

4th edition

Biostatistics

第四版

生物統計學

楊惠齡、林明德 編著



Biostatistics

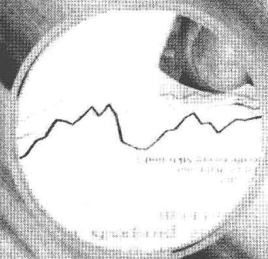
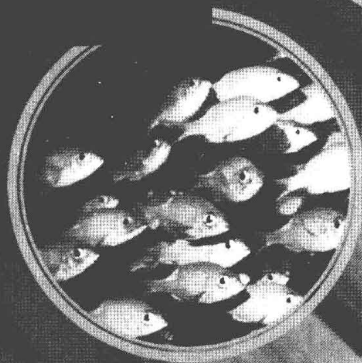
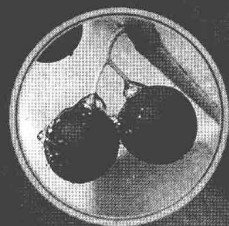
4th edition

Biostatistics

第四版

生物統計學

楊惠齡、林明德 編著



statistics

國家圖書館出版品預行編目資料

生物統計學 / 楊惠齡、林明德 編著. -- 初版.
-- 臺北縣中和市：新文京開發，民 94
面；公分

ISBN 986-150-176-2 (平裝)

1. 生物學 – 統計 2. 數理生物學
3. EXCEL (電腦程式)

360.13

93005578

生物統計學

(書號：B071e4)

編著者	楊惠齡 林明德
出版者	新文京開發出版股份有限公司
地址	台北縣中和市中山路二段 362 號 8 樓 (9 樓)
電話	(02) 2244-8188 (代表號)
F A X	(02) 2244-8189
郵撥	1958730-2
初版	中華民國 88 年 8 月 30 日
二版	中華民國 90 年 2 月 10 日
三版	中華民國 92 年 6 月 25 日
四版	中華民國 94 年 6 月 10 日

有著作權 不准翻印

建議售價：435 元

法律顧問：蕭雄淋律師

ISBN 986-150-176-2



第四版前言

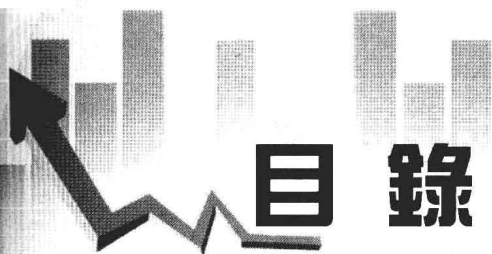
統計學是一門應用很廣且有用的學科，無論是商業、工業、或是生物醫學上的研究，只要是有關數字的量化研究，皆需要用統計方法來加以整理、分析，呈現研究的結果。一般人皆認為數學不好，統計也就學不好，其實不然，統計是以數學為基礎而加以應用的一門學科，是與數學有關係，但兩種學科之間的學習並不是絕對相關的。因此，在學習統計時，只需要有基本的數學計算能力及清晰的思考力即可，且在處理資料時，能夠做適當的整理及分析，則無論是何種研究的問題，都是可以迎刃而解的。

本書是為統計初學者而編寫之教材，以深入淺出的方式，對基本的統計方法及原理做一簡單的介紹，並藉由例題的說明，讓讀者對統計方法及統計公式的計算，有進一步的認識及瞭解。而統計軟體的使用，更是學習統計時不可或缺的工具，因此，本書使用普遍且操作簡易的 Excel 套裝軟體，來配合課本相關的內容，在方法的使用上加以介紹，希望讀者在研習統計方法及原理之餘，配合各章 Excel 套裝軟體的步驟說明，實際操作練習，從中比較其中異同，相信定能收事半功倍之效。

