Q_1:　$a=4$　$m=6$　$n=am=24$　$SS_T=10$　$SS_E=2.5$

$SS_A=SS_T-SS_E=10-2.5=7.5$

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

Q_2:

水平		总和	均值
1	$x_1 \cdots x_m$	X	$\bar{x}$
2	$y_1 \cdots y_m$	Y	$\bar{y}$
汇总		Z	$\bar{z}$

二样本独立t检验

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

检验统计量为 $t=\dfrac{\bar{x}-\bar{y}}{S_w\sqrt{\frac{1}{m}+\frac{1}{m}}}=\dfrac{\bar{x}-\bar{y}}{\sqrt{\cancel{\frac{m-1}{2m-2}}(S_x^2+S_y^2)}\cdot\sqrt{\frac{2}{m}}}=\dfrac{(\bar{x}-\bar{y})\sqrt{m}}{\sqrt{S_x^2+S_y^2}}$

在显著性水平α下

若　$|t|\geq t_{1-\alpha/2}(2m-2)$ 则拒绝原假设，反之接受

单因子方差分析模型

$a=2$　$n=am=2m$　$SS_A=m(\bar{x}-\bar{z})^2+m(\bar{y}-\bar{z})^2$

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

检验统计量为 $F_A=\dfrac{SS_A/1}{SS_E/2m-2}=\dfrac{[m(\bar{x}-\bar{z})^2+m(\bar{y}-\bar{z})^2](2m-2)}{\sum_{i=1}^{m}(x_i-\bar{x})^2+\sum_{i=1}^{m}(y_i-\bar{y})^2}$

在显著性水平α下

若　$F_A\geq F_{1-\alpha}(a-1,n-a)=F_{1-\alpha}(1,2m-2)$，则拒绝$H_0$，反之接受$H_0$

查看两个统计量

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

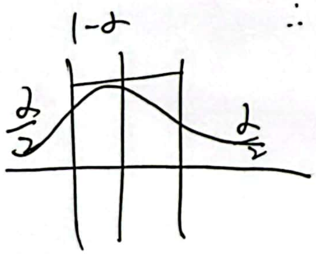
$$= \frac{[m(\frac{\bar{x}-\bar{y}}{2})^2 + m(\frac{\bar{y}-\bar{x}}{2})^2](2m-2)}{\sum_{i=1}^{m}(x_i-\bar{x})^2 + \sum_{i=1}^{m}(y_i-\bar{y})^2} = \frac{\cancel{2}[m(\frac{\bar{x}-\bar{y}}{\cancel{2}})^2](\cancel{2}m-\cancel{2})}{(m-1)S_x^2 + (m-1)S_y^2}$$

$$= \frac{m(\bar{x}-\bar{y})^2\cancel{(m-1)}}{\cancel{(m-1)}(S_x^2+S_y^2)} = \frac{m(\bar{x}-\bar{y})^2}{S_x^2+S_y^2}$$ 而 $t = \frac{(\bar{x}-\bar{y})\sqrt{m}}{\sqrt{S_x^2+S_y^2}}$ 可知 $F_A = t^2$

接下来需证明 $|t| \geq t_{1-\alpha/2}(2m-2)$ 与 $F_A \geq F_{1-\alpha}(1, 2m-2)$ 等价

assume $t \sim t(2m-2)$ $\quad F_A \sim F(1, 2m-2)$ $\quad F_A = t^2$

令 $P(F_A < x_0^2) = 1-\alpha \Rightarrow \therefore F_{1-\alpha}(1, 2m-2) = x_0^2$



$\therefore P(t^2 < x_0^2) = 1-\alpha = P(-x_0 < t < x_0)$

$\therefore P(t < x_0) = 1-\alpha+\frac{\alpha}{2} = 1-\frac{\alpha}{2}$

$\therefore t_{1-\frac{\alpha}{2}}(2m-2) = x_0$

$\therefore [t_{1-\frac{\alpha}{2}}(2m-2)]^2 = F_{1-\alpha}(1, 2m-2)$

因此 $|t| \geq t_{1-\frac{\alpha}{2}}(2m-2)$ 与 $F_A \geq F_{1-\alpha}(1, 2m-2)$ 等价

$\therefore$ 在这种情况下，one-way ANOVA 模型与二样本独立 t 检验等价

Q3

(1) $Y_{ij} = \mu_i + \varepsilon_{ij}$, $\begin{cases} i = 1, 2, \cdots a \\ j = 1, 2, \cdots m_i \end{cases}$

