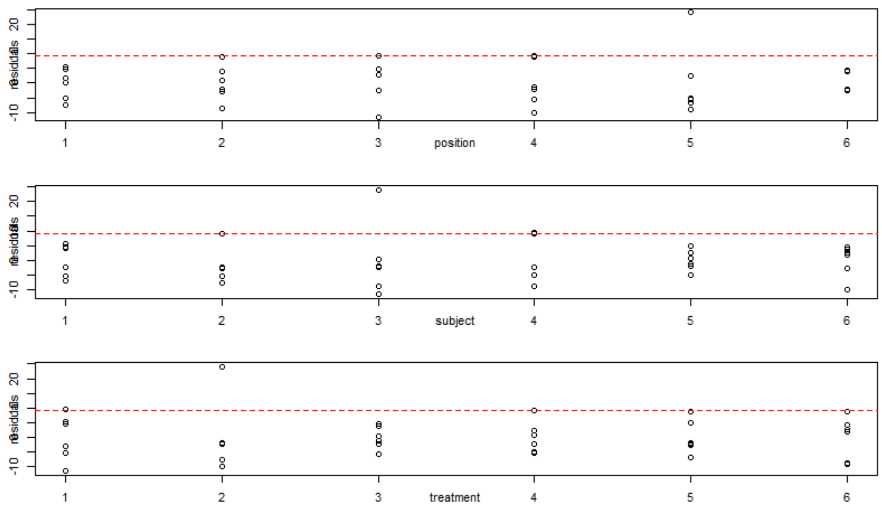
```
      Analysis of Variance Table

Response: y
          Df Sum Sq Mean Sq F value    Pr(>F)    
position   5  219.9   43.98  0.5148    0.7619    
subject    5 5407.9 1081.58 12.6600 1.248e-05 ***
treatment  5  250.3   50.05  0.5858    0.7107    
Residuals 20 1708.7   85.43                      

```

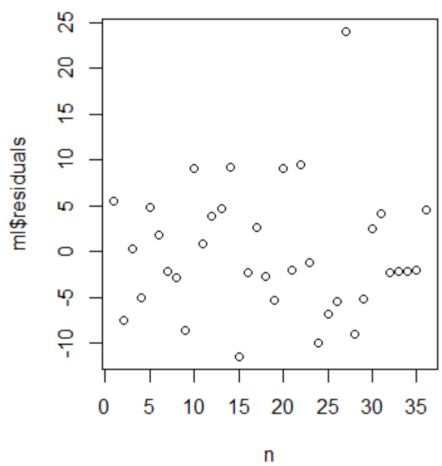
   And the graphical ANOVA



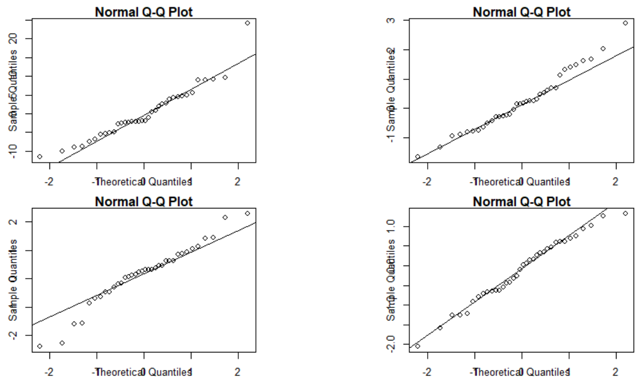
(d) The effects of positions are 0.
The effect of treatments are 0.
The effect of at lest one subject is not 0.

(e) We have





According to these figures, we can find there is an outlier under position 5, subject 3 and treatment B.
The distribution of errors are rather uniform along the variables and the index, so the independence of observations may be true.
Draw the qqplots of residuals and three sets of real standard normal variables.



The deviation of residuals are not worse than the real ones, so we accept that the normality is satisfied.

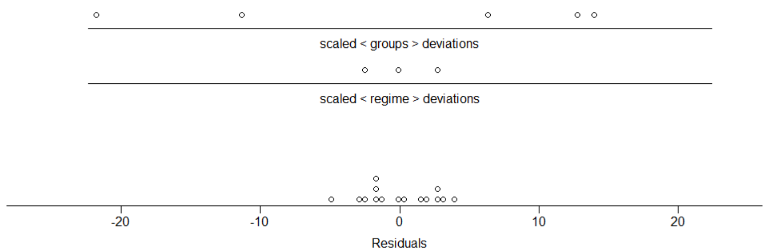
2. (a) The LSE under contrast treatment is

```
    Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   12.467      2.320   5.374 0.000667 ***
groups2        7.667      2.773   2.765 0.024490 *
groups3       20.000      2.773   7.213 9.13e-05 ***
groups4       24.667      2.773   8.896 2.02e-05 ***
groups5       25.333      2.773   9.136 1.66e-05 ***
regime2       -2.800      2.148  -1.304 0.228625
regime3       -1.600      2.148  -0.745 0.477628
```

According to the t-test, the coefficients of regimes may have no effect on the weight lost. So we investigate ANOVA:

