统计方法 理论作业1 10211900416 郭夏辉

1、解：根据题意 $a=4$ $m=6$ $n=am=24$

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

2、(1) 解：两组数据总均值 $\frac{m\bar{x}+m\bar{y}}{2m}=\frac{\bar{x}+\bar{y}}{2}$

检验统计量 $F=\dfrac{\frac{SS_A}{f_A}}{\frac{SS_E}{f_E}}=\dfrac{(2m-2)SS_A}{SS_E}$

其中 $SS_A=m\left[\bar{x}-\left(\frac{\bar{x}+\bar{y}}{2}\right)\right]^2+m\left[\bar{y}-\left(\frac{\bar{x}+\bar{y}}{2}\right)\right]^2=2m\left(\frac{\bar{x}-\bar{y}}{2}\right)^2$

$SS_E=\sum_{i=1}^{m}(x_i-\bar{x})^2+\sum_{i=1}^{m}(y_i-\bar{y})^2$ $\quad \frac{SS_E}{\delta^2}\sim\chi^2(n-a)=\chi^2(2m-2)$

在 H_0 成立时，$\frac{SS_A}{\delta^2}\sim\chi^2(a-1)=\chi^2(1)$ $\quad F\overset{H_0}{\sim}F(1,2m-2)$

拒绝域 $\{F\mid F>F_{1-\alpha}(1,2m-2)\}$ 若F落入拒绝域，则拒绝 H_0。

否则接受 H_0。

(2) 证明：在二样本独立 t 检验中，检验统计量

$t=\dfrac{\bar{x}-\bar{y}}{S_W\sqrt{\frac{1}{m}+\frac{1}{m}}}$ 其中 $S_W=\frac{m-1}{2m-2}S_x^2+\frac{m-1}{2m-2}S_y^2$

$=\dfrac{\sqrt{m}(\bar{x}-\bar{y})}{\sqrt{S_x^2+S_y^2}}\sim t(n-a)=t(2m-2)$ 拒绝域 $\{|t|>t_{1-\frac{\alpha}{2}}(2m-2)\}$

而在单因子方差分析中的 $F=\dfrac{(2m-2)SS_A}{SS_E}=\dfrac{(2m-2)\cdot 2m\cdot\frac{1}{4}(\bar{x}-\bar{y})^2}{(m-1)S_x^2+(m-1)S_y^2}$

$=\dfrac{m(\bar{x}-\bar{y})^2}{S_x^2+S_y^2}=t^2$

由此可得

$|t|>t_{1-\frac{\alpha}{2}}(2m-2)\Leftrightarrow F>F_{1-\alpha}(1,2m-2)$

拒绝域是一样的，故两者等价，原式得证。

3、(1). 解：单因子方差分析均值模型：

$$y_{ij} = \mu_i + \varepsilon_{ij} \quad 其中\ \varepsilon_i \overset{iid}{\sim} N(0, \delta^2)$$

μ_i 为第i个水平的总体均值. $\quad i=1,2\cdots a \quad j=1,2\cdots m_i$

(2). 解：原假设 $H_0: \mu_1 = \mu_2 = \cdots = \mu_a$

VS.

备择假设 H_1：存在 $i \neq j$ 且 $i,j \in [1,a]$ 使得 $\mu_i \neq \mu_j$

(3). 解：检验统计量

$$F = \frac{\frac{SSA}{f_A}}{\frac{SS_E}{f_E}} = \frac{\frac{SSA}{a-1}}{\frac{SS_E}{n-a}} = \frac{\sum_{i=1}^{a} m_i - a}{a-1} \cdot \frac{SS_A}{SS_E}$$

其中 $SSA = \sum_{i=1}^{a}\sum_{j=1}^{m_i} (\bar{y}_i - \bar{y})^2 = \sum_{i=1}^{a} m_i (\bar{y}_i - \bar{y})^2$

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

(4). 解：方差分析表

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

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

4.(1) 解：根据题意 $a=7 \quad m=4 \quad n=am=28$

$$y_{\cdot\cdot}=\sum_{i=1}^{a}\sum_{j=1}^{4}y_{ij}=4\times(6.3+6.2+6.7+6.8+6.5+7+7.1)=186.4$$

$$\bar{y}_{\cdot\cdot}=\frac{186.4}{28} \qquad SS_A=m\sum_{i=1}^{a}(\bar{y}_{i\cdot}-\bar{y}_{\cdot\cdot})^2\approx \cancel{4}\ 2.789$$

由于 $S_i^2=\frac{\sum_{j=1}^{m}(y_{ij}-\bar{y}_{i\cdot})^2}{m-1}$ 故 $SS_E=\sum_{i=1}^{a}\sum_{j=1}^{m}(y_{ij}-\bar{y}_{i\cdot})^2$

$$=(4-1)(0.81^2+0.92^2+1.22^2+0.74^2+0.88^2+0.58^2+1.05^2)$$

$$=17.2554$$

$$F=\frac{SS_A\cdot\frac{1}{a-1}}{SS_E\cdot\frac{1}{n-a}}=\frac{28-7}{7-1}\frac{SS_A}{SS_E}\approx 0.5657$$

查表 $F_{1-\alpha}(a-1,n-a)=F_{0.95}(6,21)=2.57$

$F<F_{1-\alpha}(a-1,n-a)$ 故接受原假设 $H_0: \mu_1=\mu_2=\cdots=\mu_7$

可认为7种纤维强度无显著差异

(2) 解：由(1)，各纤维之间无强度的明显差异，各种纤维可认为是n组均值相等的等方差正态分布 $N(\mu,\delta^2)$

δ^2 未知，要估计 μ 的 0.95 置信区间，，用 t 检验

$$\sqrt{n}\frac{\bar{y}_{\cdot\cdot}-\mu}{S_y}=\cancel{\sqrt{n}\frac{(\bar{y}_{\cdot\cdot}-\mu)}{n-1}}\sim t(n-1)=t(27)$$

然后 $S_y^2=\frac{\sum_{i=1}^{a}\sum_{j=1}^{m}(y_{ij}-\bar{y}_{\cdot\cdot})^2}{n-1}=\frac{SS_T}{n-1}=\frac{SS_A+SS_E}{n-1}=0.74239$

故 μ 的 0.95 置信区间为

$$\left[\bar{y}_{\cdot\cdot}-\frac{S_y}{\sqrt{n}}t_{0.975}(27),\ \bar{y}_{\cdot\cdot}+\frac{S_y}{\sqrt{n}}t_{0.975}(27)\right]$$

代入，可得 ~~$\mu\in[6.159,\ 7.155]$~~ $\mu\in[6.323,6.997]$

还有一种想法，就是我不能把这些数据看作来自相同的样本(即便参数一致)

$$\frac{\frac{\bar{y}_{\cdot\cdot}-\mu}{\sqrt{\frac{\delta^2}{n}}}}{\sqrt{\frac{SS_E}{\delta^2(n-a)}}}=\frac{\sqrt{n}(\bar{y}_{\cdot\cdot}-\mu)}{\sqrt{\frac{SS_E}{n-a}}}\sim t(n-a)=t(21)$$

这样 μ 的 0.95 置信区间为 $\left[\bar{y}_{\cdot\cdot}-t_{0.975}(n-a)\sqrt{\frac{SS_E}{n(n-a)}},\ \bar{y}_{\cdot\cdot}+t_{0.975}(n-a)\sqrt{\frac{SS_E}{n(n-a)}}\right]$

代入，可得 $\mu\in$ ~~[6.134, 7.181]~~ ~~[6.120, 7.194]~~ [6.302, 7.013]