本書各章之後皆附有習題，供讀者演練之用，及解答以供研習參考，藉以提高學習上的效果。希望藉由此書的學習，讓初學者在統計學上有一基本的認識，在資料的分析、整理及統計方法的計算上有一初步的瞭解，對統計的學習不再有所畏懼。

楊惠齡、林明德 謹識

民國 94 年 4 月



目錄

Chapter 1 緒 論

1-1 什麼是生物統計學	2
1-2 資料的取得	3
1-3 資料之性質	10
1-4 EXCEL 與生物統計	13
習 題	23

Chapter 2 資料的整理

2-1 次數分配表的編製	26
2-2 相對及累積次數分配表的編製	30
2-3 統計圖	32
2-4 EXCEL 與統計圖	37
習 題	53

Chapter 3 敘述統計

3-1 算術平均數(Mean).....	56
3-2 中位數(Median).....	63
3-3 衆數(Mode).....	67

3-4	全距(Range).....	70
3-5	標準差(Standard Deviation).....	72
3-6	變異係數(Coefficient of Variation).....	75
3-7	相關係數(Coefficient of Correlation).....	78
3-8	EXCEL 與敘述統計	82
習 題		90

Chapter 4 ▽ 機率分配

4-1	二項分配(Binomial Distribution)	95
4-2	卜瓦松分配(Poisson's Distribution)	97
4-3	常態分配(Normal Distribution).....	100
4-4	t 分配(Student's t-distribution)	108
4-5	χ^2 分配(Chi-Square Distribution)	112
4-6	F 分配(F-Distribution)	115
4-7	EXCEL 與機率	118
習 題		130

Chapter 5 ▽ 抽樣分配

5-1	樣本平均數的抽樣分配.....	134
5-2	兩樣本平均數差的抽樣分配	137

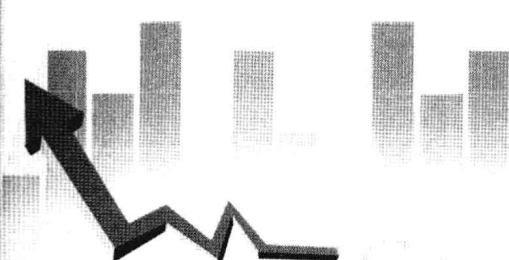
5-3	樣本比例的抽樣分配	139
5-4	兩樣本比例差的抽樣分配	141
習題		143

Chapter 6 ▽ 區間估計

6-1	母體平均數 μ 的區間估計	147
6-2	母體平均數差 $\mu_1 - \mu_2$ 的區間估計	154
6-3	母體比例的區間估計	167
6-4	兩母體比例差的區間估計	169
6-5	母體變異數 σ^2 的區間估計	171
6-6	兩母體變異數比 σ_1^2 / σ_2^2 的區間估計	174
6-7	EXCEL 與區間估計	177
習題		185

Chapter 7 ▽ 假設檢定

7-1	假設檢定的意義	188
7-2	母體平均數的假設檢定	191
7-3	兩母體平均數差的假設檢定	200
7-4	母體比例的假設檢定	208
7-5	兩母體比例差的假設檢定	211
7-6	母體變異數的假設檢定	215
7-7	兩母體變異數比的假設檢定	219



Contents

7-8 EXCEL 與假設檢定.....	222
習 題.....	235

Chapter 8 ▽ 次數分析

8-1 適合度檢定	238
8-2 獨立性檢定	241
習 題.....	247

Chapter 9 ▽ 變異數分析

9-1 變異數分析的意義	250
9-2 單因子變異數分析	251
9-3 二因子變異數分析	257
9-4 EXCEL 與變異數分析	265
習 題.....	271

