回归分析 Homework 03

姓名: 雍崔扬
学号: 21307140051

Problem 1

Data on productivity improvements for a sample of firms producing electronic computing equipment.
The firms were classified according to the **level of their average expenditures** for research **(low, moderate, high)**.
The results of the study follow (productivity improvement is measured on a scale from 0 to 100).
Apply the ANOVA model.

Level	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Low	7.6	8.2	6.8	5.8	6.9	6.6	6.3	7.7	6.0			
Moderate	6.7	8.1	9.4	8.6	7.8	7.7	8.9	7.9	8.3	8.7	7.1	8.4
High	8.5	9.7	10.1	7.8	9.6	9.5						

- **(a)** Write down the model and state clearly the assumptions.
- **(b)** Obtain the fitted values.
- **(c)** Obtain the residuals. Do they sum to zero?
- **(d)** Obtain the ANOVA table.
- **(e)** Test whether or not the mean productivity improvement differs according to the level of research and development expenditures.
  Control α at 0.05.
  State the hypotheses, decision rule, and conclusion.

α=0.05

Fα k1 k2	1	2	3	4	5	6	8	12	24	∞
1	161.4	199.5	215.7	224.6	230.2	234.0	238.9	243.9	249.0	254.3
2	18.51	19.00	19.16	19.25	19.30	19.33	19.37	19.41	19.45	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.84	8.74	8.64	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.04	5.91	5.77	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.82	4.68	4.53	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.15	4.00	3.84	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.73	3.57	3.41	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.44	3.28	3.12	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.23	3.07	2.90	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.07	2.91	2.74	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	2.95	2.79	2.61	2.40
12	4.75	3.88	3.49	3.26	3.11	3.00	2.85	2.69	2.50	2.30
13	4.67	3.80	3.41	3.18	3.02	2.92	2.77	2.60	2.42	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.70	2.53	2.35	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.64	2.48	2.29	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.59	2.42	2.24	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.55	2.38	2.19	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.51	2.34	2.15	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.48	2.31	2.11	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.45	2.28	2.08	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.42	2.25	2.05	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.40	2.23	2.03	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.38	2.20	2.00	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.36	2.18	1.98	1.73
25	4.24	3.38	2.99	2.76	2.60	2.49	2.34	2.16	1.96	1.71
26	4.22	3.37	2.98	2.74	2.59	2.47	2.32	2.15	1.95	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.30	2.13	1.93	1.67
28	4.20	3.34	2.95	2.71	2.56	2.44	2.29	2.12	1.91	1.65
29	4.18	3.33	2.93	2.70	2.54	2.43	2.28	2.10	1.90	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.27	2.09	1.89	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.18	2.00	1.79	1.51
60	4.00	3.15	2.76	2.52	2.37	2.25	2.10	1.92	1.70	1.39
120	3.92	3.07	2.68	2.45	2.29	2.17	2.02	1.83	1.61	1.25
∞	3.84	2.99	2.60	2.37	2.21	2.09	1.94	1.75	1.52	1.00

Part (a)

Write down the model and state clearly the assumptions.

Solution:
因子 A 代表过去三年平均研究经费，有三个因子水平:

$$A_1 = \text{Low}$$
$$A_2 = \text{Moderate}$$
$$A_3 = \text{High}$$

模型: $y_{ij} = \mu_i + \varepsilon_{ij}$
其中:

$$
\begin{gathered}
i = 1, 2, 3 \\
j = 1, \ldots, n_i \text{ for any } i = 1, 2, 3 \\
n_1 = 9 \\
n_2 = 12 \\
n_3 = 6
\end{gathered}
$$

而随机噪音 ε_{ij} 满足独立性假设、同方差假设和正态假设:

$$
\{\varepsilon_{ij}\} \overset{\text{i.i.d.}}{\sim} N(0, \sigma^2)
$$

Part (b)

Find the parameter estimates.

Solution:

$$
\begin{aligned}
\hat{\mu}_1 &= \bar{y}_{1\cdot} = \frac{1}{n_1} \sum_{j=1}^{n_1} y_{1j} = \frac{61.9}{9} \approx 6.88 \\
\hat{\mu}_2 &= \bar{y}_{2\cdot} = \frac{1}{n_2} \sum_{j=1}^{n_2} y_{2j} = \frac{97.6}{12} \approx 8.13 \\
\hat{\mu}_3 &= \bar{y}_{3\cdot} = \frac{1}{n_3} \sum_{j=1}^{n_3} y_{3j} = \frac{55.2}{6} = 9.20 \\
s^2 &= \frac{1}{n-3} \sum_{i=1}^{3} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i\cdot})^2 \approx 0.64
\end{aligned}
$$

Part (c)

Obtain the residuals.
Do they sum to zero?