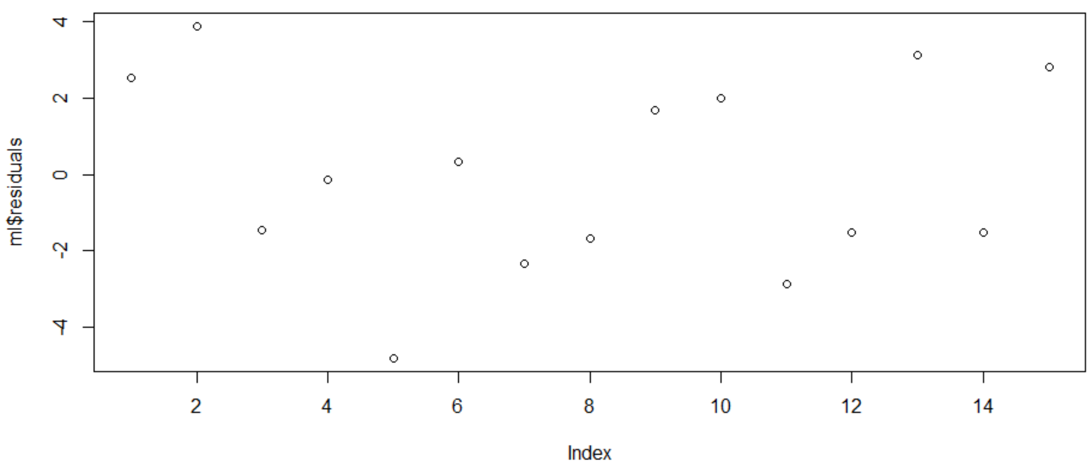
```
    Analysis of Variance Table

Response: y
          Df  Sum Sq Mean Sq F value    Pr(>F)
groups     4 1507.73  376.93 32.6821 5.274e-05 ***
regime     2   19.73    9.87  0.8555    0.4606
Residuals  8   92.27   11.53
```

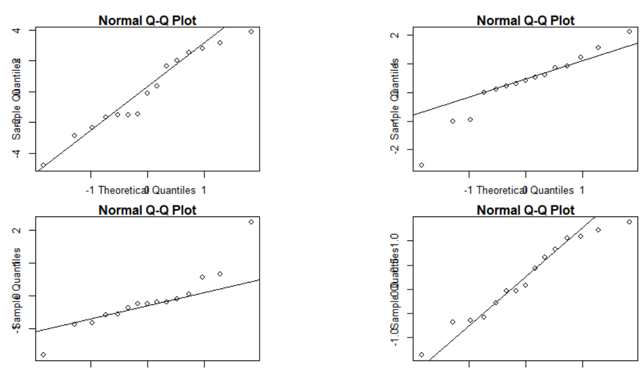
and the anova plot:



So we can conclude that the effects of different regimes are 0 under significance level 0.05.

Then we check the properties of residuals by plot the residuals and qqplot





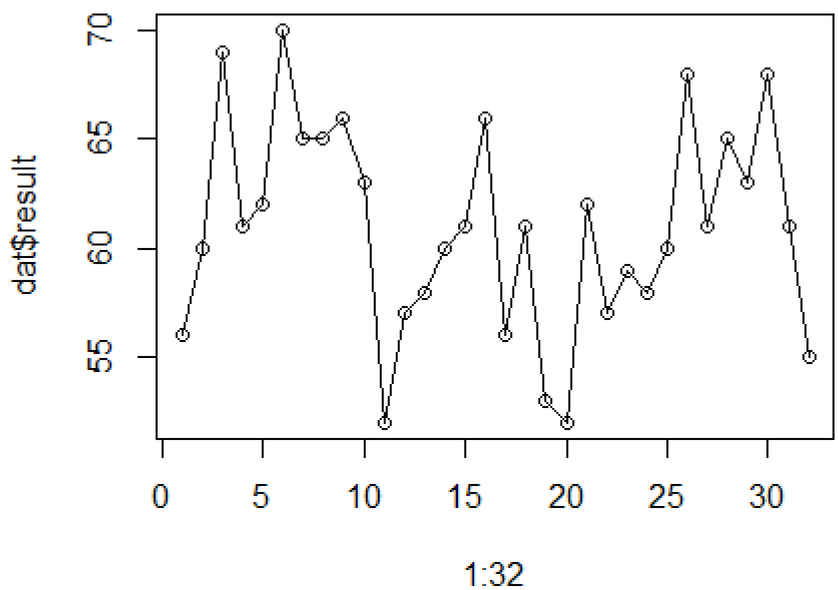
The trend of these plots suggests that NID assumption is satisfied.

(b) We compare the model involving effect of average weight and the a complex model involving average weight and groups.

```
    Analysis of Variance Table

Response: y
          Df  Sum Sq Mean Sq  F value    Pr(>F)
w          1 1371.41 1371.41 122.4474 6.238e-07 ***
groups     3  136.32   45.44   4.0572   0.03987 *
Residuals 10  112.00   11.20
```

So average weights and groups both affect the weight lost.

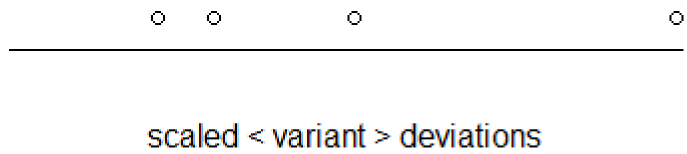
3. (a) Draw the results

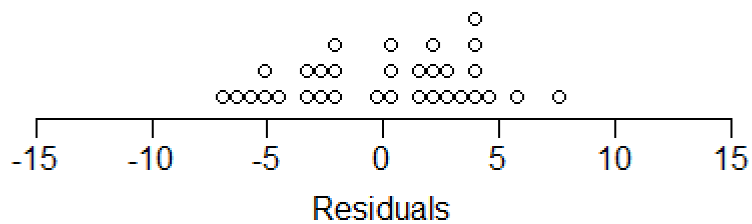


The experiment is randomized.

