理论题 1

- 考虑一个因子有 4 种不同的水平，在各个水平下，我们进行了 6 次重复实验。已计算 $SS_T = 10$, $SS_E = 2.5$，请写出完整的 ANOVA 表。

$a=4$, $m=6$, $n=am=24$.

来源	平方和 SS	自由度 df	均方和 MS	F值
因子 A	$SS_A = 7.5$	3	$MS_A = \frac{SS_A}{a-1} = 2.5$	$F_A = \frac{MS_A}{MS_E} = 20$
误差 E	$SS_E = 2.5$	20	$MS_E = \frac{SS_E}{n-a} = 0.125$	
总和	$SS_T = 10$	23		

理论题 2

- 假设我们有两组独立的数据

$$\text{第一组}: x_1, x_2, \cdots, x_m,$$
$$\text{第二组}: y_1, y_2, \cdots, y_m.$$

假定 $x_i \overset{\text{i.i.d}}{\sim} N(\mu_1, \sigma^2)$ 且 $y_i \overset{\text{i.i.d}}{\sim} N(\mu_2, \sigma^2)$。其中，$\sigma^2$ 是未知常数。检验问题为

$$H_0 : \mu_1 = \mu_2 \quad \text{vs} \quad H_0 : \mu_1 \neq \mu_2.$$

- 用单因子方差分析模型来解决这个假设检验问题；
- 证明：在这种情况下，单因子方差分析模型与二样本独立 t 检验是等价的。（提示：考虑两个检验统计量分布之间的关系）

1) 单因子方差分析：

$\because x_i \overset{i.i.d}{\sim} N(\mu_1, \sigma^2)$

$\therefore \bar{x} \sim N(\mu_1, \frac{\sigma^2}{m})$. 同理 $\bar{y} \sim N(\mu_2, \frac{\sigma^2}{m})$.

$\therefore \bar{x} + \bar{y} \sim N(\mu_1 + \mu_2, \frac{2\sigma^2}{m})$

$\therefore \bar{y}_{\cdot} = \frac{\sum x_i + \sum y_i}{2m} \sim N(\frac{\mu_1 + \mu_2}{2}, \frac{\sigma^2}{2m})$.

$\therefore$ 组间偏差平方和 $SS_A = [(\bar{y}_{\cdot} - \bar{x})^2 + (\bar{y}_{\cdot} - \bar{y})^2] \cdot m$

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

$$= 2 \cdot (\frac{\bar{x}-\bar{y}}{2})^2 \cdot m$$

$$= \frac{m}{2}(\bar{x}-\bar{y})^2$$

组内偏差平方和 $SS_E = \sum_{i=1}^{m}(x_i - \bar{x})^2 + \sum_{i=1}^{m}(y_i - \bar{y})^2$

$$= (m-1)S_x^2 + (m-1)S_y^2$$

$$= (m-1)(S_x^2 + S_y^2).$$

$\therefore F = \frac{SS_A / 1}{SS_E / (2m-2)} = \frac{m(m-1)(\bar{x}-\bar{y})^2}{(m-1)(S_x^2+S_y^2)} = \frac{m(\bar{x}-\bar{y})^2}{S_x^2+S_y^2} \sim F(1, 2m-2)$.

拒绝域 $W = \{F \geq F_{1-\alpha}(1, 2m-2)\}$.

$\therefore$ 在显著性水平 α 下，若 $F \geq F_{1-\alpha}(1, 2m-2)$，则拒绝原假设 H_0，反之则接受原假设.

2) one-way ANOVA 中，检验统计量为：

$$F = \frac{SS_A/(a-1)}{SS_E/(n-a)} = \frac{m(\bar{x}-\bar{y})^2}{S_x^2+S_y^2} \sim F(1, 2m-2)$$

二样本独立t检验中，检验统计量为：

$$t = \frac{\bar{x}-\bar{y}}{\sqrt{\frac{1}{2}S_x^2+\frac{1}{2}S_y^2} \cdot \sqrt{\frac{2}{m}}} = \frac{\sqrt{m}(\bar{x}-\bar{y})}{\sqrt{S_x^2+S_y^2}} \sim t(2m-2)$$

$\because F = t^2$

$\therefore$ 在这种情况下 one-way ANOVA 与二样本独立t检验是等价的.

理论题 3

- 假设我们有数据如下

y_{11}	y_{12}	$\cdots$	y_{1m_1}
y_{21}	y_{22}	$\cdots$	y_{2m_2}
$\vdots$	$\vdots$		$\vdots$
y_{a1}	y_{a2}	$\cdots$	y_{am_a}

注意：这组数据中每组的重复次数是不相等的。

- 写出符合此数据的单因子方差分析模型；
- 写出原假设与备择假设；
- 写出检验统计量；
- 写出方差分析表；
- （选做）写出检验统计量分布的推导过程。

1) one-way ANOVA 模型：

$y_{ij} = \mu_i + \varepsilon_{ij}$, $\begin{cases} i = 1, 2, \ldots, a \\ j = 1, 2, \ldots, m_i \end{cases}$，其中 $\varepsilon_i \overset{i.i.d}{\sim} N(0, \sigma^2)$

2) $H_0: \mu_1 = \mu_2 = \cdots = \mu_\alpha$

v.s. $H_1: \exists\, i \neq j \in \{1, 2, \ldots, a\}$, s.t. $\mu_i \neq \mu_j$.

3) $F = \dfrac{SS_A/(a-1)}{SS_E/(n-a)}$，其中 $SS_A = \sum_{i=1}^{a} m_i (\bar{y}_i - \bar{y})^2$

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

4)

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