立信統計叢書

高級統計學

金國寶著

(一九五一年初版)

立信會計圖書用品社出版

立 信 統 計 叢 書

高 級 統 計 學

金 國 寶 著

（一九五一年初版）

立信會計圖書用品社出版

立信統計叢書

高級統計學

全一冊

每冊人民幣二萬四千元

著者　金國寶

發行所　立信會計圖書用品社
上海河南中路三三九號
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

一九五一年八月初版
一九五一年十二月再版　（滬）

3001—5000（東）

自　　序

這本書是吾在復旦大學統計專修科的講義稿。1949 年先成五章，其後又續成幾章，去歲暑假期中始將全稿完成，共計十四章。一年以來，屢有增損。內容偏重於應用一方面，故書中例題亦以取材於工業農業者居多，尤儘量採用本國統計資料，以增讀者興趣。所有過分高深的理論部分多未涉及。

惟是統計之學精微博大。十九世紀以來，歐洲諸大師迭有發見。最近蘇聯諸專家尤有重大的貢獻，統計學幾乎成爲建設社會主義國家最重要的工具了。淺學如吾，學業久荒，對斯學精微未能窺見萬一，竟以未成熟的作品出而問世，豈不貽笑方家？

但是吾另外有一個想法。古語說得好，"集腋成裘"，"衆擎易舉"。這本書若不印出，祇是吾一個人在那裏摸索；印了出來，倒可以向全國人討教。一方面擬將不必要的部分盡量精簡，一方面工業農業貿易交通等等的資料，除有關機密者外，希望各方面多多供給；俾於再版之時，內容益求充實，嘉惠學者不少。果能取得全國專家的合作與指教，那末這本書的印行，眞可說是拋磚引玉了。

金國寶

1951 年 7 月 27 日

高級統計學

目　次

第一章　頻數分配

第一節　平均數……1
第二節　變差，標準差……8
第三節　平均差……14
第四節　兩極差，四分位差與概差……17
第五節　動差與累差……17
第六節　薛伯氏校正數……22

第二章　機率與機率分配

第一節　排列與組合……26
第二節　機率……30
第三節　機率分配：期望值……35
第四節　連續試驗與二項分配……38

第三章　三種標準分配

第一節　二項分配……44
第二節　卜瓦松分配……47
第三節　常態分配……54
第四節　常態分配之特質……56
第五節　常態曲線之功用……59
第六節　常態曲線之縱線與面積……65
第七節　常態曲線之配合……67

第四章　二元分配：響應與繫聯

第一節　動差……70

第二節 直線響應……72
第三節 繫聯係數與標準誤……74
第四節 單位之轉換……78
第五節 分組資料之計算……79
第六節 由響應方程式所得之估計値……82
第七節 變量和或差之變差……83
第八節 變換式……84

第五章 曲線響應

第一節 繫聯表中之行列……89
第二節 繫聯比……90
第三節 繫聯比與繫聯係數之其他關係……93
第四節 連續分配……95
第五節 多項式響應……96
第六節 繫聯指數……98
第七節 其他曲線響應……100
第八節 二變數之常態分配……101
第九節 常態繫聯曲面之性質……104

第六章 多元分配:複繫聯與淨繫聯

第一節 響應面……108
第二節 標準誤……113
第三節 複繫聯係數……115
第四節 淨繫聯係數……117
第五節 標準誤公式之研究……120
第六節 響應係數公式之研究……121
第七節 其他定理……122
第八節 高斯乘數法……124