(b) The results of ANOVA are

```
      Analysis of Variance Table

Response: result
          Df Sum Sq Mean Sq F value   Pr(>F)
variant    3 254.38  84.792  5.1003 0.006092 **
Residuals 28 465.50  16.625
```





So we may conclude the effects of treatments are not all zero.

(c) Find the LSE without intercept

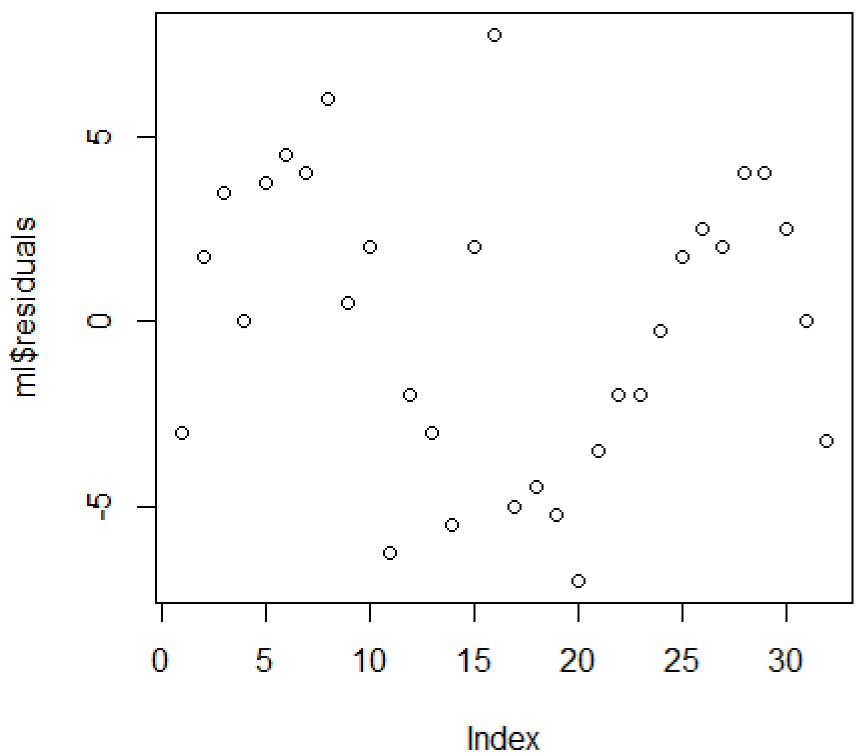
```
      Call:
lm(formula = result ~ variant - 1, data = dat)

Residuals:
   Min     1Q Median     3Q    Max 
-7.000 -3.062  0.250  2.750  7.750 

Coefficients:
         Estimate Std. Error t value Pr(>|t|)    
variantA   61.000      1.442   42.31   <2e-16 ***
variantB   58.250      1.442   40.41   <2e-16 ***
variantC   59.000      1.442   40.93   <2e-16 ***
variantD   65.500      1.442   45.44   <2e-16 ***

```

So the 95% confidence intervals are

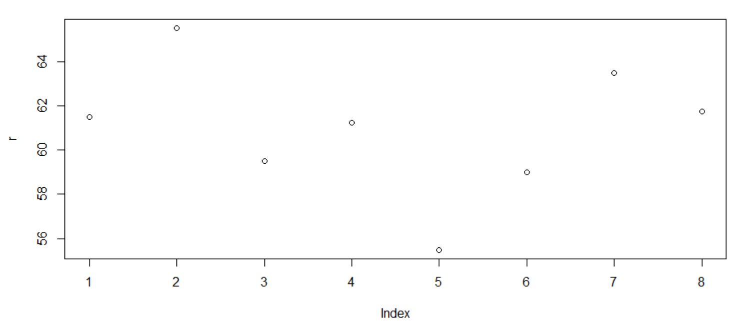
$$m_A \in (61 - 3 * 1.442, 61 + 3 * 1.442)$$

$$m_B \in (58.25 - 3 * 1.442, 58.25 + 3 * 1.442)$$
$$m_C \in (59 - 3 * 1.442, 59 + 3 * 1.442)$$
$$m_D \in (65.5 - 3 * 1.442, 65.5 + 3 * 1.442)$$

(d) We have the residuals



(e) The averages of blocks are



The process is approximately within the confidence intervals of the mean of treatments. It seems that the variation is somehow eliminated by randomization.

4. (a) We consider the following model:

$$Y_i = \varepsilon_i,\ \text{where } \varepsilon_i \sim \mathcal{N}(0,1) \quad \varepsilon_i\text{'s satisfy NID assumption}$$

i. H_0 is true. H_1 is not true.
ii. $lm(y \sim T + P + C + E + H + r)$ is true.
iii. $lm(y \sim T)$ is true.
iv. β is a 0 vector of 31 dimension. X is a 32-by-31 matrix. For each row,

$$X_i = (1, \boldsymbol{T}_i, \boldsymbol{P}_i, \boldsymbol{C}_i, \boldsymbol{E}_i, \boldsymbol{H}_i, \boldsymbol{r}_i),$$

