

《抽樣方法》

一、解釋下列名詞：（每小題5分，共10分）

（一）抽樣誤差（Sampling Error）

（二）非抽樣誤差（Nonsampling Error）

試題評析	本題是有關抽樣誤差與非抽樣誤差之解釋名詞，考古題中也出現過，講義中也有特別針對這些誤差進行解釋，只要熟讀講義及作答流暢，就可以拿到高分。
考點命中	《高點·高上抽樣方法講義》第一回，趙治勳編撰，頁5。

答：

（一）抽樣誤差：

抽樣調查中只是利用樣本資訊去推論母體參數，或多或少會產生估計誤差，即估計量與母體參數之間存在差異，故抽樣誤差是不可避免的。抽樣誤差是可以利用公式計算誤差大小並且可以控制的，一般常用的抽樣誤差型式，如：

1. 絕對誤差(absolute error)： $|\hat{\theta} - \theta|$

2. 相對誤差(absolute error)： $|\frac{\hat{\theta} - \theta}{\theta}|$

3. 均方誤差(mean square error of an estimator)： $MSE(\hat{\theta}) = E(\hat{\theta} - \theta)^2 = V(\hat{\theta}) + [E(\hat{\theta}) - \theta]^2$

（二）非抽樣誤差：

抽樣誤差與非抽樣誤差最大之不同就是抽樣誤差是可以利用公式計算誤差大小並且可以控制的，但非抽樣誤差是不可量化的。非抽樣誤差不是來自於抽樣過程中之隨機性，而是由於其他原因造成估計量與母體參數之間存在差異，這些原因例如：

1. 無答覆性誤差：底冊未編完整或調查對象避不答覆。
2. 答覆性誤差：調查表中規定事項不明確使得問答雙方理解不一致、或訪員的訪問技巧不足、或受訪者口齒不清、或受訪者對於問題產生記憶誤差。
3. 工作效率低：數字寫錯、單位弄錯或小數點定位錯。
4. 估計技術錯誤：用錯統計分析方法。

二、請敘述簡單隨機抽樣與事後分層隨機抽樣的異同處。（10分）

試題評析	本題是兩個抽樣法之比較，雖然兩個抽樣方法在課堂上都有說明過，但這類申論題沒有標準答案，考生只能盡量敘述清楚，讓閱卷老師肯定你的論述，但最起碼兩個抽樣方法於講義中都有介紹，要拿到高分應該不困難。
考點命中	《高點·高上抽樣方法講義》第一回，趙治勳編撰，第二章頁10與第六章頁36。

答：

（一）簡單隨機抽樣(SRS)：

從抽樣母體N中抽取n個作為樣本，其中每一組可能的樣本均有相等的被抽出機率，再利用樣本資訊去推論母體參數。然而，在實際抽取上可以使用亂數表進行抽樣。

（二）事後分層隨機抽樣：

有些時候，研究者要等到樣本選取之後，才能夠真正地瞭解分層標準之正確性，這時可以利用樣本所呈現之訊息再對於估計量進行修正，這種方法就稱為事後分層隨機抽樣法。此法會先從抽樣母體N中抽取n個，再由這些樣本結構進行分層，然後於各層內以SRS抽取樣本，再以各層所得之抽樣對於母體參數進行推估。

（三）兩者比較：

【版權所有，重製必究！】

事後分層於一開始從母體 N 中抽取 n 個，實為一個簡單隨機抽樣法，而事後分層與SRS最大差異在於事後分層會再以之前SRS取得之樣本資訊分層抽樣。

三、何謂隨機性母體 (Random Population)、順序性母體 (Ordered Population)、週期性母體 (Periodic Population) ? (15分)

試題評析	本題是三種母體的解釋名詞，這三種母體在講義中都有出現過，只要認真聽課，要拿到高分不難，建議考生在作答時可以搭配圖形說明。
考點命中	《高點·高上抽樣方法講義》第一回，趙治勳編撰，頁24。