Y_{ij} 表示在因子的第i种水平下所观测到的第j个响应变量

μ_i 为第i个水平下的均值

ε_{ij} 为因子在第i种水平下的第j个响应变量的随机误差

ε_{ij} iid $N(0, \sigma^2)$

(2)

$H_0: \mu_1 = \mu_2 = \cdots = \mu_a$

H_1: 存在在两种水平下 $(i \neq j)$ 的均值不相等，即 $\mu_i \neq \mu_j \ (i \neq j)$

(3)

$$F_A = \frac{MS_A}{MS_B} = \frac{SS_A/(a-1)}{SS_E/(\sum_{i=1}^{a} m_i - a)}$$

$$SS_A = \sum_{i=1}^{a} m_i (\bar{Y}_i - \bar{Y})^2 \qquad SS_E = \sum_{i=1}^{a} \sum_{j=1}^{m_i} (Y_{ij} - \bar{Y}_i)^2$$

第1组： Y_{11} , Y_{12} , $\cdots$, Y_{1m_1} $\quad \bar{Y}_1$

⋮

第a组： Y_{a1} Y_{a2} , $\cdots$ Y_{am_a} $\quad \bar{Y}_a$

汇总 $\quad \bar{Y}$

(4)

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

Q4:

4

7

6.3	0.81
6.2	0.92
6.7	1.22
6.8	0.74
6.5	0.88
7.0	0.58
7.1	1.05
6.6571	

$a=7 \quad m=4 \quad n=am=28$

① $F_A = \dfrac{MS_A}{MS_B} = \dfrac{SS_A/(7-1)}{SS_E/(28-7)} = \dfrac{SS_A/6}{SS_E/21} = \dfrac{7SS_A}{2SS_E}$

$SS_A = 4\times[(6.3-6.6571)^2 + (6.2-6.6571)^2 + \cdots + (7.1-6.6571)^2]$

$= 2.8045$

$SS_E = \sum_{i=1}^{7}\sum_{j=1}^{4}(y_{ij}-\bar{y}_i)^2 = 3\times(0.81^2+0.92^2+\cdots 1.05^2) = 17.2554$

$\therefore F_A \approx 0.5688 <$ 查表 $F_{0.95}(6,21)$

$\therefore$ 未落在拒绝域内，接受 H_0，认为7种纤维强度无显著差异

② 因为不显著，所以由 7X4 变为 28X1 $\quad a=1 \quad m=28 \quad n=am=28$

$[\bar{y} - t_{1-\alpha/2}\hat{\sigma}/\sqrt{28},\ \bar{y} + t_{1-\alpha/2}\hat{\sigma}/\sqrt{28}]$

$\bar{y} = \dfrac{6.3+6.2+6.7+6.8+6.5+7.0+7.1}{7} = 6.6571 = \hat{\mu}$

$\hat{\sigma}^2 = \dfrac{SS_E}{n-a} = MS_E = \dfrac{SS_E}{28-1} = \dfrac{SS_E}{27}$

$\because$ 对于同一组数据，$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{m}(y_{ij}-\bar{y}_{..})^2$ 是不变的，而在本小问中，

$\bar{y}_{i.} = \bar{y}_{..}$，$\therefore SS_T = SS_A + SS_E =$ (第一问求过的) $17.2554 + 2.8045 = 20.0599$

$\hat{\sigma} = \sqrt{20.0599/27} = 0.862 \qquad t_{0.975}(27) = 2.0518$

$\therefore [6.6571 - 2.0518\times 0.862/\sqrt{28},\ 6.6571 + 2.0518\times 0.862/\sqrt{28}]$

$\therefore$ 平均强度的置信水平为 0.95 的置信区间为 $[6.3229, 6.9913]$