where

$$\boldsymbol{T}_i = (\mathbb{1}_{[T_i=1]}, \mathbb{1}_{[T_i=2]}, \mathbb{1}_{[T_i=3]}, \mathbb{1}_{[T_i=4]}),$$
$$\boldsymbol{P}_i = (\mathbb{1}_{[P_i=1]}, \mathbb{1}_{[P_i=2]}, \mathbb{1}_{[P_i=3]}, \mathbb{1}_{[P_i=4]}),$$
$$\boldsymbol{C}_i = (\mathbb{1}_{[C_i=1]}, \mathbb{1}_{[C_i=2]}, \mathbb{1}_{[C_i=3]}, \mathbb{1}_{[C_i=4]}, \mathbb{1}_{[C_i=5]}, \mathbb{1}_{[C_i=6]}, \mathbb{1}_{[C_i=7]}, \mathbb{1}_{[C_i=8]}),$$
$$\boldsymbol{E}_i = (\mathbb{1}_{[E_i=1]}, \mathbb{1}_{[E_i=2]}, \mathbb{1}_{[E_i=3]}, \mathbb{1}_{[E_i=4]}, \mathbb{1}_{[E_i=5]}, \mathbb{1}_{[E_i=6]}, \mathbb{1}_{[E_i=7]}, \mathbb{1}_{[E_i=8]}),$$
$$\boldsymbol{H}_i = (\mathbb{1}_{[H_i=1]}, \mathbb{1}_{[H_i=2]}, \mathbb{1}_{[H_i=3]}, \mathbb{1}_{[H_i=4]}),$$
$$\boldsymbol{r}_i = (\mathbb{1}_{[r_i=1]}, \mathbb{1}_{[r_i=2]}).$$

However with contrast sum, we use another code system.
Now β is a 0 vector of 23 entries.
X is a 32-by-23 matrix. The entries of the first column are 1. For block T, if $T_i = j, j = 1,2,3$ then the corresponding j-th entry is 1 and the others are 0; if $T_i = 4$, then all entries are -1. Similar for H and P.
For the first three columns of block C, if $C_i = j, j = 1,2,3$ then the corresponding j-th entry is 1 and the others are 0; if $C_i = 4$, then all entries are -1. For the last three columns of block C, if $C_i = j+1, j = 4,5,6$ then the corresponding j-th entry is 1 and the others are 0; if $C_i = 8$, then all entries are -1. Similar for E.
For block r, if X_i is in replication I, it is 1, otherwise it is -1.

After first run,

$$\hat{\beta} = \begin{bmatrix} 0.31 \\ 0.02 \\ 0.21 \\ -0.36 \\ -0.11 \\ 0.33 \\ 0.01 \\ -0.01 \\ -0.42 \\ 0.18 \\ -0.57 \\ 0.84 \\ -0.59 \\ 0.52 \\ -0.09 \\ 0.19 \\ 0.39 \\ -0.56 \\ 0.04 \\ -0.09 \\ -0.26 \\ -0.11 \\ 0.08 \end{bmatrix}$$

v. I expect not to reject null hypothesis.

vi. After 100 runs of complete model, we have the mean and sd of $\hat{\beta}$ are

$$\begin{bmatrix} -0.01 \\ -0.02 \\ 0.01 \\ -0.00 \\ 0.05 \\ -0.02 \\ -0.03 \\ -0.03 \\ 0.04 \\ -0.02 \\ 0.08 \\ -0.08 \\ 0.04 \\ 0.00 \\ -0.07 \\ -0.03 \\ -0.06 \\ 0.06 \\ -0.02 \\ 0.08 \\ -0.08 \\ 0.03 \\ -0.04 \end{bmatrix}, s = \begin{bmatrix} 0.36 \\ 0.31 \\ 0.28 \\ 0.31 \\ 0.31 \\ 0.29 \\ 0.28 \\ 0.28 \\ 0.30 \\ 0.30 \\ 0.81 \\ 0.58 \\ 0.63 \\ 0.75 \\ 0.63 \\ 0.69 \\ 0.46 \\ 0.40 \\ 0.42 \\ 0.45 \\ 0.47 \\ 0.45 \\ 0.35 \end{bmatrix}$$

The type 1 errors under 0.05 level are

$$\begin{bmatrix} 0.03 \\ 0.03 \\ 0.03 \\ 0.07 \\ 0.02 \\ 0.04 \end{bmatrix}$$

For incomplete model, we have

$$\bar{\hat{\beta}} = \begin{bmatrix} -0.02 \\ -0.02 \\ 0.01 \\ -0.00 \end{bmatrix}, s = \begin{bmatrix} 0.19 \\ 0.31 \\ 0.28 \\ 0.31 \end{bmatrix}$$

The type 1 error under 0.05 level is 0.04.

(b) We consider the following model

$$Y = \sum_{i=1}^{4} \sqrt{i}\mathbb{1}_{[T=i]} + \sum_{j=1}^{4} j\mathbb{1}_{[P=j]} + \sum_{k=1}^{4} k^2\mathbb{1}_{[H=k]} + \sum_{m=1}^{4} \sin m\mathbb{1}_{[C=m]} + \sum_{n=1}^{4} \cos n\mathbb{1}_{[E=n]} + \varepsilon_{ijkmn},$$

$$\varepsilon_{ijkmn} \sim \mathcal{N}(0,1) \quad \varepsilon_{ijkmn}\text{'s satisfy NID assumption.}$$

i. H_0 is not true. H_1 is true.

ii. $lm(y \sim T + P + C + E + H + r)$ is true.

iii. $lm(y \sim T)$ is not true.

