CONTEMPORARY EDUCATIONAL SERIES

THE FUNDAMENTALS OF STATISTICS

BY

L. L. THURSTONE, M. E., Ph. D.

TRANSLATED BY

JENNINGS P. CHU, Ph. D.

THE COMMERCIAL PRESS, LTD., SHANGHAI

美國塞斯頓原著
朱君毅博士譯述

現代教育名著

教育統計學綱要

商務印書館發行

現代教育名著

教育統計學綱要

此書有著作權翻印必究

中華民國十七年八月初版

(每冊定價大洋壹元伍角
外埠酌加運費匯費)

原著者　美國塞斯頓

譯述者　朱君毅博士

發行兼印刷者　上海寶山路 商務印書館

發行所　上海及各埠 商務印書館

Contemporary Educational Series

THE FUNDAMENTALS OF STATISTICS

By

L. L. THURSTONE, M. E., Ph. D.

Translated by

JENNINGS, P. CHU, Ph. D.

1st ed., Aug., 1928

Price: $1.50, postage extra

THE COMMERCIAL PRESS, LTD., SHANGHAI

一五五二分

編輯主任原序

當今之世,吾人欲瀏覽教育書報,若不明瞭統計之名詞與方法,未有能愉快而入勝者。十五年來,教育文字之性質已大受變遷。昔日書籍之關於教育方法與教育行政者,圖表絕少,而名詞之如相關,中數,衆數,離中差數,次數表,次數面,均方差,機誤曲線,百分等級等,尤不多覯。今也不然,任取教育書籍若干種,其含有統計方法與名詞者,十居八九,苟讀者昧於統計之學,則未有能澈底了解其意義者。

昔日之教育家,嘗漠視統計事實而侈談教育問題,今則非其時矣。昔日之憑理性,經驗與常識,而不顧實際情形者,今則不復能取信於人矣。最近研究教育問題者,必採用精確之方法以蒐集材料,明示事實之分配,詳述手續之經過,然後供獻其研究之結果,務使讀者了然於胸中,不稍有疑義焉。

教育事實,本甚複雜。一種問題之發生,常與多種原因相錯綜,苟不條分縷析,追本溯源,將何從而探求其特性乎。彼缺乏經驗者,對於此種繁複之問題,往往束手無策,卽令其所得事實,均屬可靠,而組織無方,終未能顯示事實之趨勢與眞詮。故欲從事於調查,研究者,必先洞悉其方法,然後

能免於錯誤,而結果亦不致乖謬叢生矣。

今日出版書籍之討論統計方法者,其取材常專門而高深,令人問津無從,望洋興歎,惜哉。塞斯頓君之作,簡明切要非尋常者可比。凡學校教師,研究人員,倘能取而讀之,摭其精華,則對於專門書籍,瀏覽無阻,而教育上之研究,亦得施行有方。至於本書舉例之確切,猶其餘事耳,按此書可爲統計方法入門之書,著者嘗以統計方法,教授學習教育心理之學生,故凡初學者對於統計之困難,不特洞悉無遺,並能解除而輔助之。

凡欲了解教育研究之結果,或有志於研究教育者,鄙人敢以此簡要之書介紹之。

威斯康新大學　　　　渥雪
(M. V. O'Shea)

著者原序

余嘗授大學院學生以統計方法與原則,歷七載而此書告成。凡大學院學生之選習心理者,當其在大學時,大都研究經濟文學語言等科,缺乏高深之數學知識,甚有中學程度之數學亦遺忘殆盡者;但此輩數學之知識,雖較缺漏,而其對於統計學之理論的方面之批評力,常較豐富。余對於此種批評態度,常加以鼓勵,而對於盲從之行為,則竭力阻止之。蓋初學者不難於學習計算相關及繪圖作表,而難於澈底了解其意義。此書之目的,卽在闡明統計學之眞義焉。

是書之成,余不能不表謝忱於余之學生,因書中之內容,皆余平日教授若輩時之心得也。

統計書籍中之有助於拙作者甚多。余所用之相關紙,係桑戴克相關表之變相。第三十圖第三十八圖及其他百分表與相關表之材料,多由「教育研究雜誌」(Journal of Educational Research)中之拙文中採取,發行者慨然許可,余甚感之。第十九表與第二十表之材料,係由闞而生所編輯之"統計學家與生命學家用表"內薛伯氏(Sheppard)所作之長表中採取。此二表雖甚簡,但已足供初學之用。欲求工作