附錄一 標準常態分配表	276
附錄二 t 分配表	277
附錄三 χ^2 分配表	278
附錄四 F 分配表	279
附錄五 習題解答	287

Chapter

1



緒 論



本章摘要

- 1-1 什麼是生物統計學
- 1-2 資料的取得
- 1-3 資料之性質
- 1-4 EXCEL 與生物統計





1-1 什麼是生物統計學

在電視上，看到女性瘦身的廣告，面對動人的敘述，瘦身前後的照片對照，相信很多女性，內心不禁動盪，真的有那麼好的效果嗎！但是偶而也在報章雜誌上，看到其負面的報導，到底這只是那幾個人的片面效果呢？還是它真的有整體性的效果呢？那麼事情的真相到底是什麼呢？

面對經濟的不景氣，常見的使用指標就是失業率，當我們看到媒體報導說近幾個月失業率有下降，所以我國的景氣有復甦的跡象。可是為何股票的指數仍然跌跌不休呢？失業率是否和經濟指數有關連呢？

像上面這些問題，都是平常我們在日常生活中可能會碰到或想到的問題，如何解決呢？統計學就是一個很好的工具。

那麼統計學是什麼呢？簡單的說，就是依據目的將觀察或測量到的資料，加以處理以及利用處理後的資料加以分析，以便做判斷及推論的一門學問。

以功能來分，統計學可分為二個主要部份，敘述統計(Descriptive Statistics)及推論統計(Inference Statistics)。就像每年夏天一到，就會聽到登革熱的病情再現，因此根據各地的醫療院所回報的資料，加以收集，整理，分析，解釋，指出六月份南部有 17 個案例，中部有 5 個案例，北部有 2 個案例，這就是敘述統計。而連續觀察 4 週之後，發現南部增加為 235 個案例，中部增為 201 個案例，北部增為 67 個案例，根據這些資料，利用一些理論及方法，做出疫情似乎有從南向北蔓延的趨勢，今年可能會爆發大流行，政府宜及早加以應對，這就是推論統計。

如果由學術的觀點來看統計學的發展，可分為理論統計和應用統計。理論統計係以數學的方法推演統計的原理，證明各種統計公式和定理，說明其來源，以理論研究為主，適合於數學基礎較好的理工科學生研究；而應用統計則著重在討論統計的方法，如何在實務上加以應用。由於應用的



方向不同，而有人口、政治、經濟、教育、商業、工程、生物、醫學、…等統計學，這也就是為什麼會有生物統計學產生的原因。

生物統計學(Biostatistics)係應用統計學的一支，著重在研究對象為有機物或有生命的東西（包括動植物、人類及昆蟲等）發生變化現象者。

然而雖有各科統計學之區分，其原理及方法均大同小異，其主要的目的還是在於以較有效的方法用少數資料去推測一般事物的真相。

很多人一談到統計學就認為那是一門難度很高不易學習的學科，其實不然，統計學的方法是簡單的數學公式的應用，對於應用統計學，只要有高中代數的程度就綽綽有餘，甚至只要將蒐集或試驗所得的資料，以簡單數學公式處理後，解釋資料的性質即可。只不過在碰到問題時，要如何根據試驗的資料性質及目的，慎選適當的統計方法，來導出正確的結論，則是一門學問。首先，應該要瞭解統計學專有名詞的含意及概念，其次，要瞭解各種統計方法的功用，以避免選用不適當的統計方法。

1-2 資料的取得



一般說來，統計可說是利用少數的資料對有興趣的母體做推論的一種有效方法，而此處所謂的**母體(population)**或稱為群體，可說是研究者所欲研究事物對象（數值，人員，測量等）的全體，而從母體內取出的部份個體就稱為**樣本(sample)**。

舉例來說，有一陣子，幼兒感染手口足病非常嚴重，當我們想知道台中縣所有幼稚園學童感染的情況，以便對是否需要停課作一決定，因此就抽了 10 所幼稚園學童調查其平均感染率。在此母體就是台中縣所有幼稚園學童，而樣本就是被抽出的 10 所幼稚園學童。

