林陳 Chap 13 ANOVA 習題

A.一般練習題

13.3 某知名飲料公司欲瞭解其所生產的飲料容量是否因機器的種類而不同，分別在 4 台機器獨立隨機抽取 5 罐飲料，如下表所示：

機器一	機器二	機器三	機器四
7	5	3	5
6	8	4	3
4	7	6	4
4	6	3	7
6	4	6	5

①請作出飲料容量是否因機器的種類而不同之 ANOVA 表。

②在 $\alpha = 0.05$ 下，4 種不同型機器所生產的飲料之平均容量是否相同？

③承②，檢定時所需之假設為何？

④母體變異數的點估計值為何？

解

①令 $\overline{Y}_1$ ，$\overline{Y}_2$ ，$\overline{Y}_3$ 和 $\overline{Y}_4$ 分別為樣本平均數，$\overline{\overline{Y}}$ 為總和樣本平均數。S_1^2 ，S_2^2 ，S_3^2 和 S_4^2 為樣本變異數。已知 $k=4$ ，樣本數為：$n_1 = n_2 = n_3 = n_4 = 5$ 。由題目可知，$\overline{Y}_1 = 5.4$ ，$\overline{Y}_2 = 6$ ，$\overline{Y}_3 = 4.4$ 和 $\overline{Y}_4 = 4.8$ ；$S_1^2 = 1.8$ ，$S_2^2 = 2.5$ ，$S_3^2 = 2.3$ 和 $S_4^2 = 2.2$ ；$\overline{\overline{Y}} = 5.15$

$$SSF = \sum_{i=1}^{k} n_i(\overline{Y}_i - \overline{\overline{Y}})^2 = 5(5.4-5.15)^2 + 5(6-5.15)^2 + 5(4.4-5.15)^2 + 5(4.8-5.15)^2 = 7.35$$

$$MSF = \frac{SSF}{k-1} = \frac{7.35}{4-1} = 2.45$$

$$SSE = \sum_{i=1}^{k} (n_i - 1)S_i^2 = (5-1)(1.8) + (5-1)(2.5) + (5-1)(2.3) + (5-1)(2.2) = 35.2$$

$$MSE = \frac{SSE}{n-k} = \frac{35.2}{20-4} = 2.2 \text{ ， } F = \frac{MSF}{MSE} = \frac{2.45}{2.2} = 1.11$$

故ANOVA表如下

變異來源	平方和 (*SS*)	自由度 (*df*)	平均平方和 (*MS*)	*F*值
因子變異	7.35	3	2.45	1.11
隨機變異	35.2	16	2.2	
總和	42.55	19		

②設 μ_1、μ_2、μ_3 和 μ_4 分別表示四種不同型機器所生產的飲料之平均容量，故可設立兩個假設：$H_0:\mu_1=\mu_2=\mu_3=\mu_4$，$H_1:\mu_i$不全等。檢定統計量由ANOVA表可知，$F=1.11$，小於臨界值 $F_{3,16,0.05}=3.24$，落在接受域，故不拒絕虛無假設。結論為：「四種不同型機器所生產的飲料之平均容量相同」。

③檢定時所需之假設為：假設四種不同型機器所生產的飲料容量具有相同變異數的常態分配，且分別獨立隨機抽取樣本。

④母體變異數的點估計值為 $MSE=2.2$。

13.4 下表為一隨機集區設計之實驗的樣本資料：

集間	處理		
	甲	乙	丙
A	3	5	2
B	8	7	6
C	7	6	5

①請完成此 ANOVA 表：

變源	SS	自由度	MS	F
集區	21.56	2		
處理				
隨機				
總和	30.23	8		

②若欲檢定各處理間之平均數是否相同，請寫下虛無假設。

③承②，檢定統計量為何？

④承③，檢定結果為何？（假設 $\alpha=0.05$）

解

① $SSTR=b\sum_{j=1}^{k}(\overline{Y}_{.j}-\overline{\overline{Y}})^2=3[(6-5.44)^2+(6-5.44)^2+(4.33-5.44)^2]=5.56$