精密,仍以參看薛氏之表爲宜。

余意此書不特對於初學統計者有用,即一般研究心理測驗而不明統計學者亦可作他山之助。學者既明瞭集中量數,離中量數與相關量數之後,更宜進而研習猶爾(Yule),鮑萊(Bowley),愛爾得登(Elderton),湯姆生(Thompson),與凱萊(Kelley)諸氏之作,以資深造。數學不深者,宜讀猶爾(Yule)之書,其完美詳備,固無待余之贅言也。

海內學者,倘能不吝教益,以匡不逮,則幸甚矣。

一九二四年於芝加哥　　　　塞斯頓
(L. L. Thurstone)

譯者自序

美國芝加哥大學教授塞斯頓博士所著之「教育統計學綱要」,爲美國「實驗教育叢書」之一。內容簡明切要,而其討論圖示法之詳盡,尤非他書所能及。爰譯成漢文,以供吾國研究教育問題者之參考焉。

本書名稱,原爲「統計學綱要」,顧書中所載之方法與舉例,均屬教育性質,故改稱爲「教育統計學綱要」。

書中所用名詞,悉採於拙著「統計與測驗名詞漢譯」及「教育統計學」二書。

譯此書時,薛君天漢,襄助最力,特此誌謝。

民國十六年正月朱君毅序於北京清華學校

目 錄

圖次

表　次

教育統計學綱要

第一章

次數表

吾人若將有數目之事實,集合而統計之,則其第一步工作,當爲分類。假設試行心理測驗於三百學生而評定其試卷。或問張生之分數爲多少,一查而知爲79。又問此分數之地位如何。苟不知在79分以上者共若干人,79分以下者共若干人,吾人決不能答此問題。若他生之分數均在79分以下者,則張生之分數爲極高;若他生之分數均在79分以上者,則張生之分數爲極低;若300人中有150人之分數在79分以上而其他150人之分數在79分以下者,則張生之分數,適爲平均。由此足見試行測驗,評定分數,猶未能盡善盡美,必也將分數分類列表,然後知得九十分者幾人,得八十分者幾人。此表謂之**次數表**(Frequency Table)。

司華司麻(Swarthmore College)大學,曾舉行一年級生智力測驗。其分數詳**第一表**。每一數目代表每個學生之分

數。欲知該級學生,在79分以上者共若干人,在79分以下者共若干人,則非細察全表不可。但事實上苟能照第一圖排列,則此項手續,可以免矣。

62	129	95	123	81	93	105	95	96	80
123	60	72	86	108	120	57	113	65	108
109	84	121	60	84	128	100	72	119	103
77	91	51	100	63	107	76	82	110	63
104	107	63	117	116	86	115	62	122	92
69	116	82	95	72	121	52	80	100	85
94	84	123	42	90	91	81	116	73	79
100	79	101	98	110	95	67	77	91	95
79	92	73	83	74	125	101	82	71	75
125	56	86	98	106	72	117	89	99	86
87	90	80	131	102	117	98	74	101	82
110	137	99	65	113	85	82	90	102	57
139	74	149	114	74	102	69	134	78	106
75	106	85	103	78	106	102	94	108	90

第一表. 司華司麻(Swarthmore College)大學一年級生智力測驗分數

編製次數表之方法如下:

1. 在事實張❶(Data Sheet)上作三縱目,爲分數(Score)表列(Tabulation)次數(Frequency)如第一圖。

2. 先讀第一表之分數,每讀一個分數時,在第一圖

❶ 事實張爲一劃有縱行之張,用以記載事實,但記載事實時,每一縱行上,必寫縱目。

上作一劃記,每作四個劃記時,作一斜劃橫穿之,則計數較易,如第一圖。

分數	表列	次數
0-9		
10-19		
20-29		
30-39		
40-49	|	1
50-59	卌	5
60-69	卌 卌 ||	12
70-79	卌 卌 卌 卌 |	21
80-89	卌 卌 卌 卌 |||	23
90-99	卌 卌 卌 卌 |||	23
100-109	卌 卌 卌 卌 卌	25
110-119	卌 卌 ||||	14
120-129	卌 卌 |	11
130-139	||||	4
140-149	|	1
150-159		
160-169		
學生數＝		140

第一圖　司華司麻(Swarthmore)大學一年級生智力測驗分數之次數表

3.　將每橫行劃記相加,而書其總數於次數欄下。

4.　將次數縱行相加,爲全體總數。惟此項總數,應與

第一表次數之總數相脗合。

既作以上各節,即能解答下列諸問題:

1. 分數在70與79之間者共幾人? (21)

2. 分數在60與69之間者共幾人? (12)

3. 分數在99以上者共幾人? (55)

4. 分數在70以下者共幾人? (18)

5. 分數在20與49之間者共幾人? (1)

6. 分數在80與89之間者,佔全級百分之幾? (16%)

7. 分數在40與49之間者,佔全級百分之幾? (28%)

