# 《統計學概要》

一、某大學統計學系之大一統計學課程，依學生高中時期之數學程度分成甲、乙兩班，甲班學生40人，其統計學期末考試之平均成績為35分，標準差為5分；乙班學生60人，其平均成績為80分，標準差為4分。

(一)試問兩班全體學生統計學期末考試之平均成績為多少分？(6分)

(二)甲班學生之考試成績頗不理想，故老師決定每位學生之成績均「乘以2後再減5分」，試問經調整分數後，該班學生成績之平均數及標準差分別為何？(8分)

(三)乙班學生中有1人夾帶小抄舞弊，其成績為60分，經開會決議應以0分計算，試問乙班統計學期末考試成績之平均數及標準差應修正為多少分？(6分)

| | |
|---|---|
| **試題評析** | 本題涉及部份資料改變對平均數與變異數之影響，考古題中也有相關題目，且課本中也有收錄，考生只要小心計算，獲得高分應該不難。 |
| **考點命中** | 《高點‧高上統計學講義》第一回，趙治勳編撰，頁64例19。 |

**答**：

(一) $\mu_{全體}=\dfrac{N_1\mu_1+N_2\mu_2}{N_1+N_2}=\dfrac{40(35)+60(80)}{40+60}=62$ (分)

(二) 令 $X_i$ 表第i位學生調整分數前之得分且平均數為 $\mu_1$ 與標準差為 $\sigma_1$

$Y_i$ 表第i位學生調整分數後之得分且平均數為 $\mu_1^*$ 與標準差為 $\sigma_1^*$

$Y_i=2X_i-5$

$$\mu_1^*=\frac{\sum Y_i}{40}=\frac{\sum(2X_i-5)}{40}=2\mu_1-5=2(35)-5=65\text{(分)}$$

$$\sigma_1^*=\sqrt{\frac{\sum(Y_i-\mu_1^*)^2}{40}}=\sqrt{\frac{\sum(2X_i-5-(2\mu_1-5))^2}{40}}$$

$$=2\sqrt{\frac{\sum(X_i-\mu_1)^2}{40}}=2(5)=10\ \text{(分)}$$

<另解>令 $X$ 表甲班學生調整分數前之得分，$X\sim(35,5^2)$

$Y$ 表甲班學生調整分數後之得分

$Y=2X-5$

$E(Y)=E(2X-5)=2E(X)-5=2(35)-5=65$ (分)

$\sqrt{V(Y)}=\sqrt{V(2X-5)}=\sqrt{4V(X)}=2\sqrt{V(X)}=2(5)=10$ (分)

(三) 令 $X_i$ 表乙班學生修正分數前之得分且平均數為 $\mu_2$ 與標準差為 $\sigma_2$

$X_{60}$ 表該位作弊學生修正分數前得分，$X_{60}=60$

$Y_i$ 表乙班學生修正分數後之得分且平均數為 $\mu_2^*$ 與標準差為 $\sigma_2^*$

$Y_i=X_i, i=1,2,\cdots,59$

$Y_{60}$ 表該位作弊學生修正分數後之得分，$Y_{60}=0$

$$\mu_2=\frac{\sum_{i=1}^{60}X_i}{60}=\frac{\sum_{i=1}^{59}X_i+X_{60}}{60}=\frac{\sum_{i=1}^{59}X_i+60}{60}=80$$

$$\Rightarrow\sum_{i=1}^{59}X_i=80(60)-60=4740$$

$$\sigma_2^2=\frac{\sum_{i=1}^{60}X_i^2-N_2\mu_2^2}{60}=\frac{\sum_{i=1}^{59}X_i^2+X_{60}^2-60(80)^2}{60}=\frac{\sum_{i=1}^{59}X_i^2+(60)^2-60(80)^2}{60}=4^2$$