Solution:
residual $\hat{\varepsilon}_{ij} = y_{ij} - \bar{y}_{i\cdot}$

Level	1	2	3	4	5	6	7	8	9	10	11	12
Low	0.72	1.32	-0.08	-1.08	0.02	-0.28	-0.58	0.82	-0.88			
Moderate	-1.43	-0.03	1.27	0.47	-0.33	-0.43	0.77	-0.23	0.17	0.57	-1.03	0.27
High	-0.70	0.50	0.90	-1.40	0.40	0.30						

The row sums are zero (if the computational error is omitted)

Part (d) $(\star)$

Obtain the ANOVA table.

Solution:

Sum of Squares	Value	Degree of Freedom	Mean Squares	F-Value
$\mathrm{SST} = \sum_{i=1}^{3} \sum_{j=1}^{n_j} (y_{ij} - \bar{y}_{\cdot\cdot})^2$	35.4874	$n - 1 = 26$	$\mathrm{MST} = \frac{\mathrm{SST}}{26} = 1.3549$	$F = \frac{\mathrm{MSA}}{\mathrm{MSE}} = 15.7205$
$\mathrm{SSE} = \sum_{i=1}^{3} \sum_{j=1}^{n_j} (y_{ij} - \bar{y}_{i\cdot})^2$	15.3622	$n - 3 = 24$	$\mathrm{MSE} = \frac{\mathrm{SSE}}{24} = 0.6401$	
$\mathrm{SSA} = \sum_{i=1}^{3} n_i (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2 = \mathrm{SST} - \mathrm{SSE}$	20.1252	$3 - 1 = 2$	$\mathrm{MSA} = \frac{\mathrm{SSA}}{2} = 10.0626$	

Part (e)

Test whether or not the mean productivity improvement differs according to the level of research and development expenditures.
Control α at 0.05.
State the hypotheses, decision rule, and conclusion.

Solution:
零假设和备择假设为:

$$H_0 : \mu_1 = \mu_2 = \mu_3 \quad \text{v.s.} \quad H_1 : \exists\, i, j \text{ such that } \mu_i \neq \mu_j$$

检验统计量为:

$$F = \frac{\text{MSA}}{\text{MSE}} = 15.7205 \overset{H_0}{\sim} F_{2,24}$$

设第一类型错误概率界限 $\alpha = 0.05$
查表得 $F_{2,24}$ 的 $1-\alpha$ 分位数 $F_{2,24}(\alpha) \approx 3.40$
因此 $F \approx 15.72 > 3.40 = F_{2,24}(\alpha)$，于是我们拒绝零假设 $H_0 : \mu_1 = \mu_2 = \mu_3$
即认为因子 A "过去三年平均研究经费" 对利润增长 (productivity improvement) 有解释作用.

Problem 2

Consider a one-way ANOVA model.
By rewriting the model using the regression approach,
prove that $\text{SSE} \sim \sigma^2 \chi^2_{(n-a)}$ and is independent of $\bar{y}_{i\cdot}$

Solution:
考虑单因子方差分析的多元线性回归形式:

$$y = \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(a)} \end{bmatrix} = \begin{bmatrix} 1_{n_1} & & & \\ & 1_{n_2} & & \\ & & \ddots & \\ & & & 1_{n_a} \end{bmatrix}_{n\times a} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_a \end{bmatrix} + \begin{bmatrix} \varepsilon^{(1)} \\ \varepsilon^{(2)} \\ \vdots \\ \varepsilon^{(a)} \end{bmatrix} = X\mu + \varepsilon$$

