多元分析：第十周作业

蒋翌坤 20307100013

《实用多元统计分析》P267：6.23

$$\mathbf{B} = \sum_{l=1}^{g} n_l(\bar{\mathbf{x}}_l - \bar{\mathbf{x}})(\bar{\mathbf{x}}_l - \bar{\mathbf{x}})^T$$

$$\mathbf{W} = \sum_{l=1}^{g}\sum_{j=1}^{n_l}(\mathbf{x}_{lj} - \bar{\mathbf{x}}_l)(\mathbf{x}_{lj} - \bar{\mathbf{x}}_l)^T$$

于是，鸢尾花的宽度测量值单因子 MANOVA 表如下表所示：

来源	**平方和**	**自由度**
处理	$\mathbf{B} = \begin{bmatrix} 11.34 & -22.93 \\ -22.93 & 80.41 \end{bmatrix}$	2
残差	$\mathbf{W} = \begin{bmatrix} 16.96 & 4.81 \\ 4.81 & 6.16 \end{bmatrix}$	147
总和	$\mathbf{B}+\mathbf{W} = \begin{bmatrix} 28.31 & -18.12 \\ -18.12 & 86.57 \end{bmatrix}$	149

可以通过以下方式来构造两个响应的均值分量差的 95% 联合置信区间：

$$\tau_{ki} - \tau_{li} \in \bar{x}_{ki} - \bar{x}_{li} \pm t_{n-g}\Big(\frac{\alpha}{pg(g-1)}\Big)\sqrt{\frac{w_{ii}}{n-g}(\frac{1}{n_k} + \frac{1}{n_l})}$$

于是，可以得到响应的均值分量差的 95% 联合置信区间如下表所示：

差值	**下限**	**上限**
$\tau_{31} - \tau_{11}$	-0.64	-0.27
$\tau_{21} - \tau_{11}$	-0.84	-0.48
$\tau_{31} - \tau_{21}$	$2 \cdot 10^{-2}$	0.38
$\tau_{32} - \tau_{12}$	1.67	1.89
$\tau_{22} - \tau_{12}$	0.97	1.19
$\tau_{32} - \tau_{22}$	0.59	0.81

通过协方差矩阵相等性的检验可以判断假设 $\Sigma_1 = \Sigma_2 = \Sigma_3$ 的合理性。

$$\Lambda = \prod_l \Big(\frac{|\mathbf{S}_l|}{|\mathbf{S}_p|}\Big)^{(n_l-1)/2} = 6.71 \times 10^{-9}$$

$$C = (1-u) - 2\ln\Lambda = 37.01$$

其中，

$$u = \Big[\sum_l \frac{1}{n_l - 1} - \frac{1}{n-k}\Big]\frac{2p^2+3p-1}{6(p+1)(g-1)} = 0.017$$

$$v = \frac{1}{2}p(p+1)(g-1) = 6$$

由于 $C = 37.01 > \chi_6^2(1-0.05) = 12.59$,所以在 $\alpha = 0.05$ 的显著性水平下,拒绝假设 $\Sigma_1 = \Sigma_2 = \Sigma_3$。因此，假设 $\Sigma_1 = \Sigma_2 = \Sigma_3$ 不合理。

《实用多元统计分析》P273: 6.31

a

根据数据，可以构建双因子 MANOVA 表如下所示：

来源	平方和	自由度
因子 1	$\mathrm{SSP}_{\mathrm{factor1}} = \begin{bmatrix} 0.7 & -10.6 & 7.1 \\ -10.6 & 162.0 & -108.4 \\ 7.13 & -108.4 & 72.5 \end{bmatrix}$	1
因子 2	$\mathrm{SSP}_{\mathrm{factor2}} = \begin{bmatrix} 196.1 & 365.1 & 42.6 \\ 365.1 & 1089.0 & 414.6 \\ 42.6 & 414.6 & 284.1 \end{bmatrix}$	2
因子 1 * 因子 2	$\mathrm{SSP}_{\mathrm{interact}} = \begin{bmatrix} 205.1 & 363.6 & 107.7 \\ 363.6 & 780.6 & 254.2 \\ 107.7 & 254.2 & 85.9 \end{bmatrix}$	2
残差	$\mathrm{SSP}_{\mathrm{residual}} = \begin{bmatrix} 104.2 & 49.3 & 76.4 \\ 49.3 & 352.1 & 121.9 \\ 76.4 & 121.9 & 94.8 \end{bmatrix}$	30

X 效应对应的 Λ^* 和检验量如下所示：

$$\Lambda_X^* = \frac{|\mathrm{SSP}_{\mathrm{interact}}|}{|\mathrm{SSP}_X + \mathrm{SSP}_{\mathrm{residual}}|}$$

$$\text{检验量：} -\Big[gb(n-1) - \frac{p+1-\mathrm{df}(X)}{2}\Big]\ln\Lambda^*$$

其中，$\mathrm{df}(X)$ 表示 X 效应对应的自由度，可以得到关于因子 1（地块）效应，因子 2（品种）效应，地块–品种交互效应的 Λ^*、检验量及临界值如下所示：

效应	Λ_X^*	检验量	$\chi^2_{p\cdot \mathrm{df}(X)}(\alpha)$
因子 1	0.106	10.07	7.81
因子 2	0.012	21.93	12.59
因子 1 * 因子 2	0.074	12.99	12.59

因此，在 $\alpha = 0.05$ 的显著性水平下，地块效应，品种效应，地块–品种交互效应均显著。

b

从残差可以判断多元正态性，如果残差过大，说明总体可能不服从正态。可是，从题目数据来看，每个总体对应的样本量只有 2 个，即使残差很大（即两样本观测值相差很大）也不足以说明总体不服从正态。因此，可以假设总体服从正态，即 MANOVA 假定是满足的。

c

由（a）的结论可以知道地块效应，品种效应都不可加。三个独立的双因子一元 ANOVA 如下所示：

变量	来源	平方和	F value
x_1	因子 1 * 因子 2 残差	205 104	5.9
x_2	因子 1 * 因子 2 残差	780 352	6.6
x_3	因子 1 * 因子 2 残差	86 94	2.7

由于 $F_{(g-1)(b-1),gb(n-1)}(\alpha) = 3.13$ 比变量 x_1，x_2 的 F value 都小，因此，对于 x_1，x_2，交互效应均显著，对于 x_3，交互效应不显著。

d

通过 Bonferroni 联合置信区间，可以得到地块 2 中各品种的 95% 置信区间：

品种	x_1 置信区间	x_2 置信区间	x_3 置信区间
5	[7,362]	[-221,481]	[-161,261]
6	[102,297]	[81,246]	[-63,177]
8	[172,229]	[36,303]	[24,108]

从 Bonferroni 联合置信区间中可以看出，由于各个品种的置信区间中都有区间被其他品种的置信区间包含，所以没有一个品种的每个指标均优于其余两个品种。

附录

解答题目所使用的代码及输出请见：
https://thisiskunmeng.github.io/multivariate/hw10.html