$$\Rightarrow\sum_{i=1}^{59}X_i^2=4^2(60)+60(80)^2-(60)^2=3181360$$

$$\mu_2^*=\frac{\sum_{i=1}^{60}Y_i}{60}=\frac{\sum_{i=1}^{59}Y_i+Y_{60}}{60}=\frac{\sum_{i=1}^{59}X_i+0}{60}=\frac{4740}{60}=79\text{ (分)}$$

$$\sigma_2^*=\sqrt{\frac{\sum_{i=1}^{60}Y_i^2-N_2\mu_2^{2*}}{60}}=\sqrt{\frac{\sum_{i=1}^{59}Y_i^2+Y_{60}-N_2\mu_2^{2*}}{60}}=\sqrt{\frac{\sum_{i=1}^{59}X_i^2+0-60(79)^2}{60}}$$

$$=\sqrt{\frac{3181360+0-60(79)^2}{60}}=\sqrt{115}=10.7238\text{ (分)}$$

二、5位某校統計系大一學生之微積分成績(X)及統計學成績(Y)如下表所示：

| | | | | | |
|---|---|---|---|---|---|
| x | 30 | 68 | 24 | 31 | 49 |
| y | 63 | 80 | 58 | 64 | 71 |

今已求得：$\Sigma x=202$，$\Sigma y=336$，$\Sigma x^2=9,462$，$\Sigma y^2=22,870$，$\Sigma xy=14,185$。

(一)試求微積分成績及統計學成績之相關係數。(6分)

(二)今擬以微積分成績推估統計學成績，試據相關資料配一迴歸直線。(8分)

(三)試求微積分成績及統計學成績之Spearman等級相關係數(Spearman's rankcorrelation coefficient)。(6分)

| | |
|---|---|
| 試題評析 | 本題(一)(二)兩小題有關簡迴歸之計算題。考古題中常有，拿滿分並無困難。但(三)的等級相關係數實屬冷門。雖然講義中有相同的範例，但考生可能沒有特別注意到。 |
| 考點命中 | (一)(二)《高點・高上統計學講義題型加強課試題》第六回，趙治勳編撰，頁20例2。<br>(三)《高點・高上統計學講義題型加強課試題》第六回，趙治勳編撰，頁96。 |



**答**：

可得 $SS_X=9462-\frac{202^2}{5}=1301.2, SS_Y=22870-\frac{336^2}{5}=290.8$

$$SS_{XY}=14185-\frac{(202)(336)}{5}=610.6$$

(一) $r=\frac{SS_{XY}}{\sqrt{SS_X}\sqrt{SS_Y}}=\frac{610.6}{\sqrt{1301.2}\sqrt{290.8}}=0.9926$

(二) $\hat{\beta}_1=\frac{SS_{XY}}{SS}=\frac{610.6}{1301.2}=0.4693$, $\hat{\beta}_0=\overline{Y}-\hat{\beta}_0\overline{X}=\frac{336}{5}-0.4693\times\frac{202}{5}=48.2403$

$\therefore \hat{y}=48.2403+0.4693x$

(三)先計算X,Y之等級(rank)

| | | | | | |
|---|---|---|---|---|---|
| X | 2 | 5 | 1 | 3 | 4 |
| Y | 2 | 5 | 1 | 3 | 4 |
| $d=X-Y$ | 0 | 0 | 0 | 0 | 0 |

再計算Spearman等級相關係數

$$r_s=1-\frac{6\sum d_i^2}{n(n^2-1)}=1-\frac{6(0)}{5(5^2-1)}=1$$

三、設有兩袋，第Ⅰ袋中裝有3紅球(R)，2白球(W)，1藍球(B)；第Ⅱ袋中裝有4白球，2藍球。今擬自袋子中以不放回方式摸取2球。試問：

(一)自第Ⅰ袋中摸出2白球之機率為何？(5分)

(二)自第Ⅱ袋中摸出2白球之機率為何？(5分)