$SSE=SST-SSTR-SSBK=30.23-5.56-21.56=3.11$ ，

$$MSTR=\frac{SSTR}{k-1}=\frac{5.56}{3-1}=2.78$$

$$MSBK=\frac{SSBK}{b-1}=\frac{21.56}{3-1}=10.78$$ ，

$$MSE=\frac{SSE}{(k-1)(b-1)}=\frac{3.11}{(3-1)(3-1)}=0.7775$$

$F_{處理}=\frac{MSTR}{MSE}=\frac{2.78}{0.7775}=3.58$，$F_{集區}=\frac{MSBK}{MSE}=\frac{10.78}{0.7775}=13.86$

變源	SS	自由度	MS	F
集區	21.56	2	10.78	13.86
處理	5.56	2	2.78	3.58
隨機	3.11	4	0.7775	
總和	30.23	8		

②虛無假設為 $H_0:\mu_{甲}=\mu_{乙}=\mu_{丙}$ 。

③檢定統計量為3.58。

④因檢定統計量 $F_{處理}=3.58$ 小於臨界值 $F_{2,4,0.05}=6.94$，落在接受域，故不拒絕虛無假設。結論為：「各處理間之平均數相同」。

13.5 某甲發現其所任職公司的人員，平常都用四種中文輸入法（注音、倉頡、大易、速成）之一來做文書處理，他想知道這 4 種輸入法的輸入速度是否相同。經隨機抽選部份人員，調查其每分鐘打字的字數後，某甲得到下列變異數分析表：

變異來源	自由度	平方和	平均平方和	F
輸入法			104	
隨機變異				
總變異	43	1908		

①請完成上述之變異數分析表。

②試檢定這 4 種輸入法的輸入速度是否相同（$\alpha=5\%$）？

解

①

變異來源	自由度	平方和	平均平方和	F
輸入法	3	312	104	2.61
隨機	40	1,596	39.9	
總合	43	1,908		

② $F=2.61<F_{4\text{-}1,40,0.05}=2.84$，四種輸入法的輸入速度相同。

③共同變異數的信賴區間 $=(\frac{1596}{\chi^2_{40,0.025}},\frac{1596}{\chi^2_{40,0.975}})=(26.9,65.3)$

查表七 P.701 得 $\chi^2_{40,0.025}=59.3417$，$\chi^2_{40,0.975}=24.4331$ 。

13.6 設統一集團為觀察 7-11 便利超商在大都會，城市及鄉下之銷售額有無差

異。則自 3 個區域隨機抽出幾家超商，並記錄如下：

大都市	10	13	14	16	17	
城市	5	8	10	11	12	14
鄉鎮	6	8	9	13		

①請建立 ANOVA 表。

②檢定區域別是否影響銷售額（$\alpha = 0.05$）。

③試做城市銷售額之 95%信賴區間。

解

①
$$\begin{cases} H_0 : \mu_1 = \mu_2 = \mu_3 \\ H_1 : \mu_i \text{不全等} \end{cases}$$

$$\bar{X}_1 = \frac{70}{5} = 14, \bar{X}_2 = \frac{60}{6} = 10, \bar{X}_3 = \frac{36}{4} = 9$$

$$S_1^2 = \frac{30}{4} = 7.5, S_2^2 = \frac{50}{5} = 10, S_3^2 = \frac{26}{3}, \bar{X} = 11.0667$$

$$\therefore SSF = (14 - 11.0667)^2 \times 5 + (10 - 11.0667)^2 \times 6 + (9 - 11.0667)^2 \times 4 = 66.9333$$

$$SSE = 7.5 \times 4 + 10 \times 5 + 3 \times \frac{26}{3} = 106$$

ANOVA 表

變異來源	SS	df	MS	F
SSF	66.9333	2	33.467	3.79
SSE	106	12	8.833	
SST	172.9343	14		

②因為 $F = 3.79 < F_{2,12,0.05} = 3.89$

因此接受 H_0，三者相等。

③ $10 - t_{12,0.025}\sqrt{8.833}\sqrt{1/6} \le \mu_2 \le 10 + t_{12,0.025}\sqrt{8.833}\sqrt{1/6}$

$7.354 \le \mu_2 \le 12.646$

13.7 自 4 次大型考試的成績中各抽出 50 個成績得（同一試題，因此假定 $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2$）：

