

作业一：方差分析

Q1

算错了.....

来源	平方和 SS	自由度 df	均方和 MS	F 值
因子 A	$SS_A = 7.5$	3	$MS_A = \frac{SS_A}{a-1} = 2.5$	$F_A = \frac{MS_A}{MS_E} = 2$
误差 E	$SS_E = 2.5$	2	$MS_E = \frac{SS_E}{n-a} = 1.25$	
总和	$SS_T = 10$	5		

Q2

(1) 已知因子数 $a = 2$, 重复次数为 m , 样本量 $n = am = 2m$.

第一组数据的均值 $\bar{x} = \frac{1}{m} \sum_{i=1}^m x_i$, 第二组数据的均值 $\bar{y} = \frac{1}{m} \sum_{i=1}^m y_i$

所有响应变量的均值

$$\bar{z} = \frac{\sum_{i=1}^m x_i + \sum_{i=1}^m y_i}{2m} = \frac{m\bar{x} + m\bar{y}}{2m} = \frac{\bar{x} + \bar{y}}{2}$$

假设 $H_0: \mu_1 = \mu_2$, $H_1: \mu_1 \neq \mu_2$,

$$SS_A = m [(\bar{x} - \bar{z})^2 + (\bar{y} - \bar{z})^2]$$

$$SS_E = \sum_{i=1}^m (x_i - \bar{x})^2 + \sum_{i=1}^m (y_i - \bar{y})^2 = \sum_{i=1}^m [(x_i - \bar{x})^2 + (y_i - \bar{y})^2]$$

检验统计量

$$\begin{aligned} F_A &= \frac{SS_A/(a-1)}{SS_E/(n-a)} = \frac{SS_A}{\frac{1}{2m-2} SS_E} \\ &= \frac{m [(\bar{x} - \bar{z})^2 + (\bar{y} - \bar{z})^2]}{\frac{1}{2m-2} \sum_{i=1}^m [(x_i - \bar{x})^2 + (y_i - \bar{y})^2]} \end{aligned}$$

若检验假设 H_0 成立,

$$F_A \sim F(a-1, n-a)$$

在显著性水平 α 下, 如果

$$F_A \geq F_{1-\alpha}(1, 2m-2)$$

则拒绝原假设, 反之则接受原假设

(2) 由上文知, 单因子方差分析模型的检验统计量

$$F_A = \frac{m [(\bar{x}_{\cdot} - \bar{z})^2 + (\bar{y}_{\cdot} - \bar{z})^2]}{\frac{1}{2m-2} \sum_{i=1}^m [(x_i - \bar{x}_{\cdot})^2 + (y_i - \bar{y}_{\cdot})^2]}$$