其中:

$$y^{(i)} = \begin{bmatrix} y_{i,1} \\ y_{i,2} \\ \vdots \\ y_{i,n_i} \end{bmatrix} \quad \varepsilon^{(i)} = \begin{bmatrix} \varepsilon_{i,1} \\ \varepsilon_{i,2} \\ \vdots \\ \varepsilon_{i,n_i} \end{bmatrix} \quad (i = 1, \dots, a)$$

$$n = n_1 + n_2 + \cdots + n_a$$

$$\{\varepsilon_{i,j}\} \overset{\text{i.i.d.}}{\sim} N(0, \sigma^2)$$

设计矩阵 $X = 1_{n_1} \oplus \cdots \oplus 1_{n_a} \in \mathbb{R}^{n\times a}$
我们定义投影矩阵为:

$$
\begin{aligned}
H &:= X(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}} \\
&= \begin{bmatrix} 1_{n_1} & & & \\ & 1_{n_2} & & \\ & & \ddots & \\ & & & 1_{n_a} \end{bmatrix} \begin{bmatrix} n_1 & & & \\ & n_2 & & \\ & & \ddots & \\ & & & n_a \end{bmatrix}^{-1} \begin{bmatrix} 1_{n_1} & & & \\ & 1_{n_2} & & \\ & & \ddots & \\ & & & 1_{n_a} \end{bmatrix}^{\mathrm{T}} \\
&= \begin{bmatrix} \frac{1}{n_1}1_{n_1}1_{n_1}^{\mathrm{T}} & & & \\ & \frac{1}{n_2}1_{n_2}1_{n_2}^{\mathrm{T}} & & \\ & & \ddots & \\ & & & \frac{1}{n_a}1_{n_a}1_{n_a}^{\mathrm{T}} \end{bmatrix}
\end{aligned}
$$

根据多元线性回归中的结论我们有:

$$
\operatorname{tr}(I_n - H) = n - \operatorname{tr}\left(X(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}\right) = n - \operatorname{tr}\left((X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}X\right) = n - \operatorname{tr}(I_a) = n - a
$$

$$
\Rightarrow
$$

$$
\mathrm{SSE} = \|y - \hat{y}\|^2 = \|y - Hy\|^2 = \|(I_n - H)y\|^2 = y^{\mathrm{T}}(I_n - H)y \sim \sigma^2\chi^2_{(n-a)}
$$

要证明 SSE 和 $\bar{y}_{i\cdot}$ 相互独立，只需证明 $(I_n - H)y$ 和 $\bar{y}_{i\cdot}$ 不相关即可 (因为正态假设下不相关等价于独立)
记 $\bar{y} = [\bar{y}_{1\cdot}, \ldots, \bar{y}_{a\cdot}]^{\mathrm{T}} = (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}y$

$$
\begin{aligned}
\operatorname{Cov}((I_n - H)y, \bar{y}) &= \operatorname{Cov}((I_n - H)y, (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}y) \\
&= (I_n - H)\operatorname{Cov}(y, y)[(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}]^{\mathrm{T}} \\
&= (I_n - H) \cdot \sigma^2 I_n \cdot X(X^{\mathrm{T}}X)^{-1} \quad (\text{note that } HX = X \Rightarrow (I_n - H)X = 0_{n\times a}) \\
&= 0_{n\times a}
\end{aligned}
$$

因此 SSE 与 $\bar{y} = [\bar{y}_{1\cdot}, \ldots, \bar{y}_{a\cdot}]^{\mathrm{T}}$ 相互独立，即 SSE 和 $\bar{y}_{i\cdot}$ 相互独立.

Problem 3

A study was conducted to determine the joint effects of temperature and concentration of herbicide on absorption of 2 herbicides on a commercial charcoal material.
The data are given in the table.
There were 2 observations for each treatment.

Temperature \ **Concentration**	20	40
10	0.28	0.38
10	0.278	0.392
55	0.266	0.332
55	0.258	0.334

- **(a)** Consider a linear regression model of the absorption on temperature and concentration (as quantitative independent variables)
  - ① Write down the regression model. State the assumptions.
  - ② Find the least square estimates.

- ③ Obtain the ANOVA table and hence test whether there is a regression between the absorption and the 2 independent variables at 5% level of significance.

- **(b)** Consider a two-way ANOVA model with interaction.
  - ① Write down the two-way ANOVA model.
    State the assumptions.
    Find the parameter estimates.
  - ② Obtain the ANOVA table.
    Is there a regression between the dependent variable and the 2 factors at the 5% level of significance?
  - ③ Test whether or not the two factors interact; use $\alpha = 0.05$.
    State the alternatives, decision rule and conclusion.
  - ④ Test whether or not the main effects for the two ingredients are present.
    Use $\alpha = 0.05$ in each case and state the alternatives, decision rule and conclusion.
    Is it meaningful here to test main factor effects? Explain.