iv. β is a vector of 31 dimension and

$$\begin{bmatrix} 0 \\ \sqrt(1) \\ \vdots \\ \sqrt(4) \\ 1 \\ \vdots \\ 4 \\ 1^2 \\ \vdots \\ 4^2 \\ \sin 1 \\ \vdots \\ \sin 4 \\ \sin 1 \\ \vdots \\ \sin 4 \\ \cos 1 \\ \vdots \\ \cos 4 \\ \cos 1 \\ \vdots \\ \cos 4 \\ 0 \\ 0. \end{bmatrix}$$

X is a 32-by-31 matrix. For each row,

$$X_i = (1, \boldsymbol{T}_i, \boldsymbol{P}_i, \boldsymbol{C}_i, \boldsymbol{E}_i, \boldsymbol{H}_i, \boldsymbol{r}_i),$$

where

$$\boldsymbol{T}_i = (\mathbb{1}_{[T_i=1]}, \mathbb{1}_{[T_i=2]}, \mathbb{1}_{[T_i=3]}, \mathbb{1}_{[T_i=4]}),$$
$$\boldsymbol{P}_i = (\mathbb{1}_{[P_i=1]}, \mathbb{1}_{[P_i=2]}, \mathbb{1}_{[P_i=3]}, \mathbb{1}_{[P_i=4]}),$$
$$\boldsymbol{C}_i = (\mathbb{1}_{[C_i=1]}, \mathbb{1}_{[C_i=2]}, \mathbb{1}_{[C_i=3]}, \mathbb{1}_{[C_i=4]}, \mathbb{1}_{[C_i=5]}, \mathbb{1}_{[C_i=6]}, \mathbb{1}_{[C_i=7]}, \mathbb{1}_{[C_i=8]}),$$
$$\boldsymbol{E}_i = (\mathbb{1}_{[E_i=1]}, \mathbb{1}_{[E_i=2]}, \mathbb{1}_{[E_i=3]}, \mathbb{1}_{[E_i=4]}, \mathbb{1}_{[E_i=5]}, \mathbb{1}_{[E_i=6]}, \mathbb{1}_{[E_i=7]}, \mathbb{1}_{[E_i=8]}),$$
$$\boldsymbol{H}_i = (\mathbb{1}_{[H_i=1]}, \mathbb{1}_{[H_i=2]}, \mathbb{1}_{[H_i=3]}, \mathbb{1}_{[H_i=4]}),$$
$$\boldsymbol{r}_i = (\mathbb{1}_{[r_i=1]}, \mathbb{1}_{[r_i=2]}).$$

However with contrast sum, we use another code system.
Now β is a vector of 23 entries

$$\begin{bmatrix} 0 \\ \sqrt(1) \\ \vdots \\ \sqrt(3) \\ 1 \\ \vdots \\ 3 \\ 1^2 \\ \vdots \\ 3^2 \\ \sin 1 \\ \vdots \\ \sin 3 \\ \sin 1 \\ \vdots \\ \sin 3 \\ \cos 1 \\ \vdots \\ \cos 3 \\ \cos 1 \\ \vdots \\ \cos 3 \\ 0. \end{bmatrix}$$

X is a 32-by-23 matrix. The entries of the first column are 1. For block T, if $T_i = j, j = 1, 2, 3$ then the corresponding j-th entry is 1 and the others are 0; if $T_i = 4$, then all entries are -1. Similar for H and P.
For the first three columns of block C, if $C_i = j, j = 1, 2, 3$ then the corresponding j-th entry is 1 and the others are 0; if $C_i = 4$, then all entries are -1. For the last three columns of block C, if $C_i = j + 1, j = 4, 5, 6$ then the corresponding j-th entry is 1 and the others are 0; if $C_i = 8$, then all entries are -1. Similar for E.
For block r, if X_i is in replication I, it is 1, otherwise it is -1.

After first run, the estimated coefficients of complete model is

$$\begin{bmatrix} 11.56 \\ -0.62 \\ 0.14 \\ 0.42 \\ -6.99 \\ -3.53 \\ 2.11 \\ -1.28 \\ -0.55 \\ 0.52 \\ 0.38 \\ -0.26 \\ 0.34 \\ 0.68 \\ 1.01 \\ 0.24 \\ 0.83 \\ 0.38 \\ -1.14 \\ 0.91 \\ 0.26 \\ -0.78 \\ -0.02 \end{bmatrix}$$

The estimated coefficients of incomplete model is

$$\hat{\beta} = \begin{bmatrix} 11.56 \\ -1.02 \\ -0.22 \\ 0.17 \end{bmatrix}$$

v. I expect to reject null hypothesis for T, P, H, C, E but not to reject the hypothesis for r.

vi. After 100 runs of complete model, we have the mean and sd of $\hat{\beta}$ are

$$\begin{bmatrix} 11.44 \\ -0.56 \\ -0.14 \\ 0.23 \\ -1.51 \\ -0.51 \\ 0.47 \\ 0.58 \\ 0.62 \\ -0.13 \\ 0.50 \\ 0.66 \\ -0.14 \\ 0.95 \\ -0.01 \\ -0.66 \\ 0.92 \\ -0.11 \\ -0.59 \\ -6.49 \\ -3.48 \\ 1.49 \\ -0.01 \end{bmatrix}, s = \begin{bmatrix} 0.18 \\ 0.31 \\ 0.31 \\ 0.32 \\ 0.32 \\ 0.30 \\ 0.32 \\ 0.48 \\ 0.49 \\ 0.43 \\ 0.45 \\ 0.42 \\ 0.48 \\ 0.47 \\ 0.48 \\ 0.51 \\ 0.43 \\ 0.47 \\ 0.47 \\ 0.36 \\ 0.30 \\ 0.33 \\ 0.18 \end{bmatrix}$$

The type 2 errors under 0.05 level are

$$\begin{bmatrix} 0.77 \\ 0.00 \\ 0.52 \\ 0.63 \\ 0.00 \\ 0.96 \end{bmatrix}$$

For incomplete model, we have

$$\begin{bmatrix} 11.44 \\ -0.56 \\ -0.14 \\ 0.23 \end{bmatrix}, s = \begin{bmatrix} 0.18 \\ 0.31 \\ 0.31 \\ 0.32 \end{bmatrix}$$

The type 2 error under 0.05 level is 1.