$\bar{X}_1 = 68, \bar{X}_2 = 74, \bar{X}_3 = 70, \bar{X}_4 = 68 \quad S_1 = 11, S_2 = 12, S_3 = 8, S_4 = 10$

試檢定 $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$（$\alpha = 0.05$）。

解

設檢定假設為：

$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$

$H_1 : H_0$不真(μ_i不全等)

$$\overline{X} = \frac{68+74+70+68}{4} = \frac{280}{4} = 70$$

$$MSE = \frac{SSE}{49\times 4} = \frac{49\times(11^2+12^2+8^2+10^2)}{50\times 4 - 4} = 107.25$$

又 $SSF = 50(68-70)^2 + 50(74-70)^2 + 50(70-70)^2 + 50(68-70)^2 = 1200$

$$MSF = \frac{1200}{4-1} = 400$$

ANOVA 表

變異來源	SS	df	MS	F
SSF	1200	3	400	3.7296
SSE	21021	196	107.25	
SST	22221	199		

$$F = \frac{MSF}{MSE} = \frac{400}{107.25} = 3.73$$

查 F 臨界值表，F 的自由度：$V_1 = 4-1$，$V_2 = 4\times 50 - 4 = 196$ 。查表得臨界值：$F_{3,196,0.05} \approx 2.60$ ，檢定統計量 F 值=3.73 大於 2.60，落於拒絕域，因此拒絕 H_0 ，即四次考試成績不相等。

B. 應用題

13.13 台北、台中和高雄地區的房屋價格是否有差異？現由 3 個地區各選 6 間成交成屋，計算其每坪的平均售價，結果如下表：

樣本觀察值	台北地區	台中地區	高雄地區
1	31.5	4.3	29.7
2	20.3	10.9	17.1
3	29.4	16.1	14.9
4	19.1	11.3	10.2
5	23.8	8.5	6.5
6	21.6	18.9	18.2
樣本平均數	24.28	11.67	16.10
樣本變異數	25.7	27.4	64.0

（資料來源：參考信義房屋網站資料虛擬。）

①請建立 ANOVA 表。

②檢定地區別是否影響房屋每坪的價格。（$\alpha = 0.05$）

③試求台北和台中房屋每坪價格差異的 95%單一信賴區間。

解

令 μ_1，μ_2，μ_3 分別為三個地區房屋每坪的平均售價(萬元/坪)。令 $\overline{Y}_1$，$\overline{Y}_2$，$\overline{Y}_3$ 分別為相對應的樣本平均數，$\overline{\overline{Y}}$ 為總和樣本平均數。S_1^2，S_2^2，S_3^2 為相對應的樣本變異數。已知 $k=3$，樣本數為 $n_1=n_2=n_3=6$。因為 3 組樣本的樣本數都是 6，故總樣本平均數為：

$$\overline{\overline{Y}}=\frac{\sum_{i=1}^{k}\sum_{j=1}^{n_i}Y_{ij}}{nk}=\frac{\sum_{i=1}^{k}\sum_{j=1}^{n_i}Y_{ij}/n}{k}=\frac{\sum_{i=1}^{k}\overline{Y}_i}{k}=\frac{24.28+11.67+16.10}{3}=17.35$$

①先算 *MSF* 與 *MSE*。根據(13.5)式可得 *SSF* 為：

$$SSF=\sum_{i=1}^{k}n_i(\overline{Y}_i-\overline{\overline{Y}})^2$$

$$=6(24.28-17.35)^2+6(11.67-17.35)^2+6(16.10-17.35)^2=491.1$$

因此根據公式可得：

$$MSF=\frac{SSF}{k-1}=\frac{491.1}{3-1}=245.55$$

依公式可得 *SSE* 為：

$$SSE=\sum_{i=1}^{k}(n_i-1)S_i^2=(6-1)(25.7)+(6-1)(27.4)+(6-1)(64.0)=585.5$$

因此根據公式可得：

$$MSE=\frac{SSE}{\sum n_i-k}=\frac{585.5}{15}=39.03$$

於是可得檢定統計量 *F* 值(公式)為：

$$F=\frac{MSF}{MSE}=\frac{245.55}{39.03}=6.29$$

綜合上面的結果可得 ANOVA 表如下：

變異來源	平方和(*SS*)	自由度(*df*)	平均平方和(*MS*)	*F* 值
因子	491.6033	2	245.8017	6.31122
隨機	584.2	15	38.94678	
總和	1075.805	17		

