104年公務人員升官等考試、104年關務人員升官等考試、104年交通事業公路、港務人員升資考試試題

等　別：薦任
類　科：統計
科　目：抽樣方法

一、解釋下列名詞：
　㈠機率抽樣（Probability Sampling）
　㈡非機率抽樣（Nonprobability Sampling）

【擬答】：

㈠按照機率原則，從母體中抽取一定數目的單位元素作為樣本進行觀察，機率抽樣使母體中每個單位都有一定的機率被選入樣本，其抽樣步驟稱為機率抽樣。

㈡樣本不按照機率理論予以抽出，而是由抽樣者之主觀抽出或自願樣本。

二、如果要以電話抽樣調查方式研究某縣市的生活狀況調查，請描述要如何進行抽樣工作，才可以抽出一組簡單隨機樣本。

【擬答】：

早期較無效率的方式為藉亂數表在電話簿中抽出隨機樣本；因為電話簿的排列並沒有系統性的偏差，在電腦尚未普及之前，為了節省工作負擔起見，調查一般均使用系統抽樣。

而目前民調單位習慣採「電話號碼尾數隨機撥號」，以隨機亂數產生受訪者。故實際執行方式為先選定欲訪問的分區號碼(或系統業者碼)，再以電腦隨機亂數的方式選定後尾數，便可抽出簡單隨機樣本。

三、請回答下列問題：
　㈠何謂分層隨機抽樣？
　㈡何謂事後分層法？
　㈢何謂雙重抽樣法？
　㈣請寫出三種方法之差異。

【擬答】：

㈠將母體按某種標準分為若干子母體，這些子母體就稱為層。再從各層中，利用簡單隨機抽樣取出樣本，然後將各層隨機樣本合併起來即成一組分層隨機樣本。這種抽樣程序稱為分層隨機抽樣。

㈡在實務上的抽樣問題，有時並無完整的分層母體資訊，我們只能在樣本被選取之後，才可能設定抽樣單位到正確的層。換句話說，在分層隨機抽樣時，若無各分層母體的大小資訊，此時我們若僅瞭解(或利用樣本來估計)各分層的配置 下，在一組樣本選取後才進行分層，此方式稱為事後分層。

㈢雙重抽樣是指在抽樣時分兩次抽取樣本的一種抽樣方式。首先抽取一個初步樣本，並搜取一些簡單專案以獲得有關總體的資訊；然後，在此基礎上再進行深入抽樣。所以分層雙重抽樣是在分層隨機抽樣法中執行雙重抽樣的抽樣法。首先第一重是從抽樣母體中，採用簡單隨機抽樣抽出第一重抽樣單位，其分布在不同的分層。然後再從每一層第一重抽樣單位中以簡單隨機抽樣法分別抽出第二重抽樣單位。此時的樣本即為分層雙重抽樣所選取出的樣本。

㈣ *1.* 分層隨機抽樣母體平均數估計

$$\bar{y}_{st} = \frac{N_1\times\bar{y}_1 + N_2\times\bar{y}_2 + \cdots + N_L\times\bar{y}_L}{N} = \sum_{h=1}^{L} W_h\bar{y}_h \xrightarrow{估計} \bar{Y}$$

$$s_{\bar{y}_{st}} = \sqrt{\sum_{h=1}^{L} W_h^2(1-f_h)\frac{s_h^2}{n_h}} = \frac{1}{N}\sqrt{\sum_{h=1}^{L} N_h(N_h-n_h)\frac{s_h^2}{n_h}}$$

可知分層隨機抽樣需有分層母體大小資訊，即 N_h。

2. 事後分層法母體平均數估計

$$\bar{y}_{pst} = \sum_{h=1}^{L} A_h \cdot \bar{y}_h \xrightarrow{\text{估計}} \bar{Y}$$

$$s_{\bar{y}_{pst}} = \sqrt{\left(\frac{N-n}{Nn}\right)\sum_{i=1}^{L} A_h \cdot s_h^2 + \frac{1}{n^2}\sum_{i=1}^{L}(1-A_h)\cdot s_h^2}$$

若缺少分層母體大小資訊，而尚瞭解各分層的配置$(A_h = N_h/N)$，亦或利用樣本推估配置時使用。

3. 分層雙重抽樣法母體平均數估計

$$\bar{y}_{dst} = \sum g_h \bar{y}_h = \sum \frac{n_h'}{n'} \cdot \bar{y}_h \xrightarrow{\text{估計}} \bar{Y}$$

$$s_{\bar{y}_{dst}} = \sqrt{\frac{n'}{n'-1}\sum_{h=1}^{L}\left[g_h^2 - \frac{g'}{n'}g_h\right]\frac{s_h^2}{n_h} + \frac{g'}{n'-1}\sum_{h=1}^{L} g_h(\bar{y}_h - \bar{y}_{st})^2}$$