Part (a)

(a) Consider a linear regression model of the absorption on temperature and concentration (as quantitative independent variables)

- ① Write down the regression model. State the assumptions.
- ② Find the least square estimates.
- ③ Obtain the ANOVA table and hence test whether there is a regression between the absorption and the 2 independent variables at 5% level of significance.

Solution:
模型 $y = X\beta + \varepsilon$

$$y = \begin{bmatrix} 0.28 \\ 0.278 \\ 0.38 \\ 0.392 \\ 0.266 \\ 0.258 \\ 0.332 \\ 0.334 \end{bmatrix} \quad X = \begin{bmatrix} 1 & 10 & 20 \\ 1 & 10 & 20 \\ 1 & 10 & 40 \\ 1 & 10 & 40 \\ 1 & 55 & 20 \\ 1 & 55 & 20 \\ 1 & 55 & 40 \\ 1 & 55 & 40 \end{bmatrix} \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \\ \varepsilon_7 \\ \varepsilon_8 \end{bmatrix} \quad \begin{cases} n = 8 \\ p = 2 \end{cases}$$

假设:

- ① 线性关系假设: $\mathrm{E}[y|X] = X\beta$
- ② 同方差假设: $\mathrm{Var}(\varepsilon_i) = \sigma^2 \ (i = 1, \dots, 8)$
- ③ 不相关假设: $\mathrm{Cov}(\varepsilon_i, \varepsilon_j) = 0 \ (\forall \ i \neq j)$
- ④ 正态假设: ε_i 服从正态分布.

总而言之，我们有:

$$\varepsilon \sim N(0_n, \sigma^2 I_n)$$

模型参数的点估计量:

- β 的最小二乘估计量:

$$\hat{\beta} = (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}y = \begin{bmatrix} 0.207 \\ -0.000778 \\ 0.00445 \end{bmatrix}$$

- σ^2 的无偏估计量:

$$s^2 = \frac{1}{n-p-1}\|y-\hat{y}\|^2 = \frac{1}{n-3}y^{\mathrm{T}}(I_n - H)y = 0.000151$$
$$\text{where } H := X(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}$$

线性回归显著性检验:

$$H_0: \beta_1 = \beta_2 = 0 \quad \text{v.s} \quad H_1: \exists\, i = 1,2 \text{ such that } \beta_i \neq 0$$

检验统计量:

$$F = \frac{\text{MSR}}{\text{MSE}} \overset{H_0}{\sim} F_{p,n-p-1} = F_{2,5}$$

ANOVA Table:

Sum of Squares	Value	Degree of Freedom	Mean Squares	F-Value
$\text{SST} = \|y - \bar{y}1_n\|^2$	0.019048	$n-1=7$	$\text{MST} = \frac{\text{SST}}{7} = 0.002721$	$F = \frac{\text{MSR}}{\text{MSE}} = 60.489418$
$\text{SSE} = \|y-\hat{y}\|^2$	0.000756	$n-p-1=5$	$\text{MSE} = \frac{\text{SSE}}{5} = 0.000151$	
$\text{SSR} = \|\hat{y} - \bar{y}1_n\|^2 = \text{SST} - \text{SSE}$	0.018292	2	$\text{MSR} = \frac{\text{SSA}}{2} = 0.009146$	

设第一类型错误概率界限为 $\alpha = 0.05$
查表可知 $F_{2,5}(\alpha) \approx 5.79$
由于 $F = 60.49 > 5.79 = F_{2,5}(\alpha)$，故我们拒绝零假设 $H_0: \beta_1 = \beta_2 = 0$
即温度和浓度对吸收率有解释作用.

Part (b)

(b) Consider a two-way ANOVA model with interaction.

- ① Write down the two-way ANOVA model.
  State the assumptions.
  Find the parameter estimates.
- ② Obtain the ANOVA table.
  Is there a regression between the dependent variable and the 2 factors at the 5% level of significance?
- ③ Test whether or not the two factors interact; use $\alpha = 0.05$.
  State the alternatives, decision rule and conclusion.
- ④ Test whether or not the main effects for the two ingredients are present.
  Use $\alpha = 0.05$ in each case and state the alternatives, decision rule and conclusion.
  Is it meaningful here to test main factor effects? Explain.