二样本独立 t 检验的检验统计量

$$\begin{aligned} t &= \frac{\bar{x}_1 - \bar{x}_2}{S_w \sqrt{\frac{1}{m_1} + \frac{1}{m_2}}} = \frac{\bar{x}_{\cdot} - \bar{y}_{\cdot}}{S_w \sqrt{\frac{2}{m}}} \\ S_w^2 &= \frac{m_1 - 1}{m_1 + m_2 - 2} S_1^2 + \frac{m_2 - 1}{m_1 + m_2 - 2} S_2^2 \\ &= \frac{m - 1}{2m - 2} \cdot \frac{1}{m - 1} \sum_{i=1}^m (x_i - \bar{x}_{\cdot})^2 + \frac{m - 1}{2m - 2} \cdot \frac{1}{m - 1} \sum_{i=1}^m (y_i - \bar{y}_{\cdot})^2 \\ &= \frac{1}{2m - 2} \sum_{i=1}^m [(x_i - \bar{x}_{\cdot})^2 + (y_i - \bar{y}_{\cdot})^2] \\ t &= \frac{\bar{x}_{\cdot} - \bar{y}_{\cdot}}{\sqrt{\frac{2}{m}} \sqrt{\frac{1}{2m-2} \sum_{i=1}^m [(x_i - \bar{x}_{\cdot})^2 + (y_i - \bar{y}_{\cdot})^2]}} \\ t^2 &= \frac{m \cdot \frac{1}{2} (\bar{x}_{\cdot} - \bar{y}_{\cdot})^2}{\frac{1}{2m-2} \sum_{i=1}^m [(x_i - \bar{x}_{\cdot})^2 + (y_i - \bar{y}_{\cdot})^2]} \\ (\bar{x}_{\cdot} - \bar{z})^2 + (\bar{y}_{\cdot} - \bar{z})^2 &= \bar{x}_{\cdot}^2 - 2\bar{x}_{\cdot}\bar{z} + \bar{z}^2 + \bar{y}_{\cdot}^2 - 2\bar{y}_{\cdot}\bar{z} + \bar{z}^2 \\ &= \bar{x}_{\cdot}^2 - 2\bar{x}_{\cdot} \cdot \frac{\bar{x}_{\cdot} + \bar{y}_{\cdot}}{2} + \left(\frac{\bar{x}_{\cdot} + \bar{y}_{\cdot}}{2}\right)^2 + \bar{y}_{\cdot}^2 - 2\bar{y}_{\cdot} \cdot \frac{\bar{x}_{\cdot} + \bar{y}_{\cdot}}{2} + \left(\frac{\bar{x}_{\cdot} + \bar{y}_{\cdot}}{2}\right)^2 \\ &= \frac{1}{2} \bar{x}_{\cdot}^2 - \bar{x}_{\cdot} \bar{y}_{\cdot} + \frac{1}{2} \bar{y}_{\cdot}^2 \\ &= \frac{1}{2} (\bar{x}_{\cdot} - \bar{y}_{\cdot})^2 \\ &\Rightarrow t^2 = F_A \end{aligned}$$

又由拒绝域法, 当 $|t| \geq t_{1-\alpha/2}(2m-2)$ 时拒绝原假设

由于 $t^2(2m-2) = F(1, 2m-2)$, 因此二者拒绝域相同 \Rightarrow 二者等价

Q3

(1) 因子 $a = 7$, 重复次数 $m = 4$, 样本量 $n = am = 28$

$$\text{所有响应变量的均值 } \bar{y}_{\cdot\cdot} = \frac{1}{n} \sum_{i=1}^a \sum_{j=1}^m y_{ij} = \frac{1}{a} \sum_{i=1}^a \bar{y}_{i\cdot} = \frac{233}{35}$$

$$SS_A = m \sum_{i=1}^a (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2 = \frac{488}{175}$$

$$\begin{aligned}
SS_E &= \sum_{i=1}^a \sum_{j=1}^m (y_{ij} - \bar{y}_{i.})^2 \\
&= \sum_{i=1}^a \sum_{j=1}^m (m-1) \cdot \frac{(y_{ij} - \bar{y}_{i.})^2}{m-1} \\
&= (m-1) \sum_{i=1}^a S_{y_i}^2 = 17.2554
\end{aligned}$$

检验统计量

$$F_A = \frac{SS_A/(a-1)}{SS_E/(n-a)} \approx 0.5656$$

查表得 $F_{0.95}(6, 21) \approx 2.5727$, 因此 $F_A < F_{0.95}(6, 21)$, 接受原假设, 即七种纤维强度间无显著差异。

(2) 由于七种纤维强度间无显著差异, 因此可以看作来自同一总体

总体标准差

$$S = \sqrt{\frac{1}{N-1} \sum_{i=1}^a \sum_{j=1}^m (x_{ij} - \bar{x}_{..})^2} = \sqrt{\frac{SS_T}{N-1}} = \sqrt{\frac{SS_A + SS_E}{N-1}} \approx 0.862$$

标准误

$$SE = \frac{S}{\sqrt{n}} = \frac{0.862}{\sqrt{28}} \approx 0.163$$

$$t_{1-\alpha/2}(n-1) = t_{0.975}(27) = 2.052$$

置信区间

$$[\bar{y}_{..} \pm t^* \cdot SE] = \left[\frac{233}{35} \pm (2.052 \times 0.163) \right] = [6.323, 6.992]$$

Q4

符号说明:

$y_{i.}$ 表示在第 i 个水平下响应变量的总和

$\bar{y}_{i.}$ 表示在第 i 个水平下响应变量的均值

$y_{..}$ 表示所有响应变量的总和

$\bar{y}_{..}$ 表示所有响应变量的均值

• 单因子方差分析模型

$$\begin{aligned}
y_{ij} : \mu + \alpha_i + \varepsilon_{ij}, \quad \varepsilon_{ij} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2), \quad & \begin{cases} i = 1, 2, \dots, a \\ j = 1, 2, \dots, m_i \end{cases} \\
\text{s. t.} \quad \sum_{i=1}^a \sum_{j=1}^{m_i} \alpha_i = \sum_{i=1}^a m_i \alpha_i = 0
\end{aligned}$$

- 原假设与备择假设

$$H_0: \alpha_1 = \alpha_2 = \cdots = \alpha_a = 0$$

$$H_1: \exists i \in \{1, 2, \cdots, a\}, \alpha_i \neq 0$$

- 检验统计量

- 平方和分解公式

$$\begin{aligned} SS_T &= \sum_{i=1}^a \sum_{j=1}^{m_i} (y_{ij} - \bar{y}_{..})^2 \\ &= \sum_{i=1}^a \sum_{j=1}^{m_i} ((\bar{y}_{i.} - \bar{y}_{..}) + (y_{ij} - \bar{y}_{i.}))^2 \\ &= \sum_{i=1}^a \sum_{j=1}^{m_i} [(\bar{y}_{i.} - \bar{y}_{..})^2 + 2(\bar{y}_{i.} - \bar{y}_{..})(y_{ij} - \bar{y}_{i.}) + (y_{ij} - \bar{y}_{i.})^2] \\ &= \sum_{i=1}^a m_i (\bar{y}_{i.} - \bar{y}_{..})^2 + \sum_{i=1}^a \sum_{j=1}^{m_i} (y_{ij} - \bar{y}_{i.})^2 + 2 \sum_{i=1}^a \sum_{j=1}^{m_i} (\bar{y}_{i.} - \bar{y}_{..})(y_{ij} - \bar{y}_{i.}) \end{aligned}$$

因为 $\sum_{j=1}^{m_i} (y_{ij} - \bar{y}_{i.}) = 0$, 所以

$$SS_T = \sum_{i=1}^a m_i (\bar{y}_{i.} - \bar{y}_{..})^2 + \sum_{i=1}^a \sum_{j=1}^{m_i} (y_{ij} - \bar{y}_{i.})^2$$

$$SS_A = \sum_{i=1}^a m_i (\bar{y}_{i.} - \bar{y}_{..})^2$$

$$SS_E = \sum_{i=1}^a \sum_{j=1}^{m_i} (y_{ij} - \bar{y}_{i.})^2$$

- 检验统计量

$$F_A = \frac{SS_A / (a - 1)}{SS_E / (n - a)} = \frac{\sum_{i=1}^a m_i (\bar{y}_{i.} - \bar{y}_{..})^2 / (a - 1)}{\sum_{i=1}^a \sum_{j=1}^{m_i} (y_{ij} - \bar{y}_{i.})^2 / (\sum_{i=1}^a m_i - a)}$$

- 方差分析表

来源	平方和 SS	自由度 df	均方和 MS	F 值
因子 A	$\sum_{i=1}^a m_i (\bar{y}_{i.} - \bar{y}_{..})^2$	$a - 1$	$\frac{\sum_{i=1}^a m_i (\bar{y}_{i.} - \bar{y}_{..})^2}{a - 1}$	$\frac{\sum_{i=1}^a m_i (\bar{y}_{i.} - \bar{y}_{..})^2 / (a - 1)}{\sum_{i=1}^a \sum_{j=1}^{m_i} (y_{ij} - \bar{y}_{i.})^2 / (\sum_{i=1}^a m_i - a)}$
误差 E	$\sum_{i=1}^a \sum_{j=1}^{m_i} (y_{ij} - \bar{y}_{i.})^2$	$\sum_{i=1}^a m_i - a$	$\frac{\sum_{i=1}^a \sum_{j=1}^{m_i} (y_{ij} - \bar{y}_{i.})^2}{\sum_{i=1}^a m_i - a}$	
总和	$\sum_{i=1}^a m_i (\bar{y}_{i.} - \bar{y}_{..})^2 + \sum_{i=1}^a \sum_{j=1}^{m_i} (y_{ij} - \bar{y}_{i.})^2$	$\sum_{i=1}^a m_i - 1$		