(c) We consider the following model

$$Y_{ij} = \sum_{i=1}^{4} \sqrt{i}\mathbb{1}_{[T=i]} + \sum_{i=1}^{4} i\mathbb{1}_{[P=i]} + \sum_{i=1}^{4} i^2\mathbb{1}_{[H=i]} + \sum_{i=1}^{4} \sin i\mathbb{1}_{[C=i]} + \sum_{i=1}^{4} \cos i\mathbb{1}_{[E=i]} + \varepsilon_{ij},$$

$$\varepsilon_{ij} \sim \mathcal{N}(0,1) \quad \varepsilon_{ij}\text{'s satisfy NID assumption.}$$

i. H_0 is not true. H_1 is true.

ii. $lm(y \sim T + P + C + E + H + r)$ is true.

iii. $lm(y \sim T)$ is not true.

iv. β is a vector of 31 dimension and

$$\begin{bmatrix} 0 \\ \sqrt(1) \\ \vdots \\ \sqrt(4) \\ 1 \\ \vdots \\ 4 \\ 1^2 \\ \vdots \\ 4^2 \\ \sin 1 \\ \vdots \\ \sin 4 \\ \sin 1 \\ \vdots \\ \sin 4 \\ \cos 1 \\ \vdots \\ \cos 4 \\ \cos 1 \\ \vdots \\ \cos 4 \\ 0 \\ 0. \end{bmatrix}$$

X is a 32-by-31 matrix. For each row,

$$X_i = (1, \boldsymbol{T}_i, \boldsymbol{P}_i, \boldsymbol{C}_i, \boldsymbol{E}_i, \boldsymbol{H}_i, \boldsymbol{r}_i),$$

where

$$\boldsymbol{T}_i = (\mathbb{1}_{[T_i=1]}, \mathbb{1}_{[T_i=2]}, \mathbb{1}_{[T_i=3]}, \mathbb{1}_{[T_i=4]}),$$
$$\boldsymbol{P}_i = (\mathbb{1}_{[P_i=1]}, \mathbb{1}_{[P_i=2]}, \mathbb{1}_{[P_i=3]}, \mathbb{1}_{[P_i=4]}),$$
$$\boldsymbol{C}_i = (\mathbb{1}_{[C_i=1]}, \mathbb{1}_{[C_i=2]}, \mathbb{1}_{[C_i=3]}, \mathbb{1}_{[C_i=4]}, \mathbb{1}_{[C_i=5]}, \mathbb{1}_{[C_i=6]}, \mathbb{1}_{[C_i=7]}, \mathbb{1}_{[C_i=8]}),$$
$$\boldsymbol{E}_i = (\mathbb{1}_{[E_i=1]}, \mathbb{1}_{[E_i=2]}, \mathbb{1}_{[E_i=3]}, \mathbb{1}_{[E_i=4]}, \mathbb{1}_{[E_i=5]}, \mathbb{1}_{[E_i=6]}, \mathbb{1}_{[E_i=7]}, \mathbb{1}_{[E_i=8]}),$$
$$\boldsymbol{H}_i = (\mathbb{1}_{[H_i=1]}, \mathbb{1}_{[H_i=2]}, \mathbb{1}_{[H_i=3]}, \mathbb{1}_{[H_i=4]}),$$
$$\boldsymbol{r}_i = (\mathbb{1}_{[r_i=1]}, \mathbb{1}_{[r_i=2]}).$$

However with contrast sum, we use another code system.
Now β is a vector of 23 entries

$$\begin{bmatrix} 0 \\ \sqrt(1) \\ \vdots \\ \sqrt(3) \\ 1 \\ \vdots \\ 3 \\ 1^2 \\ \vdots \\ 3^2 \\ \sin 1 \\ \vdots \\ \sin 3 \\ \sin 1 \\ \vdots \\ \sin 3 \\ \cos 1 \\ \vdots \\ \cos 3 \\ \cos 1 \\ \vdots \\ \cos 3 \\ 0. \end{bmatrix}$$

X is a 32-by-23 matrix. The entries of the first column are 1. For block T, if $T_i = j, j = 1, 2, 3$ then the corresponding j-th entry is 1 and the others are 0; if $T_i = 4$, then all entries are -1. Similar for H and P.
For the first three columns of block C, if $C_i = j, j = 1, 2, 3$ then the corresponding j-th entry is 1 and the others are 0; if $C_i = 4$, then all entries are -1. For the last three columns of block C, if $C_i = j+1, j = 4, 5, 6$ then the corresponding j-th entry is 1 and the others are 0; if $C_i = 8$, then all entries are -1. Similar for E.
For block r, if X_i is in replication I, it is 1, otherwise it is -1.