第七章 取樣問題:大樣

第一節 統計的歸納……130
第二節 質之子樣……133

第三節　大樣(大子樣)……………………………… 134

第四節　和與差之標準誤……………………………… 135

第五節　大子樣之比較……………………………… 137

第六節　量之子樣……………………………… 138

第七節　平均數之子樣分配……………………………… 139

第八節　常態母體……………………………… 143

第九節　非常態母體與平均數之子樣分配……………………………… 144

第十節　二大子樣平均數之比較……………………………… 146

第十一節　二平均數之和或差之常態性……………………………… 147

第八章　標準誤

第一節　母體與子樣之變差……………………………… 151

第二節　組頻數之標準誤……………………………… 152

第三節　分割數之標準誤……………………………… 154

第四節　標準差之標準誤……………………………… 155

第五節　二大子樣標準差之比較……………………………… 156

第六節　繫聯係數之標準誤……………………………… 157

第九章　Gamma 函數與 Beta 函數

第一節　Gamma 函數……………………………… 160

第二節　Beta 函數……………………………… 162

第三節　積分之變換……………………………… 165

第四節　Beta 函數與 Gamma 函數之關係……………………………… 166

第五節　Gamma 分配與 Gamma 變量……………………………… 168

第六節　Beta 分配……………………………… 171

第十章　χ^2 分配及其應用

第一節　χ^2 分配之性質……………………………… 177

第二節　σ^2 信限之確定……………………………… 183

第三節　χ^2 測驗之性質：子樣變差之顯著性測驗……………………………… 184

第四節　適合度測驗……………………………… 185

第五節 相聯表與獨立性之測驗……189
第六節 顏茨修正法……192
第七節 變差與標準差之分配及同質性之測驗……193
第八節 差異指數……195
第九節 同質繫聯之合併……197

第十一章 "Student" 分配

第一節 小樣(或小子樣)……200
第二節 "Student" 分配……201
第三節 母體假定平均數之測驗與限界……202
第四節 二子樣平均數之比較……205
第五節 繫聯係數之顯著性……208
第六節 響應係數之測驗……210
第七節 二響應係數之比較……212
第八節 淨繫聯係數之測驗……212
第九節 偏響應係數之測驗……214
第十節 二偏響應係數之比較……216

第十二章 其他測驗

第一節 費雪氏之 z 分配……219
第二節 F 測驗……221
第三節 二變差適合性之測驗……223
第四節 平均數同質性之測驗……224
第五節 行與列同質性之測驗……227
第六節 複繫聯係數之測驗……228
第七節 L 測驗……229
第八節 費雪氏繫聯係數之變換式……231
第九節 二繫聯之比較……235
第十節 繫聯係數估計值之合併……236

第十三章 變差分析與共差分析

第一節 母體之同質性……239

第二節　二標準之分組法……………………………………248
第三節　拉丁方(三個分組標準)………………………………253
第四節　變差分析應用於直線響應……………………………256
第五節　變差分析應用於曲線響應與複繫聯…………………258
第六節　響應之絕對標準………………………………………260
第七節　繫聯比顯著性之測驗…………………………………262
第八節　響應直線性之測驗……………………………………263
第九節　共差分析之意義與功用………………………………265
第十節　一個分組標準…………………………………………267
第十一節　乘積和之計算方法…………………………………269
第十二節　響應之消除…………………………………………273
第十三節　二個分組標準………………………………………276

第十四章　統計試驗之設計

第一節　效力與效能……………………………………………286
第二節　二種誤差………………………………………………287
第三節　田間試驗………………………………………………289
第四節　樣品檢驗………………………………………………295
第五節　分層取樣………………………………………………301

附錄甲

§1. 薛伯氏校正數之公式……………………………………305
§2. 卜瓦松分配………………………………………………307
§3. 二獨立的卜瓦松變量之和仍爲卜瓦松變量……………307
§4. 求$\int_0^\infty e^{-x^2}dx$之值……………………………308
§5. $\iint zyx^3dxdy=3r\sigma_x^3\sigma_y$等之公式……………309
§6. $\tan 2\theta=\dfrac{2r\sigma_x\sigma_y}{\sigma_x^2-\sigma_y^2}$……………………………310
§7. X^2之公式………………………………………………311
§8. 樣品檢驗…………………………………………………312