Solution:
因子 A 为温度，因子水平 $A_1 = 10, A_2 = 55$
因子 B 为浓度，因子水平 $B_1 = 20, B_2 = 40$
模型 $y_{ijk} = \mu_{ij} + \varepsilon_{ijk}\ (i,j,k = 1,2)$
同方差假设、不相关假设和正态假设可以统一表示为:

$$\{\varepsilon_{ij}\} \overset{\text{i.i.d.}}{\sim} N(0,\sigma^2)$$

Temperature \ Concentration	20	40
10	0.28	0.38
10	0.278	0.392
55	0.266	0.332
55	0.258	0.334

参数的点估计量:

- $\hat{\mu}_{11} = \bar{y}_{11\cdot} = \frac{0.28+0.278}{2} = 0.279$
- $\hat{\mu}_{12} = \bar{y}_{12\cdot} = \frac{0.38+0.392}{2} = 0.386$
- $\hat{\mu}_{21} = \bar{y}_{21\cdot} = \frac{0.266+0,258}{2} = 0.262$
- $\hat{\mu}_{22} = \bar{y}_{22\cdot} = \frac{0.332+0.334}{2} = 0.331$
- $\bar{y}_{1\cdot\cdot} = \frac{0.279+0.386}{2} = 0.333$
- $\bar{y}_{2\cdot\cdot} = \frac{0.262+0.331}{2} = 0.297$
- $\bar{y}_{\cdot 1\cdot} = \frac{0.279+0.262}{2} = 0.271$
- $\bar{y}_{\cdot 2\cdot} = \frac{0.386+0.331}{2} = 0.359$
- $\bar{y}_{\cdots} = \frac{0.333+0.297}{2} = 0.315$
- $\widehat{\mathrm{ME}}_{(A_1,B_1)} = 0.279 - 0.315 = -0.036$
- $\widehat{\mathrm{ME}}_{(A_1,B_2)} = 0.386 - 0.315 = 0.071$
- $\widehat{\mathrm{ME}}_{(A_2,B_1)} = 0.262 - 0.315 = -0.053$
- $\widehat{\mathrm{ME}}_{(A_2,B_2)} = 0.331 - 0.315 = 0.016$
- $\widehat{\mathrm{ME}}_{A_1} = 0.333 - 0.315 = 0.018$
- $\widehat{\mathrm{ME}}_{A_2} = 0.297 - 0.315 = -0.018$
- $\widehat{\mathrm{ME}}_{B_1} = 0.271 - 0.315 = -0.044$
- $\widehat{\mathrm{ME}}_{B_2} = 0.359 - 0.315 = 0.044$
- $\widehat{\mathrm{IA}}_{(A_1,B_1)} = -0.036 - 0.018 - (-0.044) = -0.010$
- $\widehat{\mathrm{IA}}_{(A_1,B_2)} = 0.071 - 0.018 - 0.044 = 0.009$
- $\widehat{\mathrm{IA}}_{(A_2,B_1)} = -0.053 - (-0.018) - (-0.044) = -0.009$
- $\widehat{\mathrm{IA}}_{(A_2,B_2)} = 0.016 - (-0.018) - (0.044) = -0.010$

Residual Table:

Temperature \ Concentration	20	40
10	0.001	-0.006
10	-0.001	0.006
55	0.004	0.001
55	-0.004	-0.001

- $s^2 = \frac{1}{2\cdot 2\cdot(2-1)} \sum_{i,j,k}^{2,2,2} (y_{ijk} - \bar{y}_{ij\cdot})^2 = 2.7 \times 10^{-5}$

双因子方差分析的 ANOVA Table:

Sum of Squares	Degree of Freedom	Mean Squares	F-Value
$\text{SST}=\sum_{i,j,k}^{2,2,2}(y_{ijk}-\bar{y}_{\cdots})^2=0.019048$	$abn-1=7$	$\text{MST}=\frac{\text{SST}}{7}=0.00271$	
$\text{SSE}=\sum_{i,j,k}^{2,2,2}(y_{ijk}-\bar{y}_{ij\cdot})^2=0.000108$	$ab(n-1)=4$	$\text{MSE}=\frac{\text{SSE}}{4}=0.000027$	
$\text{SAB}=2\sum_{i,j}^{2,2}\widehat{\text{ME}}_{(A_i,B_j)}^2=0.018804$	$ab-1=3$	$\text{MAB}=\frac{\text{SAB}}{3}=0.006268$	$F^{(1)}=\frac{\text{MAB}}{\text{MSE}}=232.1$
$\text{SSAB}=2\sum_{i,j}^{2,2}\widehat{\text{IA}}_{(A_i,B_j)}^2=0.000724$	$(a-1)(b-1)=1$	$\text{MSAB}=\frac{\text{SSAB}}{1}=0.000724$	$F^{(2)}=\frac{\text{MSAB}}{\text{MSE}}=26.8$
$\text{SSA}=4\sum_{i=1}^{2}\widehat{\text{ME}}_{A_i}^2=0.002592$	$a-1=1$	$\text{MSA}=\frac{\text{SSA}}{1}=0.002592$	$F^{(3)}=\frac{\text{MSA}}{\text{MSE}}=24$
$\text{SSB}=4\sum_{j=1}^{2}\widehat{\text{ME}}_{B_j}^2=0.015488$	$b-1=1$	$\text{MSB}=\frac{\text{SSB}}{1}=0.015488$	$F^{(4)}=\frac{\text{MSB}}{\text{MSE}}=573.6$

设第一类型错误概率界限为 $\alpha=0.05$
查表得 $F_{3,4}(\alpha)=6.59$ 和 $F_{1,4}(\alpha)=7.71$

- 根据 $F^{(1)}=\frac{\text{MAB}}{\text{MSE}}=232.1>6.59=F_{3,4}(\alpha)$ 可知吸收率显著依赖于因子 "温度" 和 "浓度"
- 交互效应存在性检验的零假设和备择假设:

$$H_0:\text{IA}_{(A_i,B_j)}=\mu_{ij}+\mu_{\cdot\cdot}-\mu_{i\cdot}-\mu_{\cdot j}=0\quad(i,j=1,2)$$
$$\Updownarrow$$
$$H_1:\exists\, i,j=1,2\text{ such that }\text{IA}_{(A_i,B_j)}\neq 0$$

  根据 $F^{(2)}=\frac{\text{MSAB}}{\text{MSE}}=26.8>7.71=F_{1,4}(\alpha)$ 可以拒绝零假设，说明因子 "温度" 和 "浓度" 存在交互效应.

- 因子 A (温度) 的主效应检验的零假设和备择假设:

$$H_0:\text{ME}_{A_i}=0\quad(i=1,2)$$
$$\Updownarrow$$
$$H_1:\exists\, i=1,2\text{ such that }\text{ME}_{A_i}\neq 0$$

  根据 $F^{(3)}=\frac{\text{MSA}}{\text{MSE}}=24>7.71=F_{1,4}(\alpha)$ 可以拒绝零假设，说明因子 "温度" 存在主效应.

- 因子 B (浓度) 的主效应检验的零假设和备择假设:

$$H_0:\text{ME}_{B_j}=0\quad(j=1,2)$$
$$\Updownarrow$$
$$H_1:\exists\, j=1,2\text{ such that }\text{ME}_{B_j}\neq 0$$

  根据 $F^{(4)}=\frac{\text{MSB}}{\text{MSE}}=573.6>7.71=F_{1,4}(\alpha)$ 可以拒绝零假设，说明因子 "浓度" 存在主效应.

- **Is it meaningful to test main factor effects?**
  Given that the interaction effect is significant, testing the main effects becomes less meaningful in isolation because the interaction implies that the effects of temperature and concentration are not independent.
  It is more appropriate to interpret the interaction effect in the context of both factors simultaneously.

