



森林統計分析

譚 啟 棟 編 著

(一九五二年初版)

立信會計圖書用品社出版



森林統計分析

譚啓棟編著

立信會計圖書用品社出版

書號 705206

森林統計分析

編著者 譚 啓 棟

出版者 立信會計圖書用品社

上海河南中路三三九號

北京西琉璃廠八二號

重慶小什字立信大樓

天津建設路一號

總發行所 中國科技圖書聯合發行所

上海中央路二四號三〇四室

印刷者 中 和 印 刷 廠

上海淮安路七二七弄三〇號

引言 1—2

正文 1—214

目錄 1—2

插頁 162—163 (滬)

1952年9月初版(務本型)

0001—1000 (中和印)

1952年10月再版

1001—2000 (同和訂)

定價 ㄆ 20,000

引 言

自解放以來，統計學書籍出版很少，細究起來，不外下列幾種原因：

一、過去寫一本書是向自己負責。不管觀點如何、立論如何、影響如何，那都是作者自己個人的事。現在呢，則是向人民負責。不是僅僅個人得失問題，而是人民的得失問題，所以寫書的人都很仔細審慎起來了。

二、我們過去研究統計的同志們，差不多全是學的資本主義的那一套。無論在立場方面，立論觀點方面，都有根本上的錯誤。也就是說我們需要時間大力的改造自己。在這學習期間也不能寫出啥東西出來。

三、蘇聯在1949年前，有一部份研究統計理論的先生們，走上了資本主義統計學者的迷途，當然他們是被批判了，為資本主義服務及形式邏輯的統計理論被清算了。偉大的社會主義十月革命的勝利，蘇聯已建立起為人民服務的統計學。雖然東北統計局有了很多新統計知識的介紹，但很有系統的東西仍然不夠多。這也是統計學書籍出版少的一個主要原因。

事實雖然如此，但在建設新中國的時代，統計人員的訓練是不能等待的。也就是說我們不可以“停工待料”地來等待蘇聯統計學更發展後再編寫統計教本。在這過渡期間我們應該編些書出來以應現時需要，這就是編這本書的動機。

編這本書有三種主要精神，把它寫在下面，是否正確，請批評。

1. 儘量吸取新的 關於認識及分析方法上，儘量吸取蘇聯的先進理論，主要的是為人民國家經濟建設服務，不是為資本家服務，我們重視實際問題的研究，不注重形式的研究。

2. 批判地接受舊的 舊的統計方法，有部份是可以採用的，不過我們要用新的立場——無產階級立場——及新的哲學方法批判的接受

它,而不是毫無保留的生吞活剝的搬過來就完事。

3. 理論要與實際結合 我並不希望這本書成一本萬妙靈藥,只要求它在森林學院近一二年中有它一定之效用就够了。我不贊成教同學們“知其然不知其所以然”的讀書方法,現在同學們也不希望這樣的讀書方法,所以我配合森林系的需要,將統計分析對森林學的應用,和計算公式(一部份未證明)寫出來。