附錄乙 參考書目……………………………………314
附表一 常態曲線下之面積……………………………316
附表二 常態曲線下之縱線……………………………317
附表三 X^2之數值……………………………………318
附表四 t 之數值……………………………………319
附表五 (甲) z 分配之5%點……………………………320
(乙) z 分配之1%點……………………………321
附表六 繫聯係數之顯著值……………………………322
附表七 繫聯係數之改變數……………………………323
附表八 F 之數值……………………………324—329
附表九 L 之數值……………………………330—331

第一章　頻數分配

第一節　平均數

茲有金陵大學1928年601行大豆之產量報告，如下表。

表 1·1　金陵大學1928年601行大豆之產量分配*

產　量　(克)	f
50－	1
80－	6
110－	17
140－	25
170－	64
200－	118
230－	145
260－	108
290－	68
320－	34
350－	10
380－	3
410－	1
440－	1
合　　計	601

表中第一行為產量，可以組限表示，亦可以組中點表示。第二行為各組組限內之行數，是曰頻數。第一行變數之數值連同第二行之頻數，是為601行大豆產量之頻數分配。設以各組之值為$X_1, X_2, \cdots, X_n$，而各組之頻數為$f_1, f_2, \cdots, f_n$，各組頻數之和等於此分配內之行數，即

$$N=f_1+f_2+f_3+\cdots+f_n=\sum_{i=1}^{n} f_i \tag{1}$$

* 王綬，實用生物統計法11頁。

式中$\sum_{i=1}^{n} f_i$有時書作$\sum_i f_i$或$\sum f_i$，其意亦同。

(1) 算術平均數　頻數分配之平均數乃此變數N個數值之算術平均數$\bar{X}$，其式如下：

$$\bar{X}=\frac{\Sigma fX}{\Sigma f}=\frac{1}{N}(f_1X_1+f_2X_2+\cdots+f_nX_n)=\frac{1}{N}\sum_{i=1}^{n}f_iX_i \quad (2)$$

就此式觀之，$\bar{X}$實是X_i各值之加權平均數，各組之頻數f_i卽其權數。

變數X之數值可爲高度，亦可爲重量或年齡，亦可爲每畝田之收穫量等等。變量無論爲何種數值，f總是X各值之頻數，而N爲其總頻數。N可由(1)式求得，其平均數$\bar{X}$可由(2)式求得。