ANOVA TABLE (with interaction)

Sum of Squares		Degree of Freedom	Mean Squares
SST	$\sum_{i,j,k}^{a,b,n}(y_{ijk}-\bar{y}_{...})^2$	$abn-1$	$\mathrm{MST}=\frac{\mathrm{SST}}{abn-1}$
SSE	$\sum_{i,j,k}^{a,b,n}(y_{ijk}-\bar{y}_{ij\cdot})^2$	$ab(n-1)$	$\mathrm{MSE}=\frac{\mathrm{SSE}}{ab(n-1)}$
SAB	$n\sum_{i,j}^{a,b}\widehat{\mathrm{ME}}_{(A_i,B_j)}^2$	$ab-1$	$\mathrm{MAB}=\frac{\mathrm{SAB}}{ab-1}$
SSAB	$n\sum_{i,j}^{a,b}\widehat{\mathrm{IA}}_{(A_i,B_j)}^2$	$(a-1)(b-1)$	$\mathrm{MSAB}=\frac{\mathrm{SSAB}}{(a-1)(b-1)}$
SSA	$bn\sum_{i=1}^{a}\widehat{\mathrm{ME}}_{A_i}^2$	$a-1$	$\mathrm{MSA}=\frac{\mathrm{SSA}}{a-1}$
SSB	$an\sum_{j=1}^{b}\widehat{\mathrm{ME}}_{B_j}^2$	$b-1$	$\mathrm{MSB}=\frac{\mathrm{SSB}}{b-1}$

$$\widehat{\mathrm{ME}}_{A_i}=\bar{y}_{i\cdot\cdot}-\bar{y}_{...}$$

$$\widehat{\mathrm{ME}}_{B_j}=\bar{y}_{\cdot j\cdot}-\bar{y}_{...}$$

$$\widehat{\mathrm{ME}}_{(A_i,B_j)}=\bar{y}_{ij\cdot}-\bar{y}_{...}$$

$$\widehat{\mathrm{IA}}_{(A_i,B_j)}=\widehat{\mathrm{ME}}_{(A_i,B_j)}-\widehat{\mathrm{ME}}_{A_i}-\widehat{\mathrm{ME}}_{B_j}=\bar{y}_{ij\cdot}+\bar{y}_{...}-\bar{y}_{i\cdot\cdot}-\bar{y}_{\cdot j\cdot}$$

Problem 4

For the one-way ANOVA model:

$$y_{ij}=\mu_i+\varepsilon_{ij}\quad\begin{cases}i=1,\dots,a\\j=1,\dots,n_i\end{cases}\text{ where }\{\varepsilon_{ij}\}\overset{\text{i.i.d}}{\sim}N(0,\sigma^2)$$

For given $c_1,\dots,c_a$, propose a test statistic for:

$$H_0:\frac{\mu_1}{c_1}=\cdots=\frac{\mu_a}{c_a}$$

then find its null distribution and rejection region.

Solution:
考虑单因子方差分析的多元线性回归形式:

$$y=\begin{bmatrix}y^{(1)}\\y^{(2)}\\\vdots\\y^{(a)}\end{bmatrix}=\begin{bmatrix}1_{n_1}&&&\\&1_{n_2}&&\\&&\ddots&\\&&&1_{n_a}\end{bmatrix}_{n\times a}\begin{bmatrix}\mu_1\\\mu_2\\\vdots\\\mu_a\end{bmatrix}+\begin{bmatrix}\varepsilon^{(1)}\\\varepsilon^{(2)}\\\vdots\\\varepsilon^{(a)}\end{bmatrix}=X\mu+\varepsilon$$

其中:

$$y^{(i)}=\begin{bmatrix}y_{i,1}\\y_{i,2}\\\vdots\\y_{i,n_i}\end{bmatrix}\quad\varepsilon^{(i)}=\begin{bmatrix}\varepsilon_{i,1}\\\varepsilon_{i,2}\\\vdots\\\varepsilon_{i,n_i}\end{bmatrix}\quad(i=1,\dots,a)$$

$$n=n_1+n_2+\cdots+n_a$$

$$\{\varepsilon_{i,j}\}\overset{\text{i.i.d.}}{\sim}N(0,\sigma^2)$$

设计矩阵 $X=1_{n_1}\oplus\cdots\oplus 1_{n_a}\in\mathbb{R}^{n\times a}$
我们定义投影矩阵为:

$$
\begin{aligned}
H &:= X(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}\\
&=\begin{bmatrix}1_{n_1} & & & \\ & 1_{n_2} & & \\ & & \ddots & \\ & & & 1_{n_a}\end{bmatrix}\begin{bmatrix}n_1 & & & \\ & n_2 & & \\ & & \ddots & \\ & & & n_a\end{bmatrix}^{-1}\begin{bmatrix}1_{n_1} & & & \\ & 1_{n_2} & & \\ & & \ddots & \\ & & & 1_{n_a}\end{bmatrix}^{\mathrm{T}}\\
&=\begin{bmatrix}\frac{1}{n_1}1_{n_1}1_{n_1}^{\mathrm{T}} & & & \\ & \frac{1}{n_2}1_{n_2}1_{n_2}^{\mathrm{T}} & & \\ & & \ddots & \\ & & & \frac{1}{n_a}1_{n_a}1_{n_a}^{\mathrm{T}}\end{bmatrix}
\end{aligned}
$$

