《抽樣方法》

試題評析
今年考題較往年靈活，只要同學能把握解題的關鍵點，依序做答，應可拿到不錯的分數。預計今年高分會較去年下降20分左右，程度好的考生預估可拿70分左右，而一般考生則可能落在50分左右，各題型分析如下：
第一題：為簡單隨機抽樣法與傳統統計學觀念兩者合併命題，不難作答。
第二題：於上課補充教材中已特別說明，更於總複習課程第三堂模擬考中的第二題有相同考題。
第三題：為分層隨機抽樣，需對分層樣本配置觀念非常清楚才能拿高分。
第四題：此題為比例機率抽樣法，曾於84高考中出現過，並非傳統常出現之抽樣考題，一般考生不易作答。

一、某次選舉甲選區共有3位候選人，現欲了解選民對候選人的支持情形，進行了一項民調，隨機訪問了750位的選民，此次民調結果如下：

	支持1號候選人	支持2號候選人	支持3號候選人	尚未決定	合計
人數	210	180	120	240	750

(一)試估計1號候選人的支持率p_1。(5分)
(二)在95%的信賴度下，1號候選人的支持率p_1是否顯著高於2號候選人的支持率p_2？(10分)
(三)若甲選區共有4500位選民且1號候選人是首次參選，現希望估計1號候選人之支持率p_1的信賴度為95%且誤差界限為0.03，則訪問750位選民是否足夠，若不夠須再訪問多少位選民方能達到此要求？(10分)

註：$z_{0.025}=1.96$

答：

(一)

$E(p_1)=P_1$，所以可得 $p_1=\dfrac{210}{750}=0.28$。

(二)此為非獨力性問題，任何一位候選人支持度變化，皆會影響到其它候選人的支持度。

$$\hat{p}_1-\hat{p}_2\pm 2\sqrt{\frac{\hat{p}_1\hat{q}_1}{n}+\frac{\hat{p}_2\hat{q}_2}{n}+2\frac{\hat{p}_1\hat{p}_2}{n}}\qquad \hat{p}_1=0.28\quad \hat{p}_2=0.24$$

$$0.28-0.24\pm 2\sqrt{\frac{0.28\times0.72}{750}+\frac{0.24\times0.76}{750}+2\frac{0.28\times0.24}{750}}$$

$[0.04\pm0.05258]$

因為信賴區間包含0，所以沒有足夠的資訊可以說明1號候選人的支持率顯著高於2號候選人的支持率。

(三)

$$n_0=\left(\frac{Zs}{B}\right)^2=\left(\frac{1.96^2\times\dfrac{750}{750-1}\times0.28\times0.72}{0.03^2}\right)=\left(\frac{0.77550056}{0.0009}\right)=861.667\cong 862$$

$$n=\frac{n_0}{1+\dfrac{n_0}{N}}=\frac{862}{1+\dfrac{862}{4500}}=\frac{862}{1+0.191555}=\frac{862}{1.191555}=723.424\cong 724$$

抽出n=724即可達到題意之要求，所以題目中抽出750是足夠的。

【參考書目】
抽樣方法講義第一回，P.37-P.38。

二、某零售商去年每月的平均營業額為880萬元，為了解今年每月的平均營業額，自其500種營業項目中抽出12種，調查去年的每月平均營業額（x）及今年1～6月的每月平均營業額（y），得下列的資料（單位：萬元）：

$\sum_{i=1}^{12} x_i = 10103$，$\sum_{i=1}^{12} y_i = 11458$，$\sum_{i=1}^{12} x_i y_i = 10196528$，$\sum_{i=1}^{12} x_i^2 = 9152367$，$\sum_{i=1}^{12} y_i^2 = 11453476$。

（一）利用迴歸估計式（regression estimator）估計今年每月的平均營業額，並求算此估計式的95%誤差界限。（10分）

（二）利用差量估計式（difference estimator）估計今年每月的平均營業額，並求算此估計式的95%誤差界限。（10分）

（三）計算迴歸估計式對差量估計式的相對效率（relative efficiency）且評估那種估計式較適合此個案。（5分）

註：$z_{0.025} = 1.96$

答：

(一)

$$\sum_{i=1}^{12}(x_i-\bar{x})^2 = \sum_{i=1}^{12} x_i^2 - \frac{\left(\sum_{i=1}^{12} x_i\right)^2}{n} = 9152367 - \frac{(10103)^2}{12} = 9152367 - 8505884.083 = 646482.917$$

$$\sum_{i=1}^{12}(y_i-\bar{y})^2 = \sum_{i=1}^{12} y_i^2 - \frac{\left(\sum_{i=1}^{12} y_i\right)^2}{n} = 11453476 - \frac{(11458)^2}{12} = 11453476 - 10940480.333 = 512995.667$$