接下來的問題是會不會因為隨便的抽樣而導致資料不具有客觀性、代表性，而影響正確的判斷，因此，再來要談的**抽樣**(sampling)則是利用適切的方法，從母體中抽出一部份樣本，作為觀察的對象。抽樣之方法可分為**簡單隨機抽樣**，**系統隨機抽樣**，**分層隨機抽樣**和**集群隨機抽樣**。

1-2-1 簡單隨機抽樣(Simple Random Sampling)

從母體中抽樣時，每一個個體都有公平的機會被抽到，這就是簡單隨機抽樣。在實際操作時，如果母體不大，譬如一個班上有 50 位同學我們要從中抽出 5 位，擔任畢業旅行的執行委員，這時可以做 50 支籤丟到籤箱中，充分攪拌後，再抽出 5 支籤即可。可是如果母體很大，就像統一發票開獎，事實上不太可能對每一張發票做一張籤，於是就用搖獎機，以號碼球，搖出中獎號碼，仍不失為一公平的辦法。

可是如果沒有搖獎機的時候怎麼辦呢？這時就可以考慮使用隨機號碼表。隨機號碼表是利用搖獎機的原理，將 0 到 9 的數字一次又一次不斷的搖出，將這些數字連在一起就成隨機號碼表。表內每一個數，0 到 9 出現的機會都一樣，整個表內 0 到 9 的數字出現的頻率也差不多。因為每個數字出現的機會都是相同的，因此隨機號碼表不論橫的、直的、斜的甚至倒的用都可以。