于是我们有:

$$
\begin{aligned}
\mathrm{SSE}_{\text{full}} &= \|y-\hat{y}\|^2\\
&=\|y-Hy\|^2\\
&=\left\|\begin{bmatrix}y^{(1)}\\ y^{(2)}\\ \vdots\\ y^{(a)}\end{bmatrix}-\begin{bmatrix}1_{n_1} & & & \\ & 1_{n_2} & & \\ & & \ddots & \\ & & & 1_{n_a}\end{bmatrix}_{n\times a}\begin{bmatrix}\bar{y}_{1\cdot}\\ \bar{y}_{2\cdot}\\ \vdots\\ \bar{y}_{a\cdot}\end{bmatrix}\right\|^2\\
&=\left\|\begin{bmatrix}y^{(1)}-\bar{y}_{1\cdot}1_{n_1}\\ y^{(2)}-\bar{y}_{2\cdot}1_{n_2}\\ \vdots\\ y^{(a)}-\bar{y}_{a\cdot}1_{n_a}\end{bmatrix}\right\|^2\\
&=\sum_{i=1}^{a}\|y^{(i)}-\bar{y}_{i\cdot}1_{n_i}\|^2\\
&=\sum_{i=1}^{a}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_{i\cdot})^2\sim\chi^2_{(n-a)}\quad(\text{where }\bar{y}_{i\cdot}=\frac{1}{n_i}\sum_{j=1}^{n_i}y_{ij})
\end{aligned}
$$

注意到零假设 $H_0:\frac{\mu_1}{c_1}=\cdots=\frac{\mu_a}{c_a}$ 可以等价表示为:

$$
H_0: C\mu=\begin{bmatrix}\frac{1}{c_1} & -\frac{1}{c_2} & & & \\ \frac{1}{c_1} & & -\frac{1}{c_3} & & \\ \vdots & & & \ddots & \\ \frac{1}{c_1} & & & & -\frac{1}{c_a}\end{bmatrix}\begin{bmatrix}\mu_1\\ \mu_2\\ \vdots\\ \mu_a\end{bmatrix}=\begin{bmatrix}0\\ 0\\ \vdots\\ 0\end{bmatrix}=0_{a-1}
$$

根据线性约束检验的结论可知:

$$\mathrm{ESS} = (C\hat{\mu}_{\text{full}})^{\mathrm{T}}[C(X^{\mathrm{T}}X)^{-1}C^{\mathrm{T}}]^{-1}(C\hat{\mu}_{\text{full}}) \sim \sigma^2\chi^2_{(a-1)}$$

$$\text{where } \hat{\mu}_{\text{full}} := \begin{bmatrix} \bar{y}_{1\cdot} \\ \bar{y}_{2\cdot} \\ \cdots \\ \bar{y}_{a\cdot} \end{bmatrix} \text{ and } C = \begin{bmatrix} \frac{1}{c_1} & -\frac{1}{c_2} & & & \\ \frac{1}{c_1} & & -\frac{1}{c_3} & & \\ \vdots & & & \ddots & \\ \frac{1}{c_1} & & & & -\frac{1}{c_a} \end{bmatrix} \in \mathbb{R}^{(a-1)\times a}$$

$$\mathrm{ESS} \perp \mathrm{SSE}_{\text{full}}$$

于是检验统计量为:

$$F := \frac{\mathrm{EES}/(a-1)}{\mathrm{SSE}_{\text{full}}/(n-a)} \overset{H_0}{\sim} \frac{\sigma^2\chi^2_{(a-1)}/(a-1)}{\sigma^2\chi^2_{(n-a)}/(n-a)} = F_{a-1,n-a}$$

当 $F > F_{a-1,n-a}(\alpha)$ 时，我们拒绝零假设 $H_0 : C\mu = 0$ (即 $H_0 : \frac{\mu_1}{c_1} = \cdots = \frac{\mu_a}{c_a}$)

The End