其中 $g' = \dfrac{N-n'}{N-1}$

當對母體資訊一無所知，此時可利用第一重樣本的配置來推估母體配置，再以第二重樣本來作分層的估計下使用。

四、請回答下列問題：

㈠分層隨機抽樣之每一層樣本的配置（Allocation），應注意那三項事情？

㈡何謂 Deming's Allocation（最優配置）？

㈢何謂 Neyman's Allocation？㈣何謂 Proportional Allocation？

【擬答】：

㈠分層隨機抽樣需根據以下分層的三種特性來決定其樣本配置，即分層母體大小、各分層變異程度、以及各分層抽樣樣本之成本來決定。

㈡ Deming's Allocation(談明配置)又稱為最優配適，考量的原則為母體大的要多抽、變異程度大的要多抽、成本大的要少抽。即 $n_h = \dfrac{\dfrac{N_h S_h}{\sqrt{c_h}}}{\sum_{h=1}^{L}\dfrac{N_h S_h}{\sqrt{c_h}}} \times n$

㈢在談明配置下，若不考慮各分層抽樣成本的高低，即為 Neyman's Allocation（紐曼配置）。

即 $n_h = \dfrac{N_h S_h}{\sum_{h=1}^{L} N_h S_h} \times n$

㈣在談明配置下，若不考慮各分層抽樣成本的高低，也不考慮各分層的變異程度，僅考慮各分層之大小時，稱為 Proportional Allocation(比例配置)。即 $n_h = \dfrac{N_h}{\sum_{h=1}^{L} N_h} \times n$

五、何謂重複系統抽樣（Repeat Systematic Sampling）？其使用時機為何？

【擬答】：

重複系統抽樣法是指在一母體中同時抽出許多系統樣本，而由這許多系統樣本組成一組樣本，此樣本即稱為重複系統樣本。

單一系統抽樣無法計算變異數，而是以簡單隨機抽樣的變異數予以取代。但二者有時無法取代，若要估計變異數時，則可採重複系統抽樣法，亦即抽出一個以上的系統樣本。重複系統抽

樣可得到多個系統樣本均數，便可用來推估母體的平均以及變異數。

六、請計算下列問題：

㈠某一工廠所生產奶茶的包裝都是以 12 包裝成一箱，現隨機抽取五箱，稱其每箱的重量，得到下列的數據：120.0、120.1、120.0、120.5、119.7 公克。假設此工廠所生產的奶茶數量相當龐大，試估計此工廠所生產的奶茶平均一包的重量及其變異數。

㈡此工廠為因應市場不同的需求，目前所生產奶茶的包裝有以 24 包、12 包及 6 包裝成一箱。此工廠每天所生產的奶茶箱數有 1,000 箱，24 包的奶茶箱數是 200 箱、12 包的奶茶箱數是 400 箱、6 包的奶茶箱數是 400 箱。因每箱奶茶的重量與每箱奶茶的包數有很大的正相關，故採用比例機率（pps）集群抽樣。現隨機抽取五箱，分別取得 24 包、12 包、12 包、6 包、6 包裝，稱其每箱的重量，分別得到下列的數據：240.0、120.1、120.0、60.25、59.785 公克，試估計此工廠所生產的奶茶平均一包的重量及其變異數。

【擬答】：

㈠ $\bar{y}_{rc}=\frac{N}{\hat{M}}\bar{y}_t=\frac{Y_1+Y_2+\cdots+Y_5}{M_1+M_2+\cdots+M_5}$

$$=\frac{120+120.1+120+120.5+119.7}{12\times5}=10.005$$

$$s_c^2=\frac{\sum_{i=1}^{n}(Y_i-\bar{y}_{rc}M_i)^2}{n-1}=\frac{\sum_{i=1}^{n}Y_i^2-2\bar{y}_{rc}\sum_{i=1}^{n}Y_iM_i+\bar{y}_{rc}^2\sum_{i=1}^{n}M_i^2}{n-1}$$

$$=\frac{72072.35-2\times10.005\times7203.6+10.005^2\times720}{5-1}=0.083$$

$$\mathrm{var}(\bar{y}_{rc})\approx\frac{1}{\bar{M}^2}(1-f)\frac{s_c^2}{n}\approx\frac{1}{12^2}\times\frac{0.083}{5}=0.000115$$

㈡ $\bar{y}_{pps}=\frac{1}{n}\sum\bar{Y}_i=\frac{1}{5}(\frac{240}{24}+\frac{120.1}{12}+\frac{120}{12}+\frac{60.25}{6}+\frac{59.785}{6})=10.0028$

$$\mathrm{var}(\bar{y}_{pps})=\frac{1}{n}\cdot\frac{\sum_{i=1}^{n}(\bar{Y}_i-\bar{\bar{y}}_t)^2}{n-1}=\frac{1}{5}\times0.000762=0.000152$$