表 1-1 隨機號碼表

10	09	73	25	33	76	52	01	35	86	34	67	35	48	76	80	95	90	91	17	39	29	27	49	45
37	54	20	48	05	64	89	47	42	96	24	80	52	40	37	20	63	61	04	02	00	82	29	16	65
08	42	26	89	53	19	64	50	93	03	23	20	90	25	60	15	95	33	47	64	35	08	03	36	59
99	01	90	25	29	09	37	67	07	15	38	31	13	11	65	88	67	67	43	97	04	43	62	76	06
12	80	79	99	70	80	15	73	61	47	64	03	23	66	53	98	95	11	68	77	12	17	17	68	33
66	06	57	47	17	34	07	27	68	50	36	69	73	61	70	65	81	33	98	85	11	19	92	91	70
31	06	01	08	05	45	57	18	24	06	35	30	34	26	14	86	79	90	74	39	23	40	30	97	32
85	26	97	76	02	02	05	16	56	92	68	66	57	48	18	73	05	38	52	47	18	62	38	85	79
63	57	33	21	35	05	32	54	70	48	90	55	35	75	48	28	46	82	87	09	83	49	12	56	24
73	79	64	57	53	02	52	96	47	78	35	80	83	42	82	60	93	52	03	44	35	27	38	84	35
98	52	01	77	67	14	90	56	86	07	22	10	94	05	58	60	97	09	34	33	50	50	07	39	98
11	80	50	54	31	39	80	82	77	32	50	72	56	82	48	29	40	52	42	01	52	77	56	78	51
83	45	29	96	34	06	28	89	80	83	13	74	67	00	78	18	47	54	06	10	68	71	17	78	17
88	68	54	02	00	86	50	75	84	01	36	76	66	79	51	90	36	47	64	93	29	60	91	10	62
99	59	46	73	48	87	51	76	49	69	91	82	60	89	28	93	78	56	13	68	23	47	83	41	13
65	48	11	76	74	17	46	85	09	50	58	04	77	69	74	73	03	95	71	86	40	21	81	65	44
80	12	43	56	35	17	72	70	80	15	45	31	82	23	74	21	11	57	82	53	14	38	55	37	63
74	35	09	98	17	77	40	27	72	14	43	23	60	02	10	45	52	16	42	37	96	28	05	26	55
69	91	62	68	03	66	25	22	91	48	36	93	68	72	03	76	62	11	39	90	94	40	05	64	18
09	89	32	05	05	14	22	56	85	14	46	42	75	67	88	96	29	77	88	22	54	38	21	45	98
91	49	91	45	23	68	47	92	76	86	46	16	28	35	54	94	75	08	99	23	37	08	92	00	48
30	38	69	45	98	26	94	03	68	58	70	29	73	41	35	53	14	03	33	40	42	05	08	23	41
44	10	48	19	49	85	15	74	79	54	32	97	92	65	75	57	60	04	08	81	22	22	20	64	13
12	55	07	37	42	11	10	00	20	40	12	86	07	46	97	96	64	48	94	39	28	70	72	58	15
63	60	64	93	29	16	50	53	44	84	40	21	95	25	63	43	65	17	70	82	07	20	73	17	90
61	19	69	04	46	26	45	74	77	74	51	92	43	37	29	65	39	45	95	93	42	58	26	05	27
15	47	44	52	66	95	27	07	99	53	59	36	78	38	48	82	39	61	01	18	33	21	15	94	66
94	55	72	85	73	67	89	75	43	87	54	62	24	44	31	91	19	04	25	92	92	92	74	59	73
42	48	11	62	13	67	34	40	87	21	16	86	84	87	67	03	07	11	20	59	25	70	14	66	70
23	52	37	83	17	73	20	88	98	37	68	93	59	14	16	26	25	22	96	63	05	52	28	25	62
04	49	35	24	94	75	24	63	38	24	45	86	25	10	25	61	96	27	93	35	65	33	71	24	72
00	54	99	76	54	64	05	18	81	59	96	11	96	38	96	54	69	28	23	91	23	28	72	95	29
35	96	31	53	07	26	89	80	93	54	33	35	13	54	62	77	97	45	00	24	90	10	33	93	33
59	80	80	83	91	45	42	72	68	43	83	60	94	97	00	13	02	12	48	92	78	56	52	01	06
46	05	88	52	36	01	39	09	22	86	77	28	14	40	77	93	91	08	36	47	70	61	74	29	41
32	17	90	05	97	87	37	92	52	41	05	56	70	70	07	86	74	31	71	57	85	39	41	18	38
69	23	46	14	06	20	11	74	52	04	15	95	66	00	00	18	74	39	24	23	97	11	89	63	38
19	56	54	14	30	01	75	87	53	79	40	41	92	15	85	66	67	43	68	06	84	96	28	52	07
45	15	51	49	38	19	47	60	72	46	43	66	79	45	43	59	04	79	00	33	20	82	66	95	41
94	86	43	19	94	36	16	81	08	51	34	88	88	15	53	01	54	03	54	56	05	01	45	11	76
98	08	62	48	26	45	24	02	84	04	44	99	90	88	96	39	09	47	34	07	35	44	13	18	80
33	18	51	62	32	41	94	15	09	49	89	43	54	85	81	88	69	54	19	94	37	54	87	30	43
80	95	10	04	06	96	38	27	07	74	20	15	12	33	87	25	01	62	52	98	94	62	46	11	71
79	75	24	91	40	71	96	12	82	96	69	86	10	25	91	74	85	22	05	39	00	38	75	95	79
18	63	33	25	37	98	14	50	65	71	31	01	02	46	74	05	45	56	14	27	77	93	89	19	36
74	02	94	39	02	77	55	73	22	70	97	79	01	71	19	52	52	75	80	21	80	81	45	17	48
54	17	84	56	11	80	99	33	71	43	05	33	51	29	69	56	12	71	92	55	36	04	09	03	24
11	66	44	98	83	52	07	98	48	27	59	38	17	15	39	09	97	33	34	40	88	46	12	33	56
48	32	47	79	28	31	24	96	47	10	02	29	53	68	70	32	30	75	75	46	15	02	00	99	94
69	07	49	41	38	87	63	79	19	76	35	58	40	44	01	10	51	82	16	15	01	84	87	69	38

* Reproduced from Table A-1 of Wifred J. Dixon and Frank J. Massey, Jr., Introduction to Statistical Analysis, 2nd edition, McGraw-Hill Book Co., New York, 1957.