答：

(一)隨機性母體：

若抽樣母體之觀察值於母體編號上沒有特殊情形出現，即隨機地分佈於整個母體上的話，則稱為隨機性母體。

(二)順序性母體：

若抽樣母體之觀察值於母體編號上有呈現出由小到大或由大到小的排序情形，則稱為順序性母體。

(三)週期性母體：

若抽樣母體之觀察值於母體編號上，呈現出一種在某間隔(通常是年)內的分布情形於往後每一個相同的間隔中一再出現，則稱為週期性母體。

在抽樣方法中，以上三種母體曾經一同出現在比較SRS與系統抽樣法之精確度上，當隨機性母體下，SRS與系統抽樣法之精確度相當；當順序性母體下，系統抽樣法之精確度較高；當週期性母體下，系統抽樣法之精確度較差。

四、依105年6月臺北市政府網站的最新戶口資料，臺北市共有12個行政區、456個里，總計有1,044,441戶。一項主計處的研究是以往戶為抽樣單位的抽樣調查，預計抽出1,068戶組成樣本。請設計三種合理的抽樣方式(簡單隨機抽樣、分層隨機抽樣、二階段抽樣)，可使每戶有相同被抽出的機率。(30分)

試題評析	本題是要敘述SRS，分層隨機與兩階段抽樣法在抽取樣本的過程，因為都在講義中說明過，要拿到滿分並不難。
考點命中	《高點·高上抽樣方法講義》第一回，趙治勳編撰，第二章頁10、第五章頁28與第十七章頁110。

答：

(一)簡單隨機抽樣(SRS)：

先把台北市1044441戶編號為1至1044441號，利用亂數表產生1068個不重複之樣本組成簡單隨機樣本。先隨機決定亂數表起始位置之列與行，從左至右每次取七位數以決定抽樣單位(註:若號碼大於1044441就不管它再往下繼續讀)直到抽出1068個介於(1~1044441)的數為止

(二)分層隨機抽樣：

先將台北市按照行政區分為12個層，再於每一個層中以SRS(過程同上亂數表)抽取樣本進而組成分層隨機樣本。

(三)二階段抽樣：

先將台北市按照行政區分為12個層，以SRS(過程同上亂數表)抽取 n 個層出來，並於之前抽取出來的 n 個層中的每一個以SRS(過程同上亂數表)抽出 $m_i, i = 1, 2, \dots, n$ 個樣本進而組成二階段隨機樣本。

五、某一廣告公司欲調查三個區域 ($N_1 = 2,500$ 戶, $N_2 = 5,000$ 戶, $N_3 = 15,000$ 戶) 住戶的薪資狀況。假設此一廣告公司想要估計這22,500戶每戶平均薪資的95%信賴區間，且要求誤差的界限為 $B = 300$, $Z = 2$ 。又由過去的研究得知各區的母體住戶薪資的標準差為 $\sigma_1 = 4,500$, $\sigma_2 = 6,000$, $\sigma_3 = 9,000$ 。若此一廣告公司是採取分層比例隨機抽樣法。

【版權所有，重製必究！】

(一)請問此廣告公司應該抽取多少樣本？(5分)

(二)這些樣本如何配置到這三個區域？(5分)

(三)你覺得以分層比例隨機抽樣法抽樣是否合適？如果不合適，你會採取何種配置法？應該抽取多大之樣本？又如何將這些樣本配置到這三個區域？(10分)

試題評析	本題是分層隨機抽樣法中總樣本數及配置之計算題，上課講義中已經有相同範例，考古題中也出現過很多，只要計算公式沒有忘記且沒計算錯誤，要拿到滿分並不難。
考點命中	《高點·高上抽樣方法講義》第一回，趙治勳編撰，頁33、34。