②設立兩個假設

$H_0:\mu_1=\mu_2=\mu_3$(地區別對房屋每坪的平均售價有影響)

$H_1:\mu_i$不全等(地區別對房屋每坪的平均售價沒有影響)

因為要比較三個常態母體的平均數，所以用 *F* 檢定量來做檢定。

顯著水準為 $\alpha=0.05$。ANOVA 檢定為右尾檢定，拒絕域在 *F* 分配的右尾。

F 分配的兩個自由度為：

分子的自由度為 $df=k-1=3-1=2$

分母的自由度為 $df = \sum n_i - k = 18 - 3 = 15$

查 F 分配機率表知，F 檢定的臨界值為 $F_{2,15,0.05} = 3.68$ 。

因檢定統計量 $F = 6.29$ 大於臨界值 $F_{2,15,0.05} = 3.68$ ，落在拒絕域，故拒絕虛無假設。結論為：「地區別影響房屋每坪的價格」。

③依公式可求得 $\mu_1 - \mu_2$ 的信賴區間如下：

查 t 值表 $t_{15,0.025} = 2.131$ 。 $MSE = 39.03$ ，$\overline{Y}_1 = 24.28$ ，$\overline{Y}_2 = 11.67$ ，於是可得：

$$(24.28 - 11.67) - 2.131\sqrt{39.03}\sqrt{\frac{2}{6}} \le \mu_1 - \mu_2 \le (24.28 - 11.67) + 2.131\sqrt{39.03}\sqrt{\frac{2}{6}}$$

$$12.61 - 7.69 \le \mu_1 - \mu_2 \le 12.61 + 7.69$$

$$4.92 \le \mu_1 - \mu_2 \le 20.3$$

在 95%信賴水準下，台北地區房屋每坪價格約比台中地區高出 4.92 萬元至 20.3 萬元。

13.14 台灣於二次大戰後經濟迅速成長，生育率降低，當時仍保有父系家庭組織觀念，產生重男輕女現象，遂產生姊妹為貼補其兄弟受教育而工作的情形。隨機抽取 1946 年至 1960 年出生之男、女性各 60 人，詢問其教育年限，得資料如下：（單位：年）

	平均數	標準差
男性	9.75	0.78
女性	7.83	1.26

（資料來源：樣本數、平均數與標準差為參考右列資料模擬。M. Weinstein, J. Cornman, and M.-C. Chang, R. Hassan, M. Stark, 2004. “Did Taiwanese Sisters Subsidize the Education of their Brothers,” Journal of Population Studies 29:1-34.）

假設兩母體變異數相等，試以下列兩種方法檢定男、女性接受教育的平均年限是否相同（ $\alpha = 0.05$ ）。

①兩母體平均數差的檢定。

②變異數分析法。

③①與②的結果是否相同？

解

①設立兩個假設

$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \ne \mu_2$$

樣本數 n=60，樣本數足夠大，$\overline{X}_1-\overline{X}_2$ 的抽樣分配趨近於常態分配，因此以 Z 分配來做漸進假設檢定。

母體變異數未知但知相等，依公式計算混合標準差

$$S_p=\sqrt{\frac{(60-1)(0.78)^2+(60-1)(1.26)^2}{60+60-2}}=1.05$$

依公式計算 Z 值：

$$Z=\frac{(9.75-7.83)-0}{1.05\sqrt{\frac{1}{60}+\frac{1}{60}}}=10.02$$

因為 $Z=10.02>Z_{0.025}=1.96$，落在拒絕域，故拒絕虛無假設，結論為男、女性接受教育的平均年限不同。

②根據公式可得下列變異數分析表：

變異來源	平方和	自由度	平均平方和	F 值
因子	110.592	1	110.59	100.72
隨機	129.564	118	1.10	
總合	240.156	119		

兩個假設與①同，因檢定統計量 $F=100.72>F_{1,118,0.05}\approx 3.92$，故拒絕 H_0，結論為男、女性接受教育的平均年限不同。

③①與②的結果相同。