(三)若擲一正常骰子，出現1點或6點，則自第Ⅰ袋中以不放回方式摸出2球，否則自第Ⅱ袋中以相同方式摸出2球。今若所摸出之2球皆為白球，試問它係摸自第Ⅰ袋之機率為何？(10分)

| | |
|---|---|
| **試題評析** | 本題涉及超幾何分配與貝氏定理。 |
| **考點命中** | 1.《高點・高上統計學講義》第三回，趙治勳編撰，頁15。<br>2.《高點・高上統計學講義》第一回，趙治勳編撰，頁89。 |

**答**：

(一) 令 $X$ 表從I袋中抽出白球之個數， $X\sim Hyper(N=6,m=2,n=2)$

$$P(X=2)=\frac{\binom{2}{2}\binom{4}{0}}{\binom{6}{2}}=0.06667$$

(二) 令$Y$表從II袋中抽出白球之個數，$Y\sim Hyper(N=6,m=4,n=2)$

$$P(Y=2)=\frac{\binom{4}{2}\binom{2}{0}}{\binom{6}{2}}=0.4$$



(三) 令$A$表骰子投出1點或6點(從I袋抽兩球)

$A^C$ 表骰子投出2點或3點或4點或5點(從II袋抽兩球)

$B$ 表抽出兩個白球

$$P(A|B)=\frac{P(A)P(B|A)}{P(A)P(B|A)+P(A^C)P(B|A^C)}$$

$$=\frac{\frac{2}{6}\times 0.06667}{\frac{2}{6}\times 0.06667+\frac{4}{6}\times 0.4}=0.07693$$

四、三種品種之馬鈴薯各以面積相等之田地若干塊作種植實驗，一段時間後得收穫之產量（單位：公噸）如下表所示：

| 品種 | $X_{if}$ | | | |
|---|---|---|---|---|
| A | 23 | 20 | 24 | 25 |
| B | 23 | 20 | 17 | 26 |
| C | 20 | 17 | 16 | 21 |

(一)若欲對三種品種生產能力是否將相同，而進行變異數分析，則對品種之收穫量有何基本之假設？（6分）

(二)求進行變異數分析時之總平方和（Total sum of squares）SST及處理平方和（Treatment sum of squares）SSTR。（8分）

(三)試在α=0.05的顯著水準下，檢定三種品種生產能力相同之假設是否成立？（6分）

| 試題評析 | 本題涉及一因子變異數分析之基本假設與計算，講義中很多相同例題，獲得滿分不難。 |
|---|---|
| 考點命中 | 《高點・高上統計學講義》第五回，趙治勳編撰，頁70。 |

**答**：

(一) 假設模型: $X_{ij}=\mu+\alpha_i+\varepsilon_{ij},\varepsilon_{ij}\overset{iid}{\sim}N(0,\sigma^2),i=1,2,3,\ j=1,2,3,4$

(1)獨立性(2)常態性(3) $E(\varepsilon_{ij})=0$ (4)變異數齊一性(5)模型正確性

(二) $SST=5410-\frac{252^2}{12}=118$

$$SSTR=\frac{92^2}{4}+\frac{86^2}{4}+\frac{74^2}{4}-\frac{252^2}{12}=42$$

(三)

| ANOVA TABLE | | | | |
|---|---|---|---|---|
| source | SS | d.f. | MS | F |
| 品種 | 42 | 2 | 21 | 2.4869 |
| Error | 76 | 9 | 8.4444 | |
| Total | 118 | 11 | | |