答：

(一)比例配置

$$n = \frac{N \sum_{h=1}^L N_h S_h^2}{VN^2 + \sum_{h=1}^L N_h S_h^2} = \frac{22500 \times 1.445625 \times 10^{12}}{22500 \times 22500^2 + 1.445625 \times 10^{12}} = 2533.9614 \approx 2534$$

$$\text{其中 } V = \left(\frac{B}{z_{\frac{\alpha}{2}}} \right)^2 = \left(\frac{300}{2} \right)^2 = 22500$$

$$\sum_{h=1}^L N_h S_h^2 = 2500 \times 4500^2 + 5000 \times 6000^2 + 15000 \times 9000^2 = 1.445625 \times 10^{12}$$

$$(二) n_1 = \frac{N_1}{N} \times n = \frac{2500}{22500} \times 2534 = 281.5555 \approx 282$$

$$n_2 = \frac{N_2}{N} \times n = \frac{5000}{22500} \times 2534 = 563.1111 \approx 563$$

$$n_3 = n - n_1 - n_2 = 2534 - 282 - 563 = 1689$$

(三)我認為比例配置法並不適合本例，因為第三區域之住戶數明顯多於其他兩個區域。我建議改採紐門配置法，計算過程如下：

$$n = \frac{\left(\sum_{h=1}^L N_h S_h \right)^2}{VN^2 + \sum_{h=1}^L N_h S_h^2} = \frac{(176250000)^2}{22500 \times 22500^2 + 1.445625 \times 10^{12}} = 2420.0263 \approx 2421$$

$$\text{其中，由(一)已得 } V = 22500, \sum_{h=1}^L N_h S_h^2 = 1.445625 \times 10^{12}$$

$$\sum_{h=1}^L N_h S_h = 2500 \times 4500 + 5000 \times 6000 + 15000 \times 9000 = 176250000$$

$$n_1 = \frac{N_1 S_1}{\sum_{h=1}^L N_h S_h} \times n = \frac{2500 \times 4500}{176250000} \times 2421 = 154.5319 \approx 155$$

$$n_2 = \frac{N_2 S_2}{\sum_{h=1}^L N_h S_h} \times n = \frac{5000 \times 6000}{176250000} \times 2421 = 412.0851 \approx 412$$

【版權所有，重製必究！】

$$n_3 = n - n_1 - n_2 = 2421 - 155 - 412 = 1854$$

六、某一個公司共有員工人數300 ($M = 300$) 位，分佈在30 ($N = 30$) 個部門，若想估計全公司員工請假天數的總和，先以簡單隨機抽樣法自此公司抽出3 ($n = 3$) 個部門，之後又從此三個部門抽出一半員工做為樣本。若抽出的3個部門的每一個部門的人數分別為 $M_1 = 18$ 、 $M_2 = 36$ 、 $M_3 = 20$ ，又各抽出一半員工，故 $m_1 = 9$ 、 $m_2 = 18$ 、 $m_3 = 10$ ，其三個抽出部門樣本的平均請假天數為 $\bar{y}_1 = 10$ ， $\bar{y}_2 = 15$ ， $\bar{y}_3 = 12$ ，及其樣本變異數為 $s_1^2 = 27$ ， $s_2^2 = 24$ ， $s_3^2 = 20$ 。

(一) 試求全公司員工請假天數總和的兩階段估計值。(5分)

(二) 試求全公司員工請假天數總和兩階段估計值標準誤的估計值。(10分)

試題評析	本題是兩階段抽樣法的計算題，上課講義中已經有相同範例，考古題中也出現過很多，只要計算公式沒有忘記且沒計算錯誤，要拿到滿分並不難。
考點命中	《高點·高上抽樣方法講義》第一回，趙治勳編撰，頁113。

答：

$$(一) \hat{Y} = \frac{N}{n} \sum_{i=1}^n M_i \bar{y}_i = \frac{30}{3} (18 \times 10 + 36 \times 15 + 20 \times 12) = 9600 \text{ (天)}$$

$$(二) s_{\hat{Y}} = \sqrt{N^2(1-f_1) \frac{s_{1b}^2}{n} + \frac{N}{n} \sum_{i=1}^n M_i (M_i - m_i) \frac{s_{2i}^2}{m_i}}$$

$$= \sqrt{30^2 \left(1 - \frac{3}{30}\right) \frac{37200}{3} + \frac{30}{3} \left[18(18-9) \frac{27}{9} + 36(36-18) \frac{24}{18} + 20(20-10) \frac{20}{10}\right]}$$

$$= 3171.9868 \text{ (天)}$$

【版權所有，重製必究！】