After first run, the estimated coefficients of complete model is

$$\hat{\beta} = \begin{bmatrix} 11.56 \\ -6.94 \\ -3.15 \\ 1.83 \\ NA \\ NA \\ NA \\ NA \\ NA \\ NA \\ -0.30 \\ -1.26 \\ 0.10 \\ NA \\ NA \\ NA \\ NA \\ NA \\ NA \\ NA \\ NA \\ NA \\ -0.02 \end{bmatrix}$$

The estimated coefficients of incomplete model is

$$\hat{\beta} = \begin{bmatrix} 11.56 \\ -1.02 \\ -0.22 \\ 0.17 \end{bmatrix}$$

v. I expect to reject null hypothesis for T, P, H, C, E but not to reject the hypothesis for r.

vi. After 100 runs of complete model, we have the mean and sd of $\hat{\beta}$ are

$$\bar{\hat{\beta}} = \begin{bmatrix} 11.44 \\ -7.11 \\ -3.50 \\ 1.44 \\ NA \\ NA \\ NA \\ 0.08 \\ -0.04 \\ 0.01 \\ NA \\ NA \\ NA \\ NA \\ NA \\ NA \\ NA \\ NA \\ NA \\ NA \\ NA \\ NA \\ -0.01 \end{bmatrix}, s = \begin{bmatrix} 0.18 \\ 0.45 \\ 0.42 \\ 0.48 \\ NA \\ NA \\ NA \\ 0.65 \\ 0.67 \\ 0.63 \\ NA \\ NA \\ NA \\ NA \\ NA \\ NA \\ NA \\ NA \\ NA \\ NA \\ NA \\ NA \\ 0.18 \end{bmatrix}$$

Since collinearity, there are only 3 variables on which we can do ANOVA: Treatment, Cycle and replication.
The type 2 errors under 0.05 level are

$$\begin{bmatrix} 0.00 \\ 0.96 \\ 0.97 \end{bmatrix}$$

For incomplete model, we have

$$\begin{bmatrix} 11.44 \\ -7.07 \\ -3.52 \\ 1.45 \end{bmatrix}, s = \begin{bmatrix} 0.18 \\ 0.34 \\ 0.32 \\ 0.33 \end{bmatrix}$$

The type 2 error under 0.05 level is 0.

5. (a) If a_i and b_i have the same sign, $(factor(ab))_i$ is 1; otherwise $(factor(ab))_i$ is -1

(b) The values are

$$\begin{bmatrix} y & a & b & c & ab & bc \\ 16.37 & -1 & -1 & -1 & 1 & 1 \\ -12.82 & 1 & -1 & -1 & -1 & 1 \\ -13.84 & -1 & 1 & -1 & -1 & -1 \\ 22.60 & 1 & 1 & -1 & 1 & -1 \\ 9.33 & -1 & -1 & 1 & 1 & -1 \\ -21.82 & 1 & -1 & 1 & -1 & -1 \\ -16.51 & -1 & 1 & 1 & -1 & 1 \\ 17.74 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

For ab, $\bar{y_+} = 16.51$, $\bar{y_-} = -16.25$. So main effect is 32.76. The interaction effect is 32.76.
For bc, $\bar{y_+} = 1.20$, $\bar{y_-} = -0.93$. So main effect is 2.13. The interaction effect is 2.13.

(c) Assume $\hat{\beta}_i'$'s are coefficients of model 2, $\hat{\beta}_i$'s are coefficients of model 1.
We suppose to have

$$\sum_{i=0}^{j} \beta_i' = \sum_{i=0}^{j} \beta_i - \sum_{i=j+1}^{5} \beta_i, i = 0, 1, 2, 3, 4, 5.$$

So we can solve it out.

$$\beta_i' = 2\beta_i, i = 1, 2, 3, 4, 5$$

$$\beta_0' = \beta_0 - \sum_{i=1}^{5} \beta_i$$

Assume $\hat{\beta}_i''$'s are coefficients of model 3.
We suppose to have

$$\beta_i'' = \beta_i' = 2\beta_i, i = 2, 3, 4, 5$$

$$\beta_0'' = \beta_0' = \beta_0 - \sum_{i=1}^{5} \beta_i$$

$$\beta_1'' = \beta_0'' + \beta_1' = \beta_0 - \sum_{i=1}^{5} \beta_i + 2\beta_1 = \beta_0 + \beta_1 - \sum_{i=2}^{5} \beta_i.$$

(d) We have

$$X = \begin{bmatrix} 1 & -1 & -1 & -1 & 1 & 1 \\ 1 & 1 & -1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 & -1 & -1 \\ 1 & 1 & 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 \\ 1 & 1 & -1 & 1 & -1 & -1 \\ 1 & -1 & 1 & 1 & -1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}, \beta = \begin{bmatrix} 0 \\ 1 \\ 2 \\ -3 \\ 16 \\ 1 \end{bmatrix}$$

(e) Model 1:

$$X = \begin{bmatrix} 1 & -1 & -1 & -1 & 1 & 1 \\ 1 & 1 & -1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 & -1 & -1 \\ 1 & 1 & 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 \\ 1 & 1 & -1 & 1 & -1 & -1 \\ 1 & -1 & 1 & 1 & -1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}, \beta = \begin{bmatrix} 0 \\ 1 \\ 2 \\ -3 \\ 16 \\ 1 \end{bmatrix}$$

Model 2:

$$X = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}, \beta = \begin{bmatrix} -17 \\ 2 \\ 4 \\ -6 \\ 32 \\ 2 \end{bmatrix}$$

Model 3:

$$X = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}, \beta = \begin{bmatrix} -17 \\ -15 \\ 4 \\ -6 \\ 32 \\ 2 \end{bmatrix}$$