這本教材作者在東北農學院試教過五班同學,每週三小時講演、三小時實習,可在一學期講完,特寫出供採用者參考。

這本書的綱要,經東北農學院數理室同志集體討論過三次,教材內容經李世達同志試教一次,經何澤洪及李世達兩同志校閱一遍,併此致謝。

作者的政治水平不够,若有差誤,請同志們提出指正及批評,我們大家要為中國人民的統計開闢一條大道,這是作者誠懇希望的。

譚 啓 棟

1952年1月於東北農學院

目 錄

引 言

第 一 章 統計資料之整理..... 1

第一節 統計分組.....1

第二節 統計表的編製.....8

第三節 統計圖的製作..... 10

討論題及練習題

第 二 章 代表數..... 23

第一節 算術平均數..... 23

第二節 中位數..... 29

第三節 衆數..... 32

第四節 幾何平均數..... 34

第五節 調和平均數..... 36

第六節 平方平均數及立方平均數..... 37

第七節 進展平均數及退縮平均數..... 39

第八節 代表數應用之討論..... 41

討論題及練習題

第 三 章 離差及偏態..... 48

第一節 離差..... 48

第二節 偏態..... 58

討論題及練習題

第 四 章 相對數及指數..... 65

第一節 相對數..... 65

第二節 指數..... 69

討論題及練習題

第 五 章 機率及機率常態曲線..... 87

第一節 大數法則.....	87
第二節 排列及組合.....	89
第三節 機率.....	91
第四節 二項分配之統計量數.....	97
第五節 常態曲線.....	100
第六節 常態曲線之配合.....	107
第七節 配合之適度測驗.....	114
討論題及練習題	
第 六 章 配合方程式的求法.....	119
第一節 最小二乘法的理論根據.....	120
第二節 直線配合方程式之求法.....	122
第三節 二次以上代數函數之配合法.....	127
第四節 對數曲線及多變數函數之配合.....	131
第五節 正則式經驗求法及解法.....	139
第六節 本章之討論及總結.....	146
討論題及練習題	
第 七 章 相關量數.....	153
第一節 直線相關.....	154
第二節 非直線相關.....	164
第三節 複相關.....	172
第四節 偏相關.....	180
討論題及練習題	
第 八 章 抽樣問題.....	188
第一節 抽樣方法.....	189
第二節 均方誤.....	193
第三節 差異顯著性測定.....	202
討論題及練習題	
主要參考書	

第一章 統計資料之整理

第一節 統計分組

統計工作的第一個步驟，是統計資料的蒐集問題，就是統計事物的觀察或調查。因為要求目的、經費、人力及觀察事物的不同，調查方法也有不同。若以調查的範圍來分，大別為兩類：

一、普查 就是全面觀察。

二、部份調查 就是非全面觀察。

統計資料蒐集後的工作就是整理統計資料，但是由這堆大量複雜的資料中，怎樣去得到我們所需要的結果呢？我們知道統計的目的是由統計分析得到的。那麼這一堆大量的資料怎樣去着手分析呢？在分析的前一步工作就是合理的整理這一堆資料。整理資料的第一步工作，也是最重要的工作，就是資料的質量分組，在資本主義國家的資料分組，只是機械的等距分組，因為等距分組在數學計算上是便利的。他們忽視了質的重要，他們忽視統計資料在實際目的上的應用，同時他們利用分組方法來混淆真象，利用分組方法說明有利於他們資本主義的偽造道理。所以分組問題在統計工作中是一件很重要的工作。若是分組錯誤了，不但不能完成所需要的結果，有的時候並且隱混真實的結果，所以統計分組不但是很重要的工作，也是最基本的工作。

分組的寫法 如調查一塊林區中的 216 棵樹，若在我們實際要求上是等距分組，組距為 3 公分。經整理之結果如下表 1。

在表之直徑分組這一系列就是分組的寫法，如 13.5—16.5 就是表示：在這一組樹的直徑是包括 13.5 公分及 13.5 公分以上不及 16.5 公分。也就是說若是一棵 13.5 公分的樹就屬於這一組；若直徑是 16.5 公分的樹就屬於下一組（即 16.5—19.5）。組距中點就是這一組的代表數。如 25.5

(表 1) 216 棵樹直徑分組

直徑分組(公分)	中點(公分)	樹 棵 數
13.5—16.5	15	1
16.5—19.5	18	3
19.5—22.5	21	15
22.5—25.5	24	33
25.5—28.5	27	61
28.5—31.5	30	41
31.5—34.5	33	26
34.5—37.5	36	22
37.5—40.5	39	5
40.5—43.5	42	3
43.5—46.5	45	1

資料來源：河田杰著森林簡易統計計算法

—28.5 組之代表直徑是 27，因為 27 是這一組的中點 $\left(\frac{28.5+25.5}{2}=27\right)$ 。第三列的樹棵數，就是說在一個直徑分組中有幾棵樹。如 19.5—22.5 組中有樹 15 棵，凡是直徑在 19.5 及 19.5 以上而不到 22.5 的樹皆納入在這一組裏。這 15 棵樹直徑和可以說是 $15 \times 21 = 315$ 公分。若把這 15 棵樹直徑實際值相加，絕不與 315 公分完全相等，因為這 15 棵樹的實際直徑值，並不皆是 21 公分。可是扯長補短的計算，相差是非常有限的。這一點說明統計並不是數學，是存在着實際的差誤的，不過數量愈多差誤亦愈小。