設有二個變數u與v，在頻數分配之中對於每一個u值，卽有v相應之一值，故此二變數之值均是成組的。設令N爲此等相應值之組數，而

$$X_i=u_i+v_i$$

則X之N值之算術平均數等於等號右方N值之算術平均數，卽

$$\bar{X}=\bar{u}+\bar{v} \quad (3)$$

此卽謂二變數和之平均數，卽等於其平均數之和。此可推廣至任何個數之變數，均是如此。又若a與b爲常數，而

$$X=au+bv$$

吾人不難證明

$$\bar{X}=a\bar{u}+b\bar{v} \quad (3')$$

$\bar{X}$之性質　$\bar{X}$有二種重要之性質，玆於下列二定理述之。

定理一　一切變量對其算術平均數離差之和爲零。

證： 今令x爲對於算術平均數之離差，卽$X-\bar{X}=x$，則

$$\sum_1^k f_i x_i = \sum_1^k f_i (X_i - \overline{X})$$

$$= \sum_1^k f_i X_i - \sum_1^k f_i \overline{X}$$

$$= \sum_1^k f_i X_i - \overline{X} \sum_1^k f_i$$

$$= N\overline{X} - \overline{X}N$$

$$= 0 \qquad (4)$$

定理二　若變量作下列之變換，新原點 X_o，單位 C，而

$$u = \frac{X - X_o}{C} \qquad (c \neq o) \qquad (5)$$

則新舊平均數 $\overline{X}$ 與 $\overline{u}$ 間有下列之關係：

$$\overline{X} = C\overline{u} + X_o \qquad (6)$$

證：依(5)，$X = Cu + X_o$

以之代入(2)式則

$$\overline{X} = \frac{1}{N} \sum_1^k f_i (cu_i + X_o)$$

$$= \frac{c}{N} \sum_1^k f_i u_i + \frac{X_o}{N} \sum_1^k f_i$$

$$= c\overline{u} + X_o$$

系：各變量對任意數值 X_o 之離差平均數若已求得，則將此加上 X_o 之代數的和，結果即爲各變量之算術平均數 $\overline{X}$，以符號表示之，則

$$\overline{X} = \frac{1}{N} \sum_1^k f_i (X_i - X_o) + X_o \qquad (7)$$

此可以上文(5)，(6)二式證明之。

此處之 X_o 有人稱之曰臨時平均數，亦有人稱之曰工作平均數，亦稱曰假定平均數。根據此式求算術平均數，甚爲便利。例如有 205，197，200，204 四數，求其算術平均數，可取 200 作爲 X_o，則對此 X_o 之離差

爲5，−3，0，4，於是 $\overline{X}=200+\frac{1}{4}(5-3+0+4)=201.5$。

統計資料若已組成頻數分配，則公式(7)可以書作下式：

$$\overline{X}=X_0+\frac{\Sigma fx'}{\Sigma f}w \tag{8}$$

式中 x' 爲對此臨時平均數 X_0 之離差，而以組距爲單位者，w 爲組距。

例一　今將第一表之統計資料依公式(8)求算術平均數，則如下表：

表 1·2　算術平均數之計算簡法

產量(克) X	頻　數 f	離　差 x'	fx'
50−	1	−6	−6
80−	6	−5	−30
110−	17	−4	−68
140−	25	−3	−75
170−	64	−2	−128
200−	118	−1	−118
230−	145	0	0
260−	108	1	108
290−	68	2	136
320−	34	3	102
350−	10	4	40
380−	3	5	15
410−	1	6	6
440−	1	7	7
合　計	601		−11

今 $X_0=245$ 克，$w=30$ 克，故

$$\overline{X}=245-\frac{11}{601}\times 30=244.45\text{ 克}$$

若頻數分配是一個連續分配，而爲頻數曲線 $Y=f(X)$ 所表示者，則其算術平均數可由下式求得：

$$\overline{X}=\frac{\int Xf(X)dX}{\int f(X)dX} \tag{9}$$

如果曲線下之面積作爲一,則 $\overline{X}$ 之値等於此式中之分子可矣。

平均數之平均數　以上吾人謂各變量均在一組中，可用一個下碼足矣,但設有兩組變量,一組中 N_1 個，另一組中 N_2 個，則吾人不得不加一下碼,以資區別。如第一組中，

$$X_{11}, X_{12}, X_{13}, \cdots, X_{1N_1},$$

而第二組中則爲

$$X_{21}, X_{22}, X_{23}, \cdots, X_{2N_2},$$

此兩組之變量各有一平均數,第一組之平均數爲

$$\overline{X}_1 = \frac{1}{N_1}\sum_{i=1}^{N_1} X_{1i} \qquad (a)$$

而第二組之平均數爲

$$\overline{X}_2 = \frac{1}{N_2}\sum_{i=1}^{N_2} X_{2i} \qquad (b)$$

定理三　若一組 N_1 個變量中之平均數爲 $\overline{X}_1$，另一組 N_2 個變量之平均數爲 $\overline{X}_2$，則二組合計之平均數 $\overline{X}$ 爲

$$\overline{X} = \frac{N_1\overline{X}_1 + N_2\overline{X}_2}{N} \qquad (10)$$

此處 $N = N_1 + N_2$

證：依上列(a),(b)二式，

$$\frac{N_1\overline{X}_1 + N_2\overline{X}_2}{N_1 + N_2} = \frac{\sum_1^{N_1} X_{1i} + \sum_1^{N_2} X_{2i}}{N_1 + N_2} = \frac{\sum_1^{N_1+N_2} X_i}{N} = \overline{X}$$

上列(10)式亦可書作

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{2} n_i\overline{X}_i \qquad N = \sum_{i=1}^{2} n_i$$

將此式普遍化,可推廣至於 k 組,故有下列之定理。

定理四　若有 N 個變量由 k 組而成,則其平均數爲

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{k} n_i\overline{X}_i \qquad (11)$$