6. (a) Let coefficients of a, b, c, ab, bc be $1, 4, 9, 16, 25$.
Under these settings, H_0^i's are not true. The model lm(yã+b+c+ab+bc) is true.
The anova table is

```
    Analysis of Variance Table

Response: y
          Df  Sum Sq Mean Sq   F value    Pr(>F)
a          1     7.8     7.8    2.5002    0.1449
b          1   409.4   409.4  131.7020 4.440e-07 ***
c          1  1172.7  1172.7  377.2756 2.856e-09 ***
ab         1  4303.7  4303.7 1384.5289 4.680e-12 ***
bc         1 10281.9 10281.9 3307.7748 6.129e-14 ***
Residuals 10    31.1     3.1

```

For b, c, ab, bc, we tend to reject null hypothesis. Since the effect of a is small here, F-test doesn't reject null hypothesis.
The mean square error of residuals is an unbiased estimations of σ^2 because the model

here is true. The others are not because H_0^i's are not true.
As for the model lm(yã+b+c), H_0^i's are not true. The model is not true.
We can obtain the anova table:

```
    Analysis of Variance Table

Response: y
          Df  Sum Sq Mean Sq F value Pr(>F)
a          1     7.8    7.77  0.0064 0.9377
b          1   409.4  409.38  0.3361 0.5728
c          1  1172.7 1172.72  0.9628 0.3459
Residuals 12 14616.6 1218.05

```

F-tests tend not to reject null hypothesis. The mean squares are not unbiased estimator of $sigma^2 = 4$.

(b) Now we set $Y_i = \varepsilon_i \sim \mathcal{N}(1, 4)$.
For model lm(yã+b+c), H_0^i's are true. The model is correct.
Anova table is as follow:

```
    Analysis of Variance Table

Response: y
          Df Sum Sq Mean Sq F value  Pr(>F)
a          1 18.796 18.7960  4.1065 0.06554 .
b          1  1.043  1.0433  0.2279 0.64163
c          1  4.909  4.9090  1.0725 0.32079
Residuals 12 54.925  4.5771

```

For each H_0^i, F-test doesn't reject null hypothesis. All means square errors are unbiased estimator of σ^2. But we trust the one with the highest degree of freedom.
For model lm(yã+b+c+ab+bc), H_0^i's are true. The model is correct.
Anova table is as follow:

```
    Analysis of Variance Table

Response: y
          Df Sum Sq Mean Sq F value  Pr(>F)
a          1 18.796 18.7960  4.0284 0.07253 .
b          1  1.043  1.0433  0.2236 0.64646
c          1  4.909  4.9090  1.0521 0.32919
ab         1  8.178  8.1775  1.7526 0.21501
bc         1  0.089  0.0889  0.0191 0.89295
Residuals 10 46.659  4.6659

```

For each H_0^i, F-test reject null hypothesis. All means square errors are unbiased estimator of σ^2. But we trust the one with the highest degree of freedom.

(c) Consider model $y = a + b + \varepsilon$ where $\varepsilon \sim \mathcal{N}(1, 2)$.
For lm(yã), we have

```
    Analysis of Variance Table

Response: y
           Df  Sum Sq Mean Sq F value    Pr(>F)
a           1  292.51 292.511  49.366 3.341e-11 ***
Residuals 198 1173.22   5.925

```

H_0^i's are false and the model is incorrect. But F-test reject null hypothesis.
For lm(y a+b), we have

```
  Analysis of Variance Table

Response: y
           Df Sum Sq Mean Sq F value    Pr(>F)
a           1 292.51 292.511  63.525 1.253e-13 ***
b           1 266.10 266.095  57.788 1.163e-12 ***
Residuals 197 907.12   4.605
```

H_0^i's are false and the model is correct. The F-tests tend to reject null hypothesis. Hence the results of anova can not be evidence to reject null hypothesis or not when the model is wrong.

(d) Consider $y = 10a - 27b + e$. Then we have

$$Y_i = 1 + (-10)\mathbb{1}_{[a_i=-1]} + 10\mathbb{1}_{[a_i=1]} + 27\mathbb{1}_{[b_i=-1]} + (-27)\mathbb{1}_{[b_i=1]} + \varepsilon_i.$$

suppose we compute LSE with contr.treatment. Then the model is

$$Y_i = 18 + 20\mathbb{1}_{[a_i=1]} - 54\mathbb{1}_{[b_i=1]} + \varepsilon_i.$$

And the estimated coefficients are

```
  Call:
lm(formula = y ~ a + b + c)

Residuals:
    Min      1Q  Median      3Q     Max
-4.5755 -1.2633 -0.1997  1.1226  6.4768

Coefficients:
            Estimate Std. Error  t value Pr(>|t|)
(Intercept)  17.3434     0.4658   37.233   <2e-16 ***
a1           20.1967     0.4658   43.359   <2e-16 ***
b1          -53.5081     0.4658 -114.872   <2e-16 ***
c1           -0.1439     0.4658   -0.309    0.758
```

The anova table is

```
  Analysis of Variance Table

Response: y
          Df Sum Sq Mean Sq    F value Pr(>F)    
a          1   8158    8158  1879.9727 <2e-16 ***
b          1  57262   57262 13195.5964 <2e-16 ***
c          1      0       0     0.0954 0.7582    
Residuals 76    330       4                      
```

H_0^i's are not true and the model is correct. So mean square error of residuals is unbiased estimation of σ^2.