大量資料的歸組方法 當歸組的組距及本質的分組項目決定了後，即將大量資料登記在各組距或項目中。若資料數量很大，整理資料到編製統計表，不是件簡單的工作。如大量的人口普查卡片，必需利用機器來分組及歸類。而通常用的則是卡片分類、劃記分類及登記表分類。卡片分類的方法，第一步將調查資料審查後填入卡片，加以校對，再依照所決定的組距或類別將卡片分開，從卡片上計數就得到了統計分

組的統計表了。爲什麼要填卡片呢？因原始調查資料是要永久保存，以備查考的，但每一種資料因目的不同，需要編製數張或數十張統計表，這是不能用原始資料分組的第一個理由。有些調查表包括項目很多，項目混雜亦不可能用原始資料分組。如人口調查，包括得很廣闊，若以戶爲調查單位，則整理一戶的人口材料非填寫戶口卡，是不可能整理的。在學校中及小型調查材料之整理多利用劃記法，第一步將決定的分組表劃好，再依照統計資料劃記號，待統計資料完畢後綜合結果，即得所需之統計表。登記法暫不介紹。

例如 1951 年春，在哈爾濱張家油房所作的農戶人口調查，我們欲瞭解每戶人口分佈情形，將人口數及戶數以劃記分類法作表如下：

(表 2) 哈爾濱市張家油房村民主屯人口分佈

人 口 數	戶 數 劃 記	戶 數
1	正下	8
2	正正下	14
3	正正正正下	24
4	正正正正正正	30
5	正正正正正正下	27
6	正正正正正正	24
7	正正正	14
8	正正	9
9	正一	6
10	一	1
11	丁	2
12	丁	2
共 計		161

劃的方法就是，根據資料實際情形在所在組內劃一記號。如張三家爲五口人，則五口人這一組內劃“一”。李四家爲八口人，則在八口人這一組內劃“一”。將所有資料劃記號完畢後，來數“正”字的多少及最後掛零數即得各組之結果。如五口人的一組有五個“正”字另二劃，就是說共計 27 戶。

屬性分組 屬性分組就是質量的分組，在資本主義國家是被忽視的。解放後為瞭解勞動人民在未土改前被剝奪學習文化的情況。在1951年作了一個張家油房的農村人口調查，結果分析如下表：

(表 3) 民主屯不識字人口數及百分比

階 級 成 份	人 口 數	不識字人口數	不識字人口佔 人口百分比
地 主	78	47	60.7
富 農	70	49	70.0
中 農	126	94	74.6
貧 農	329	252	76.6
僱 農	179	146	81.6

這張表說明了什麼？說明了貧、僱農被剝奪了學習文化的權利。在實際資料中，地主不識字的多是婦女，而貧僱農識字的全是學齡兒童。因為這一調查是1951年舉行的，也就是說貧僱農得到學習文化的機會，是解放以後的事。

若作樹木調查，按木材應用分組來說，應以主要樹種分類。因為各種樹木硬度不同，荷載力量不同，應用也是不同的，這就是在應用上，對質的分組的例子。

量的等距分組及不等距分組 統計資料怎樣整理歸組已如上節說明。為什麼要歸組呢？因為一堆差異很大、質量全異的數字，是表現不出來什麼的，也無從研究分析。如前列舉 216 棵樹之例，它們的直徑幾乎完全不同，若不分組連計算平均數都是不容易的，所以按照數量分佈情形分成三公分的等距分組，列成統計表，那麼作進一步的分析也就容易了。怎樣分組呢？只有一條道理，就是將調查所得資料，分成質量上相同的各組。這是與資本主義國家分組只注重數量上的分佈情形，甚或拿它的標準差（參閱第三章）及總量來決定分組而不考慮質的問題，完全不同的。比如在資本主義國家企業資本金的分組，他們將小資產階級及資產階級混淆了，他們故意矇蔽了剝削工人的事實；在反動政府時代，他們把土地所有權的分組，只作形式上的等距分組，這樣他們混淆了