假設有一醫院要對去年曾經住院的 500 名病患，從中抽出 10 位做家庭訪問，想得知其對醫院服務品質的看法。此時在利用隨機號碼表之前，必須先將病患加以編號 1~500，因為每一編號皆可視為一個 3 位數，因此由隨機號碼表中任意找三列或三行，逐一找出比 500 小的或一樣的 10 個三位數，即為所要的抽樣號碼。譬如說從表 1-1 中之第 2 列，第 5 行開始，找到數字為 2，用直尺以 3 位數為準，垂直劃二條線，由上而下比對（大於 500 的放棄），直至額滿為止。

	5	6	7	8 行		
2 列	2	0	4	8	→	204(✓) (1)
3 列	2	6	8	9	→	268(✓) (2)
4 列	9	0	2	5	→	902(×)
5 列	7	9	9	9	→	799(×)
6 列	5	7	4	7	→	574(×)
7 列	0	1	0	8	→	010(✓) (3)
8 列	9	7	7	6	→	977(×)
9 列	3	3	2	1	→	332(✓) (4)
10 列	6	4	5	7	→	645(×)
11 列	0	1	7	7	→	017(✓) (5)
12 列	5	0	5	4	→	505(×)
13 列	2	9	9	6	→	299(✓) (6)
14 列	5	4	0	2	→	540(×)
15 列	4	6	7	3	→	467(✓) (7)
16 列	1	1	7	6	→	117(✓) (8)
17 列	4	3	5	6	→	435(✓) (9)
18 列	0	9	9	8	→	099(✓) (10)

因此決定對編號 204，268，10，332，17，299，467，117，435，99 等 10 位病人，進行家庭訪視，如果要抽樣的數目太多，垂直的三行不夠用時，則轉到右邊的三行，以此類推，周而復始，即可得到任何數目的隨機號碼。



不過，萬一作田野調查時，身上沒帶隨機號碼表時怎麼辦呢？隨機號碼還是可由其他方法來獲得，例如我們要從 18 個地點，抽出 4 個出來，這時可用經過車輛之牌照號碼後 4 碼，譬如 0853，3167，1607，7249。

$$0853 \div 18 = 47 \text{ 餘 } 12, 12+1=13$$

$$3167 \div 18 = 175 \text{ 餘 } 17, 17+1=18$$

$$1607 \div 18 = 89 \text{ 餘 } 5, 5+1=6$$

$$7249 \div 18 = 402 \text{ 餘 } 13, 13+1=14$$

由於除以 18 之後的餘數為 0 至 17，而地點編號為 1 至 18，所以將餘數各加 1，最後就可決定對 6，13，14，18 四個地點取樣，或者考慮使用千元大鈔上的號碼也可。

『隨機』取得樣本在生物學的研究中尤其重要，例如想以草莓醬製成的誘餌陷阱，捕捉昆蟲來測量其體重時，我們應該想到通常誘捕到的昆蟲多為飢餓的昆蟲，而飢餓的昆蟲體重往往較輕，換言之，此例中飢餓而體重較輕的昆蟲有較大的機會自母體中被選取出來，此即不符「母體中每個樣本單位都有相同被選取機會」的隨機抽樣原則，由此估算出的體重也就無法代表整個母體。而有時抽樣所產生的誤差根本無法自知，但是當我們懷疑抽樣可能有偏差時，我們就應該特別注意，並於結果說明時，將之考慮在內。

1-2-2 系統隨機抽樣(Systematic Random Sampling)