$$\sum_{i=1}^{12}(x_i-\bar{x})(y_i-\bar{y}) = \sum_{i=1}^{12} x_i y_i - \frac{\sum_{i=1}^{12} x_i \sum_{i=1}^{12} y_i}{n} = 10196528 - \frac{(10103)(11458)}{12} = 10196528 - 9646681.166$$

$$= 549846.834$$

$$b_1 = \frac{\sum_{i=1}^{12}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{12}(x_i-\bar{x})^2} = \frac{549846.834}{646482.917} = 0.8505$$

$$SSE = \sum_{i=1}^{12}(y_i-\bar{y})^2 - b_1^2 \sum_{i=1}^{12}(x_i-\bar{x})^2 = 512995.667 - 0.72335025 \times 646482.917$$

$$= 512995.667 - 467633.579 = 45362.088$$

$$MSE = \frac{SSE}{n-2} = \frac{45362.088}{12-2} = 4536.2088 \ldots\ldots(1)$$

$$\mathrm{var}(\bar{y}_{lr}) = (1-f)\frac{MSE}{n} = \left(1-\frac{12}{500}\right)\frac{4536.2088}{12} = 0.976 \times 378.0174 = 368.9449$$

$$B = Z_{\alpha/2} \cdot s_{\bar{y}_{lr}} = 1.96 \times 19.207 = 37.6475$$

(二)

$$\sum_{i=1}^{12} d_i^2 = \sum_{i=1}^{12}(x_i - y_i)^2 = \sum_{i=1}^{12} x_i^2 - 2\sum_{i=1}^{12} x_i y_i + \sum_{i=1}^{12} y_i^2$$

$$= 9152367 - 2\times 10196528 + 11453476 = 20605843 - 20393056 = 212787$$

$$\sum_{i=1}^{12} d_i = \sum_{i=1}^{12}(x_i - y_i) = \sum_{i=1}^{12} x_i - \sum_{i=1}^{12} y_i = 10103 - 11458 = -1355$$

$$s_d^2 = \frac{\sum_{i=1}^{12}(d_i - \bar{d})^2}{n-1} = \frac{1}{n-1}\left[\sum_{i=1}^{12} d_i^2 - \frac{\left(\sum_{i=1}^{12} d_i\right)^2}{n}\right] = \frac{1}{12-1}\left[212787 - \frac{(-1355)^2}{12}\right]$$

$$= \frac{1}{11}[212787 - 153002.08] = 5434.9927.........(2)$$

$$\operatorname{var}(\bar{y}_{lr}) = (1-f)\frac{s_d^2}{n} = \left(1 - \frac{12}{500}\right)\frac{5434.9927}{12} = 0.976\times 452.91605 = 442.04606$$

$$B = Z_{\alpha/2}\cdot s_{\bar{y}_{lr}} = 1.96\times 21.024891 = 41.208786$$

(三)

迴歸估計式：$MSE = \dfrac{SSE}{n-2} = \dfrac{45362.088}{12-2} = 4536.2088......(1)$

差異估計式：

$$s_d^2 = \frac{\sum_{i=1}^{12}(d_i - \bar{d})^2}{n-1} = \frac{1}{n-1}\left[\sum_{i=1}^{12} d_i^2 - \frac{\left(\sum_{i=1}^{12} d_i\right)^2}{n}\right] = \frac{1}{12-1}\left[212787 - \frac{(-1355)^2}{12}\right]$$

$$= \frac{1}{11}[212787 - 153002.08] = 5434.9927.........(2)$$

因MSE=4536.2088小於$s_d^2 = 5434.9927$，所以迴歸估計式相對較優良。

三、某工廠管理者欲估計員工在面臨緊急事件時的平均反應時間。管理者認為不同性別的員工在反應時間上可能有所差異；因此，決定以性別來進行分層。已知工廠共有員工100人，其中男員工有45人，先前的研究顯示男員工的反應時間為5到20秒，女員工的反應時間為3到13秒。每位員工的調查成本是相同的。

(一)決定各層最適的樣本配置。（10分）

(二)若希望估計員工在面臨緊急事件時的平均反應時間的信賴度為95%且誤差界限為1秒，則需要調查多少位男、女員工？（10分）

(三)若每位員工的調查成本是$50元，且調查總費用的預算為$2000元，則調查多少位男、女員工才能使估計式的變異數最小。（5分）

註：$z_{0.025} = 1.96$；標準差約為全距的1／4。

答：

性 別	男 性 N_1	女 性 N_2
N=100	$N_1=45$	$N_2=55$
S_i^2 變異數	$S_1\cong\dfrac{20-5}{4}=3.75$	$S_2\cong\dfrac{13-3}{4}=2.5$