貧、僱農與中農的區劃，甚或混淆了地主與富農的區劃，這樣他們就矇蔽一切階級的剝削與被剝削關係。

等距分組是最簡單而最易計算的一種方法。雖然它是最簡便的，可是不可違背“在同組中質量完全相同”的原則。若在同組中質量不同，這一組就錯誤了，所分析的結果也是錯誤的，這一點必須注意。216 棵樹直徑分組，是以三公分等距的分組，這必定在需要目的本質上，只研究 216 棵樹粗細分佈情形方可這樣做。若因用途不同，如某一尺寸至某一尺寸可用做枕木；某一尺寸至某一尺寸可用做房屋建築，這種目的的情形下，這種分組就錯誤了，就必需按照規定尺寸分組，纔是質量完全相同。

非等距分組 分組是依質量相同情形而劃分的，分組若只着重表面質的問題，而忽略了量對質的影響也是不對的。

(表 4) 1908—1911 年期間企業主與職工死亡率
(100 人中死亡數)

年 齡 (歲)	企業主死亡率	職工死亡率
18—24	3.76	3.97
25—34	3.70	4.36
35—44	5.06	5.59
45—54	9.83	10.84
55—64	23.68	26.43
65—	75.81	82.61
按合計計算	8.81	6.84

資料來源：東北統計局譯，新統計學概論，上冊第 17 表。

在上表中若只看表面的社會成份劃分，就是只看企業主及職工的成份劃分，而忽視了年齡對死亡率的影響，也就得到相反的結果。若只按成份劃分所得到合計的結果，企業主死亡率大於職工死亡率。我們仔細研究就瞭解了一個問題，因為年老者死亡率大，年壯者死亡率小，而企業主年齡大的多，勞動者年齡大的比重小。若按年齡再分組即得真實結果，從各個年齡分組中，很顯然的，職工死亡率比企業主為高。

連續數列及不連續數列 年齡分組是不連續數列，因年齡的規定

是滿二十歲而不到二十一歲者以二十歲計，沒有 20.3 或 20.5 歲的。也就是說，只有二十及二十一歲，中間是不連續的。像這樣的分組如 18—24 就是從十八歲到二十四歲，組內包括十八歲及二十四歲。如 216 棵樹直徑的數列，就是連續數列，因為二十公分至二十一公分間，仍有很多數字。樹的直徑不只是二十及二十一公分，可以是 20.67 甚或 20.999 亦可。

非等距分組方法，是依據同質量情形下劃分的，如我們在 1950 年大汎河村的調查，將農村參加生產人口的年齡分組如下表：

(表 5) 大汎河村參加生產農民年齡分組

成份別	10—16		17—50		50—	
	男	女	男	女	男	女
地主	5	2	—	—	—	—
富農	4	5	2	1	—	—
中農	12	12	3	1	—	—
貧農	12	6	1	—	—	—
雇農	6	5	3	1	—	—

為什麼這樣分組呢？因為我們瞭解 10—16 歲為半個勞動力；17—50 歲組為全勞動力；50 歲以上以半勞動計算。為調查便利起見，這裏年齡是指的虛歲數。在這種分組中，第一組為 7 歲的組距；第二組組距為 34 歲；最後一組是不定的。雖各組組距不同，但以勞動力來說是相同的。第一組為半勞動力；第二組為全勞動力；第三組為半勞動力。也就是說在各組中質量是相同的。

二次分組 二次分組是列寧利用來揭露資本主義國家及反動政府朦蔽事實的一種分組方法。列寧曾把帝俄於 1897 年時代人口調查按職業區分的表，重新歸組了一次。把一個機械形式無意義、羅列六十五種職業類別的表，劃分成生產、半生產、不生產人口三大類別的表。由這表中知道多少人是社會中的寄生者，多少人是勞動者。

列寧再分組的表是這樣的：

(表 6) 1907 年全俄人口職業分類

單位：百萬人(男女合計)