此處 $\overline{X}_i$ 爲第 i 組之平均數，n_i 爲其頻數，而 $N=\sum_{i=1}^{k} n_i$。

系：若 $n_i=n$ 在各組中均相等，則 $N=kn$，而(11)式變爲

$$\overline{X}=\frac{1}{k}\sum_{1}^{k}\overline{X}_i \tag{12}$$

(2) 中位數　若將無數之值依其大小次序排列，如 N 爲奇數，則其中間一值爲中位數；如 N 爲偶數，則其中間二項之算術平均數爲中位數。故就 N 個數值之中位數之位次爲第 $\frac{N+1}{2}$ 項。

1，4，5，7，10，

中間之一值卽第 $\frac{5+1}{2}=3$ 項，故中位數爲 5。但下列之數列

1，4，5，7，10，12，

中位數爲第 $\frac{6+1}{2}=3.5$ 項，卽第 3 項與第 4 項之間，故取中間兩項卽 5 與 7 之算術平均數卽 $\frac{5+7}{2}=6$ 爲中位數。

但在頻數分配表中求中位數，須用插補法。此項插補法根據於一種假定，卽謂中位數所在之一組中各項之值均是均勻分配。其公式如下：

$$\dot{X}=l+\frac{\frac{1}{2}(N+1)-N_b-\frac{1}{2}}{N_m}w \tag{13}$$

此式改簡，則爲

$$\dot{X}=l+\frac{\frac{1}{2}N-N_b}{N_m}w \tag{14}$$

式中　$\dot{X}$＝中位數；

l＝中位數組之下限；

N＝總頻數；

N_m＝中位數組之頻數；

N_b＝中位數組下各組頻數之和；

w＝組距。

例二　仍用第一表之統計資料求中位數。

此處 $N=601$，中位數之位次爲 $(601+1)/2=301$。查表中行數在 230 克以下者 231 行，在 260 克以下者 376 行，故中位數必在 230-260 一組中。換言之，中位數組之下限爲 230 克。依公式(14)，

$$\dot{X}=230+\frac{300.5-231}{145}\times 30=230+\frac{69.5}{145}\times 30=244.38\text{ 克}$$

同理，四分位數，Q_1，Q_2，Q_3 乃變數之全距中將全體項數分爲四等分之各值。Q_2 卽是中位數。上四分位數與下四位數之差卽 Q_3-Q_1，名曰四分位差。十分位數乃將全體項數分成十等分，百分位數乃將全體項數分成一百等分。中位數，四分位數，十分位數，百分位數等統名曰分割數或等分位數。

若頻數分配爲連續分配，則中位數乃縱線將頻數曲線下面積分成二等分之 X 值也。若以 $\dot{X}$ 爲中位數，則

$$\int_{-\infty}^{\dot{X}} f(X)\,dX=\int_{\dot{X}}^{\infty} f(X)\,dX \tag{15}$$

(3) 衆數　衆數者乃頻數曲線下頻數最大時之 X 值。在衆數點上之縱線大於左右任何其他縱線。多數曲線在頻數分配之中心附近有一個最大值，但亦有一種曲線有二個以上之最大值者。

求衆數之法並不容易。唯一妥善之法在對於頻數分配配合一理論頻數曲線，於是求其最大頻數之 X 值。若頻數函數爲連續，且可求導微函數則可就下列二式解之卽得：

$$f'(X)=\frac{d}{dX}f(X)=0,\quad f''(X)=\frac{d^2}{dX^2}f(X)<0$$

若 $f'(X)$ 爲零，而 $f''(X)$ 大於零，則所得結果爲最小值，而此最小值之點有時名之曰反衆數。

(4) 幾何平均數　凡 N 個數量乘積之 N 次方根爲此 N 個數量之幾何平均數。計算之時，用對數爲便。若以 $X_1, X_2, \cdots, X_n$ 之幾何平均數爲 $\overline{X}$，則