(一)由柯西不等式中，可得如下

$$(c_1n_1+\dots\dots+c_nn_n)\left[\frac{N_1^2S_1^2}{n_1}+\dots\dots\dots+\frac{N_L^2S_L^2}{n_L}\right]\ge\left(\sqrt{c_1}N_1S_1+\dots\dots\dots+\sqrt{c_L}N_LS_L\right)^2$$

$C\cdot V$有最小值時,可以上式轉換成如下：

(1)分層談明配置：$n_h=\dfrac{N_hS_h/\sqrt{c_h}}{\sum_{h=1}^{L}N_hS_h/\sqrt{c_h}}\cdot n$

(2)假設 $c_1=c_2\dots\dots=c_L$，分層紐門配置：$n_h=\dfrac{N_hS_h}{\sum_{h=1}^{L}N_hS_h}\cdot n$

(3)假設 $c_1=c_2\dots\dots=c_L$ 且 $s_1=s_2\dots\dots=s_L$，分層比例配置：$n_h=\dfrac{N_h}{\sum_{h=1}^{L}N_h}\cdot n$

(4)當每位員工調查成本是相同，並假定男性與女性兩層之調查成本相同，所以可得 $c_1=c_2$，但 $S_1\ne S_2$，所以我們採用分層紐門配置來決定各層最適之樣本。

$$n_1=\frac{N_1S_1}{\sum_{i=1}^{2}N_hS_h}\cdot n=\frac{45\times3.75}{45\times3.75+55\times2.5}\cdot n=\frac{168.75}{168.75+137.5}\cdot n=\frac{168.75}{306.25}\cdot n=0.55102\cdot n$$

$$n_2=\frac{N_2S_2}{\sum_{i=1}^{2}N_hS_h}\cdot n=\frac{55\times2.5}{45\times3.75+55\times2.5}\cdot n=\frac{137.5}{168.75+137.5}\cdot n=\frac{137.5}{306.25}\cdot n=0.44897\cdot n$$

(二)承(一)中(4)假定，可得 $n=\dfrac{\sum_{h=1}^{L}\dfrac{N_h^2s_h^2}{w_h}}{\dfrac{B^2N^2}{Z^2}+\sum_{h=1}^{L}N_hs_h^2}$，並以紐門配置，可以轉成如下

$$n=\frac{\left[\sum_{h=1}^{2}N_hs_h\right]^2}{\dfrac{B^2N^2}{Z^2}+\sum_{h=1}^{2}N_hs_h^2}=\frac{(306.25)^2}{\dfrac{1^2\cdot100^2}{1.96^2}+\left(45\cdot3.75^2+55\cdot2.5^2\right)}=\frac{93789.0625}{\dfrac{10000}{3.8416}+(632.8125+343.75)}=\frac{93789.0625}{2603.082+976.5625}$$

$$=\frac{93789.0625}{3579.6445}=26.2006\cong27$$

$$n_1=\frac{N_1S_1}{\sum_{i=1}^{2}N_hS_h}\cdot 27=0.55102\cdot 27=14.87754\cong 15$$

$$n_2=\frac{N_2S_2}{\sum_{i=1}^{2}N_hS_h}\cdot 27=0.44897\cdot 27=12.12219\cong 12$$

(三)當每位員工調查成本是相同，且$50n_1+50n_2=2000$，$50(n_1+n_2)=2000$，$(n_1+n_2)=40$

$(n_1+n_2)=n=40$，男性與女性兩層之調查成本c_1與c_2未知，$50n_1=c_1$與$50n_2=c_2$

以談明配置$n_h=\dfrac{N_hS_h/\sqrt{c_h}}{\sum_{h=1}^{L}N_hS_h/\sqrt{c_h}}\cdot n$，可以轉成如下：

$$n_1=\frac{N_1S_1/\sqrt{c_1}}{\sum_{h=1}^{2}N_hS_h/\sqrt{c_h}}\cdot 40............(1)$$

$$n_2=\frac{N_2S_2/\sqrt{c_2}}{\sum_{h=1}^{2}N_hS_h/\sqrt{c_h}}\cdot 40............(2)$$

$\dfrac{(1)}{(2)}$可得如下，$\dfrac{n_1}{n_2}=\dfrac{\frac{N_1S_1}{\sqrt{c_1}}}{\frac{N_2S_2}{\sqrt{c_2}}}$，$\dfrac{n_1}{n_2}=\dfrac{N_1S_1}{N_2S_2}\cdot\dfrac{\sqrt{c_2}}{\sqrt{c_1}}$，$\dfrac{n_1}{n_2}=\dfrac{N_1S_1}{N_2S_2}\cdot\dfrac{\sqrt{50\cdot n_2}}{\sqrt{50\cdot n_1}}$，$\dfrac{n_1}{n_2}=\dfrac{N_1S_1}{N_2S_2}\cdot\dfrac{\sqrt{n_2}}{\sqrt{n_1}}$