職 業 類 別	共 計	從業人數	家庭人口
非生產人口	6.9	4.1	2.3
官吏及軍人	2.2	1.5	0.7
僧侶及自由職業者	1.6	0.7	0.9
靠利息及儲金生活者	2.2	1.3	0.9
被剝削自由者、娼妓、無業及職業不明	0.9	0.6	0.3
半生產人口	12.7	5.7	7.0
商業	5.0	1.6	3.4
運輸及通信	1.9	0.7	1.2
私營企業機關職員、僕役、日工	5.8	3.4	2.4
生 產 人 口	106.0	23.4	82.6
農業	93.7	18.2	75.5
工業	12.3	5.2	7.1
總 計	125.6	33.2	92.4

列寧把反動政府混淆剝削階級及被剝削階級的事實利用他們自己的統計資料給揭明了，這是二次分組最大的功用。在資本主義國家裏常利用模糊的分組來製造對於他們有利的統計數字。可是列寧則用他們自己的資料揭明了被隱蔽的事實。美國統計學家們爲了說明他們偽造「小農場集約性」的理論，他們把無有土地出賣勞動力的貧、雇農也劃分在土地少的資本家一組內，其他量的分組也沒表示有何質量的相同點。他們的原始統計表是這樣的：

(表 7) 1900 年 550 萬個美國農場調查

農場面積(英畝)	單位土地面積指標 (美元/英畝)			
	僱傭勞動支出	肥 料 費	家 畜 價 值	農具及機械價值
0— 3	40.30	2.36	456.76	27.57
3— 10	2.95	0.60	16.32	6.71
10— 20	1.12	0.33	8.30	2.95
20— 50	0.55	0.20	5.21	1.65
50— 100	0.46	0.12	4.51	1.47
100— 175	0.45	0.07	4.09	1.14
175— 260	0.52	0.07	3.96	1.00
260— 500	0.48	0.07	3.61	0.77
500— 1000	0.47	0.03	3.16	0.57
1000—	0.25	0.02	2.15	0.29

材料來源：上表與本表皆根據東北統計局譯，新統計學概論，表四及表三十八。

上表正說明他們所需求的“事實”，[小農場經營比較大農場集約些]。這是矇蔽真實的偽造道理，想藉此以反駁農場工業化的真理。因為在這份統計表的數字上，表示着小農場具有較多的家畜及農具、機械、肥料等。可是在僱傭勞動中就暴露出來了資本家更多剝削關係的事實，也就是資本家剝削僱傭勞動的典型實例。列寧爲了揭露被矇蔽的真理，他用二次分組的方法重新整理了資料，廢棄了極端形式的土地大小分組，根據經濟理論反映了農場的經濟力。他的二次分組結果如下表：

(表 8) 農場單位面積指標 (美元/英畝)

農場分組 (按生產物價值 (美元))	僱傭勞動支出	肥 料 費	家 畜 價 值	農 具 及 機 械 價 值
0	0.08	0.01	2.97	1.19
1— 50	0.06	0.01	1.78	0.38
50— 100	0.03	0.03	2.01	0.48
100— 250	0.11	0.05	2.46	0.62
250— 500	0.19	0.07	3.00	0.82
500—1000	0.36	0.07	3.75	1.07
1000—2500	0.67	0.08	4.63	1.21
2500—	0.72	0.06	3.98	0.72

材料來源：根據東北統計局譯，新統計學概論，第 39 表編製。

列寧二次分組的統計表說明了，農業的集約性是隨農場生產物之價值的增加而增大的。所以這二個例子說明了一件事實，同一資料由於分組方法不同，所得出的結果完全相反。因此在統計分析中，對於資料質量分析的分組方法是重要的，也是最基本的。