系統隨機抽樣又稱等距抽樣，也就是規則的從母體中，每間隔一定的距離抽取一個樣本。如果有一班級總共有 60 名學生，想從其中抽出 6 名擔任公差，系統抽樣法則先計算抽樣區間的長度，即 $60/6=10$ ，再以簡單隨機抽樣，由 1 到 10 中抽一個數，假設為 2，則 2、12、22、32、42、52 等 6 名學生即為公差。

使用此法的優點有三：

- (1) 可節省編製名冊及抽取號碼的手續，此外系統抽樣法也可用相同的間隔、時間、距離、空間作為抽樣的標準。例如飲料工廠的生產線上，品管人員常每隔一定的數目抽出一瓶，測量其容量是否合乎標準，因為機器生產的速度是固定的，也可每隔一段時間來抽測。調查河水的深度每一百公尺測量一次是用相同的距離。調查都市內土地的利用情形，若把土地分成若干相等的小塊，每隔 5 塊調查一塊，則是用相同的空間作抽樣的標準。
- (2) 使抽出的樣本單位普遍出現於母體各部份，而不過分集中。雖然簡單隨機抽樣法可使母體的各單位有相等的機會出現，可是由於機會的變化性（就像一顆骰子六面朝上的機會是相等的，可是丟 12 次之後，不見得 1~6 的數字都是出現各 2 次），樣本單位的分佈常有集中而不普遍的現象，例如自全省抽出數戶作樣本，用簡單隨機抽樣法，常會發生若干鄉鎮沒有樣本或樣本過於集中少數地區的情形。用系統抽樣法就可以避免這種現象，而使樣本均勻的散佈於各鄉鎮，以增加樣本的代表性。
- (3) 若事先把母體各單位按一定的層次排列，則系統抽樣法實在具有分層抽樣法的效果。

1-2-3 分層隨機抽樣(Stratified Random Sampling)

如果個體在母體中分布並不平均，我們可以先把性質類似的個體歸類在一起，稱為“層”，然後在每一層中，依簡單隨機抽樣法，抽出需要的樣本數，為什麼需要這麼麻煩呢？假設學校有三個系的學生修生物統計學，甲系有 60 名，乙系有 120 名，丙系有 180 名。現在欲從中抽樣 30 名來調查其反應，如果依照前述的簡單隨機抽樣法，先把學生編號 1~360，再從中抽出 30 名，萬一結果是甲系 8 名，乙系 13 名，丙系 11 名，如此一來，丙系學生的意見所佔的比例似乎和其原來人數比例不太相稱，因此



為了避免簡單隨機抽樣的樣本發生過分集中於某種特性，或缺乏某種特性的現象時，就需用到分層隨機抽樣了。

此時抽樣的方式可改為：

$$\text{甲系抽} 30 \times \frac{60}{360} = 5 \text{名}$$

$$\text{乙系抽} 30 \times \frac{120}{360} = 10 \text{名}$$

$$\text{甲系抽} 30 \times \frac{180}{360} = 15 \text{名}$$

這種以各層所佔的比例，來決定抽樣個數的方法又稱為比例抽樣法。

採用此法之理由有四：

- (1) 如母體之某些部份所要求之準確度已知時，則將各層視為獨立的母體，來處理較為有利。
- (2) 行政上的方便，各層分人負責，不但費用可減少，且準確度亦可提高。
- (3) 在母體內不同部份，抽樣的問題，可能有顯著的差異，則分層可方便做適當的調整，應用各種可行的方法來處理。
- (4) 分層通常可使樣本推算值之差異減小，亦即可使整個母體特徵的推論值的精密度提高。

1-2-4 集群隨機抽樣(Cluster Random Sampling)

集群隨機抽樣是將母體按某種標準分成若干**集群**(cluster)，然後在所有的集群中，隨機抽出數個集群，並對被抽到的集群作全面調查。例如教育部欲對全國中學生做升學調查，此時學校可視為集群（因為學校可看成是母體的縮影），則抽出幾個學校之後加以全部調查，而不必長途跋涉到每個學校去抽樣，可以節省更多的時間、人力。