$\dfrac{n_1^{3/2}}{n_2^{3/2}}=\dfrac{N_1S_1}{N_2S_2}$，$\left(\dfrac{n_1}{n_2}\right)^{3/2}=\dfrac{N_1S_1}{N_2S_2}=\dfrac{45\cdot 3.75}{55\cdot 2.5}=\dfrac{168.75}{137.5}=1.22727$，$\dfrac{n_1}{n_2}=1.146287$

$n_1=1.146287n_2$，$2.146287n_2=40$，$n_2=18.63\cong 19$，$n_1=21.36\cong 21$

【參考書目】
抽樣方法講義第二回，P.26-P.28與抽樣方法架構表中的第二頁。

四、品管人員想要了解每張電路板之晶片損壞的平均個數。由於每張電路板所包含的晶片個數不完全相同，且每張電路板上損壞晶片的個數與電路板上的晶片總數有高度的相關。因此，品管人員依照每張電路板上的晶片個數，利用依大小成比例的機率抽樣法（sampling with probabilities proportional to size, PPS）抽樣。某天生產的10張電路板的晶片個數如下：8, 12, 22, 10, 16, 24, 9, 10, 9, 30。

(一)若利用亂數表隨機選取的數字為7, 66, 30, 140，請利用PPS抽樣法抽出4張電路板為集體樣本。（12分）

(二)若挑選出的電路板為第2, 3, 5, 7張電路板，其中損壞的晶片數分別為1, 3, 2, 1。估計每張電路板之晶片損壞的平均個數，並求算此估計式的95%誤差界限。（13分）

註：$z_{0.025}=1.96$

答：

ANS：

在n=4個電路板被抽樣，我們必須隨機從000跟150之間抽4個隨機號碼，

看亂數表(A)第一行、第一列開始選07、66、30、140，這些號碼對應到電路板如下。

(1)

晶片張數	晶片個數	累積區間	依亂數表選張數
1	8	1 ~ 8	∨
2	12	9 ~ 20	
3	22	21 ~ 42	∨
4	10	43 ~ 52	
5	16	53 ~ 68	∨
6	24	69 ~ 92	
7	9	93 ~ 101	
8	10	102 ~ 111	
9	9	112 ~ 120	
10	30	121 ~ 150	∨

(2)N = 10，n =4，M = 150

晶片張數	晶片個數 m_i	損壞晶片 y_i	$z_i=\dfrac{m_i}{M}$
2	12	1	$\frac{12}{150}$
3	22	3	$\frac{22}{150}$
5	16	2	$\frac{16}{150}$
7	9	1	$\frac{9}{150}$

1.晶片損壞之總數：$\hat{A}_{PPs}=\dfrac{1}{n}\sum\dfrac{A_i}{Z_i}=\dfrac{1}{n}\sum\dfrac{A_i}{\frac{M_i}{M}}=\dfrac{1}{4}\left[\dfrac{1}{\frac{12}{150}}+\dfrac{3}{\frac{22}{150}}+\dfrac{2}{\frac{16}{150}}+\dfrac{1}{\frac{19}{150}}\right]=14.899$

2.平均每一張電路板中損壞晶片個數：$\bar{A}_{pps}=\dfrac{1}{N}\cdot\dfrac{1}{n}\sum\dfrac{A_i}{Z_i}=\dfrac{1}{10}\cdot 14.899=1.4899$

3. $\hat{\bar{A}}_{pps}=\frac{1}{N}\cdot\frac{M}{n}\cdot\sum\frac{A_i}{M_i}=\frac{1}{N}\cdot\frac{M}{n}\cdot\sum\bar{A}_i$

$$\operatorname{var}\left(\hat{\bar{A}}_{pps}\right)=\frac{1}{N^2}\cdot\frac{M^2}{n}\cdot\frac{\sum\left(\bar{A}_i-\bar{\bar{a}}_t\right)^2}{n-1}=\frac{1}{10^2}\cdot\frac{150^2}{4}\cdot\frac{0.001579}{4-1}=0.0296$$

$$s_{\hat{\bar{A}}_{pps}}=0.17204 \quad B=1.96\times0.17204=0.3372 \quad \left[\hat{\bar{A}}_{pps}\pm B\right]=\left[1.4899\pm0.3372\right]$$

【參考書目】
正課第13堂中之補充資料，說明比例機率抽樣法。