

112年公務人員高等考試三級考試試題

類 科：統計
科 目：抽樣方法
考試時間：2 小時

座號：_____

※注意：(一)可以使用電子計算器。

(二)不必抄題，作答時請將試題題號及答案依照順序寫在試卷上，於本試題上作答者，不予計分。

(三)本科目除專門名詞或數理公式外，應使用本國文字作答。

一、欲了解某一工業園區 N 家製造業者AI人才的需求狀況，下述三種抽樣設計可用以推估該工業園區有AI人才需求的業者家數比例及總需求人數：

- (1)如果園區各業者的營業規模已知，首先將業者依其營業規模分成 L 層 ($1, \dots, L$)，每層家數分別為 $N_1, \dots, N_L$ ，再由每層抽取一簡單隨機樣本，分別為 $n_1, \dots, n_L$ ，以調查業者的AI人才需求狀況。
- (2)如果園區各業者的營業規模未知，但已知園區業者營業規模的比例分別為 $W_1, \dots, W_L$ ，首先由 N 家業者抽取一簡單隨機樣本(n)調查業者規模及AI人才需求狀況後，再根據調查結果依其規模分成 L 層進行推估。
- (3)如果沒有園區業者營業規模的資訊，首先由 N 家業者抽取一簡單隨機樣本 n' ，取得營業規模資訊，而後根據營業規模資訊將 n' 家業者依其營業規模分成 L 層 ($n'_1, \dots, n'_L$)，再由每層抽取一簡單隨機樣本，分別為 $n_1, \dots, n_L$ ，調查業者的AI人才需求狀況。

(一)說明前述三種抽樣設計的抽樣方法為何？(10分)

(二)若欲估計園區AI總需求人數(Y)，分別列出對應前述三種抽樣方法的估計量(estimator)及該估計量之變異數的估計量。(15分)

二、某一縣市共4000家養雞戶分散在20個村里 (cluster)，欲透過調查了解該縣市養雞戶的所得狀況，抽樣方法可採用一階段集體抽樣 (single-stage cluster sampling) 或二階段集體抽樣 (two-stage cluster sampling)。

(一)若抽樣方法採一階段集體抽樣，首先由20個村里以簡單隨機抽樣(SRS) 抽出3個村里，就抽得的3個村里之養雞戶全數調查，調查結果村里內養雞戶數及平均年所得列於下表：

村里 (cluster) i	養雞戶數 (M_i)	村里內養雞戶 平均每戶年所得 (百萬元) ($\bar{y}_i$)
1	150	4
2	200	8
3	250	5.6

採用集體大小比率估計量 (ratio-to-size estimator, $\bar{y}_R$) 估計該縣市養雞戶平均每戶年所得 (以百萬元為單位) 及該估計量之標準誤。(10分)

(二)若抽樣方法採二階段集體抽樣，首先由20個村里以簡單隨機抽樣抽出3個村里，再就抽得的3個村里之養雞戶以簡單隨機抽樣分別抽出1/10養雞戶進行調查，調查結果養雞戶之平均所得列於下表：

村里 (cluster) i	農戶數 (M_i)	抽出養雞戶數 (m_i)	村里內抽得之養雞戶平均 每戶年所得及標準差 (百萬元) ($\bar{y}_i, s_i$)	
			$\bar{y}_i$	s_i
1	150	15	4	1
2	200	20	7	3
3	250	25	6	2

試問：(15分)

(1)本抽樣設計第一階段的抽樣單位 (primary sampling unit, PSU) 及第二階段的抽樣單位 (secondary sampling unit, SSU) 分別為何？

(2)採用不偏估計量 (unbiased estimator, $\bar{y}$) 估計該縣市養雞戶平均每戶年所得 (以百萬元為單位) 及該估計量之標準誤。

三、欲了解某區域養殖漁戶營運狀況，該區域共有800家養殖漁戶，首先由該區域抽得一個包含500戶（第一重樣本 $n'=500$ ）的簡單隨機樣本以取得養殖型態（是否為漁電共生的層別資訊），調查得知其中100戶為漁電共生戶，進而以簡單隨機抽樣由漁電共生及非漁電共生的養殖漁戶分別抽20%以調查其營運成本資訊，調查結果整理如下表：

養殖型態（層別）	第一重樣本 （戶） n'_h	第二重樣本 （戶） n_h	養殖漁戶年營運成本	
			平均年營運成本 $\bar{y}_i$ （十萬元）	標準差 s_i （十萬元）
漁電共生（I）	100	20	120	100
非漁電共生（II）	400	80	50	60
合計	500（ n' ）	100（ n ）		75（ s ）

(一)估計該區域養殖漁戶平均年營運成本（ $\bar{Y}$ ）（以十萬元為單位）及該估計量的標準誤。（10分）

(二)如果此調查總預算為44,500元，取得養殖型態的單位成本為9元，調查營運成本的單位成本為400元（ $c'=9$, $c_h=400$ ），有關標準差、各層權重、各層平均年營運成本及標準差之母體資訊分別以前述樣本資料（ s , $w'_i = \frac{n'_i}{n'}$, $\bar{y}_i$, s_i ）取代。試求：（15分）

(1)決定分層雙重抽樣的最佳抽樣設計（Optimum double sample plan），亦即求算 n' 、 n_h 。

(2)就(1)的抽樣設計求算平均年營運成本估計量（以十萬元為單位）的變異數。

四、欲了解2022年國內汽車銷售概況，就2000家汽車銷售業者進行調查，汽車銷售業者分為兩大類：國產型(I)及進口型(II)，業者家數分別為 $N_1=1500$ 及 $N_2=500$ 。抽樣方法採用分層隨機抽樣，依類別分層，從每一層分別隨機抽出10家業者進行調查。假設2021年(x)各類業者的年平均銷售量已知為： $\bar{X}_1=220$ (輛)； $\bar{X}_2=140$ (輛)。

調查結果20家業者在2021年(x)及2022年(y)的銷售量統計如下：

層別 (h)	變數	樣本均數 $\bar{y}_i, \bar{x}_i$ (輛)	比率 ($\hat{R}_h$)	樣本共變異數 S_{xyh}	樣本標準差 S_h
I	y	240	1.2	7200	100
	x	200			80
II	y	180	1.8	2200	60
	x	100			40
合計	$S_y=110, \quad S_x=90, \quad S_{xy}=9000$				

(一)利用下列估計量估計年平均銷售量($\bar{Y}$)及該估計量的變異數:(15分)

(1) $\bar{y}_{st}$ ，分層隨機抽樣結合簡單均數估計量(mean per unit estimator)。

(2) $\bar{y}_{Rs}$ ，分層隨機抽樣結合分開比率估計量(separate ratio estimator)。

(3) $\bar{y}_{Rc}$ ，分層隨機抽樣結合混合比率估計量(combined ratio estimator)。

(二)求算估計量 $\bar{y}_{st}$ 、 $\bar{y}_{Rs}$ 、 $\bar{y}_{Rc}$ 對單位均數估計量($\bar{y}$)之相對效率(relative efficiency)，並說明那個估計量具有較佳精確度。(10分)