第二節 統計表的編製

統計表 統計表是根據研究目的，整理統計資料所得之結果。同時，統計表是研究分析的依據材料，項目必需清楚，不能夠含混不明。

統計表本體分爲三部份，左方第一項稱爲縱表頭，如下表之原成份別及以下這一行就是表之縱表頭，是爲統計表第一部份。表之左上方乃

縱表頭之標題，不要在中間劃斜線。

統計表第二部份是橫表頭，如下表之騾、馬、驢、牛、戶數、共計等。

(表9) 張家油房各農戶之耕畜數

時期：1950年12月

單位：頭

原成份別	戶數	共 計		騾		馬		驢		牛	
		可供役使	不可役使	可供役使	不可役使	可供役使	不可役使	可供役使	不可役使	可供役使	不可役使
總計	32	27	4	2	—	23	4	—	—	2	—
地 主	4	2	1	1	—	1	1	—	—	—	—
富 農	3	3	1	—	—	3	1	—	—	—	—
中 農	8	8.5	1	1	—	5.5	1	—	—	2	—
貧 農	9	6.5	—	—	—	6.5	—	—	—	—	—
雇 農	8	7	1	—	—	7	1	—	—	—	—

材料來源：根據1950年在張家油房村調查結果。

說 明：(1)可供役使指可以耕作用之牲畜；

(2)不可役使指幼小不能耕作及老弱不堪耕作牲畜。

統計表第三部份是表身，就表中除兩表頭外之各行列數字。

表之起線、終線，兩表頭與表身分別線宜劃粗線或雙線，以免混淆。表之左右皆不劃線，縱行之總計及橫列之共計寫在最前或最後皆可。但因總計或共計皆為最重要數字，宜列於前面。

統計表之縱表頭為該表之主辭，橫表頭為該表之賓辭。如所列表例，乃比較土改後翻身了的貧雇農與地主牲畜數所佔百分比。由此表看出，牲畜與戶數成比例的，而與原成份無何關係。但在土改前牲畜全集中在地主及富農家中，這是個顯明的對比。

縱表頭及橫表頭分類或分組，一般是時間的、地域的數量或其他質的分別。

統計表除開表本體外，還有表首及表尾，表首包括標題、資料地點時間及單位。

標題是很重要的，要簡明的說出表的內容。一般先說明主辭，再說明賓辭及表身。有時候將資料時間及地點寫在標題裏，若標題未寫明時間及地點，在另條中必需註明，否則這張表就失去了效用了。如例表標

題爲張家油房各農戶之耕畜數，因爲資料地點已說明了，所以另一條只寫出了資料時間爲1950年12月，單位爲頭。

表尾一般也包括三種內容：

第一、資料來源 就是這統計表資料是怎樣來的或從那兒來的。

第二、統計表編製者

第三、說明 說明是在表中須加解釋的地方，宜在表中註以1、2、3…不宜劃星號，因爲說明有時很多，劃星號容易混淆，使讀者看時模糊不清。

在統計表項目中，數量若是零，宜劃一小橫線如“—”；未詳的數字可用三點如“…”來表示。

編製統計表基本規則 統計表要簡單明瞭，不能有過多的行列，行列太多，表的意義就模糊了，到最後的行列亦不易觀察。所以在一個統計表中，最好只說明一個目的。若爲完整及聯繫而不可分割的統計資料也不可任意分開另列一表。統計表若有歷史性，則必須與過去的表之項目一致，否則在時間上就不連續，失去了時間數列分析及比較的價值。

統計表的縱接及橫接 統計表之主辭及賓辭雖求精簡，若事實上，有不可分割的情形，又必須列在同一表上時，而主辭太多（就是縱表頭項目太多）一張表不能列下，需二張或二張以上之紙張方可完成，則未完成表的下線不劃，一直到表完後方劃表下線。但每一個相聯的橫表頭必須重寫，以便在這一張中明白主辭與賓辭的關係。如賓辭太多（就是橫表頭過多時）一張表不能寫完，則在未完表的右方劃一直線，表示還有連續的表，直到最後的一張表上，方不劃右方線，連續表之縱表頭需寫出，以便與賓辭互相對照和查閱表身。

第三節 統計圖的製作

統計圖 將統計資料繪成各種圖形就是統計圖。統計資料經過整理後歸納表示的方法只有兩種：一統計表、二統計圖。可是在功用上是不完全相同的。統計圖可供統計分析，如統計資料線的配合，必先看他