$H_0:\mu_1=\mu_2=\mu_3$ vs $H_1$:至少一個$\mu_i\neq\mu_j$

T.S.: $F=\frac{MSR}{MSE}\sim F_{(2,9)}$

R.R.: Reject $H_0$ at $\alpha = 0.05$ if $F^* > F_{0.05(2,9)} = 4.26$

$\because F^* = 2.4869 \qquad \therefore don't\,reject H_0$

結論：我們沒有足夠的統計證據推論三種品種生產力存在差異

五、某農業試驗場對同種作物分別施以A肥料及B肥料，其各5塊實驗田之收穫量相關資料如下（單位相同），今假設兩種肥料之收穫量服從常態分配，並具相同之變異數。

A肥料：$\bar{x}_A = 34,\ s_A^2 = \dfrac{\Sigma(x_1 - \bar{x}_1)^2}{n_A - 1} = 6$

B肥料：32　34　30　34　30

(一)試求施以A肥料之所有該作物平均收穫量的95%信賴區間。（10分）

(二)試在α=0.05的顯著水準下，檢定施以兩種肥料之所有該作物的平均收穫量是否有差異？（10分）

| | |
|---|---|
| **試題評析** | 本題涉及兩獨立常態母體平均數之假設檢定，屬於假設檢定中的基本計算題型，考生要拿滿分不難。 |
| **考點命中** | 《高點・高上統計學講義》第五回，趙治勳編撰，頁20。 |

**答**：

令 $X_1, X_2$ 分別表A,B肥料之收穫量

母體：$X_1 \sim N(\mu_1, \sigma^2) \perp X_2 \sim N(\mu_2, \sigma^2)$　假設(1) $X_1 \perp X_2$ (2)隨機樣本

樣本：$X_{11}, X_{12}, \cdots, X_{15} \overset{iid}{\sim} N(\mu_1, \sigma^2)$，$X_{21}, X_{22}, \cdots, X_{25} \overset{iid}{\sim} N(\mu_2, \sigma^2)$

點估計：$\overline{X_1} = \dfrac{\sum_{i=1}^{5} X_{1i}}{5}$ 及 $\overline{X_2} = \dfrac{\sum_{i=1}^{5} X_{2i}}{5}$

$$S_1^2 = \frac{\sum (X_{1i} - \bar{X}_1)^2}{5-1} \text{ 及 } S_2^2 = \frac{\sum (X_{2i} - \bar{X}_2)^2}{5-1}$$

(一)

樞紐量：$T = \dfrac{\overline{X_1} - \mu_1}{S_1 / \sqrt{5}} \sim t_{(4)}$

機率區間：$P(-t_{(4)0.025} \le \dfrac{\overline{X_1} - \mu_1}{S_1/\sqrt{5}} \le t_{(4)0.025}) = 0.95$



信賴區間：$P(\overline{X_1} - t_{(4)0.025}\dfrac{S_1}{\sqrt{5}} \le \mu_1 \le \overline{X_1} + t_{(4)0.025}\dfrac{S_1}{\sqrt{5}}) = 0.95$

結論：$\mu_1$ 之95%信賴區間為

$$(\overline{X_1} \mp t_{(4)0.025} \frac{S_1}{\sqrt{5}}) = (34 \mp 2.776 \frac{\sqrt{6}}{\sqrt{5}}) = (30.959, 37.041)$$

(二)

$H_0 : \mu_1 = \mu_2$ vs $H_1 : \mu_1 \neq \mu_2$

T.S.: $T = \dfrac{\overline{X_1} - \overline{X_2} - (0)}{\sqrt{S_p{}^2(\frac{1}{5} + \frac{1}{5})}} \sim t_{(8)}$ 其中 $S_p^2 = \dfrac{(5-1)S_1^2 + (5-1)S_2^2}{5+5-2} = 5$

R.R.: Reject $H_0$ at $\alpha = 0.05$ if $|T^*| > t_{0.05(8)} = 2.306$

$$\because |T^*| = \left| \frac{34 - 32 - (0)}{\sqrt{5(\frac{1}{5} + \frac{1}{5})}} \right| = 1.4142 \qquad \therefore don't\ reject\ H_0$$

結論：我們沒有足夠證據去推論兩種肥料下之平均收穫量有差異。