8. 得95分者,是高是低抑為平均? (平均)

9. 分數超過89者,佔全級百分之幾? (56%)

問題之類此者,一看次數表,即可解答。

變量(Variable)為一種含有不同價值之數量。無論何種常變之數量,均為變量。例如人口,年齡,生死律,物價,工資,氣壓表之度數,雨量之紀錄等皆是。智力測驗之分數,亦為變量。每組各人在某種測驗所得之分數為變量。每人對於某種測驗在不同時間所得之分數,亦為變量。

變量可分為繼續(Continuous)與分立(Discontinuous)兩種。凡變量之可分為極小等級者,稱之曰繼續變量。不能分者稱之曰分立變量。例如溫度之由68度而升至69度,其間可經過無數之價值,故溫度為繼續變量。火車中之貨車數目應視為分立變量,因每車分立,中間不能繼續,倘此火

車共有68輛，則由68而至69，決不能謂經過$68\frac{1}{3}$輛車與夫$68\frac{5}{7}$輛車也。

全距離(Range)爲變量價值之最大者與最小者之差，司華司廉大學學生測驗分數之全距離爲107，此卽該級最高分數與最低分數之差(卽149與42之差)。

組距(Class Interval)爲量表分爲若干相等部分之一部。若表列傭工之年齡時，應依年齡而分，如是則一年可爲一組距。若分類再求精細，則可用半年或幾月爲一組距。兒童年齡分類時，卽用此法。吾人若將年薪分類，則$100或可視爲一便利組距。如是則得$1600與$1699者，歸爲一組，其第二組則爲薪水之在$1700與$1799者；餘類推。如第一圖，以十爲一組距，在70－79之一組者，共有21人。

組限(Class Limits)．假設有一量表，自0以至100，共分爲十組距，列之如下：　0－10，　10－20，　20－30，　等，表列各數時，除十之倍數外，均無困難。然表列30時，則此數將歸20－30之一組，抑爲30－40之一組，吾人不無猶豫。欲免除此種含糊之現象，則各組距宜彼此分清界限，而無重複之弊。其法有三：

1.　上例之組距，可表示之如下：　0－9.9，　10－19.9，20－29.9，餘類推。如是則30應歸入30－39.9一組，而不能歸入他組也明矣。

2.　以上所列之組距，更可以用組距中點表示之：5，

15,25,35,餘類推。如是則表列之時,凡數目之與某組距中點相近者卽可歸入。例如27,應歸爲25之一組,而不歸入35之一組,因27與25爲較近。但此法不必採取,蓋表列10或10之倍數時,仍不免有界限不淸之弊。

3. 以上組距,亦可表讀如下: 0至少於10, 10至少於20, 20至少於30,餘類推。如是則可以免除界限不淸之弊,然終不如第一法之爲簡便。

組距次數(Class Frequency)爲某組距中所包含之次數(人數或量數)。按第一圖,學生所得之智力測驗分數,在100與109之間者,共25人。故25爲100－109一組之組距次數。然則此組距次數之總和,必等於全表之總數。

119	97	109	97	103	128	109	119	95	101	109	133	79	41	118
102	57	69	147	57	88	102	98	57	92	88	98	101	78	62
97	108	102	102	98	107	87	106	155	50	106	79	57	105	57
132	58	98	88	57	89	130	83	101	89	70	70	131	52	136
88	94	128	104	157	114	92	115	70	100	79	92	96	95	90
108	142	83	83	101	95	95	78	91	83	115	90	57	120	98
89	72	89	118	95	100	120	105	128	80	49	112	100	70	105
120	83	114	69	126	105	80	89	78	100	105	76	80	77	66
82	113	100	94	89	91	93	82	95	80	60	137	90	53	105
94	58	84	91	124	110	76	112	80	126	89	66	88	71	122
125	93	93	110	87	86	89	70	92	86	109	76	85	112	63
82	129	91	74	120	86	94	73	61	110	72	93	64	61	76
117	57	93	81	84	66	65	117	86	62	69	91	88	71	110
92	72	81	127	94	82	85	61	61	84	55	54	71	60	60
103	86	104	85	62	135	96	76	110	60	122	96	110	50	49
38	69	58	65	39	60	60	67	67	54	50	53	69	52	48
37	46	74	71	68	75	47	71	46	75	71	67	45		

第二表 駱斐耶得大學第一年生智力測驗分數

問題一　第二表係駱斐耶得大學(Lafayette College)第一年生智力測驗之分數。表中共253人,其作法與第一圖相同。試細察此表,然後解答以下諸問題:

1. 駱斐耶得大學學生之分數,在80—89之間者共幾人?
2. 其分數在110—139之間者共幾人?
3. 其分數在70以下者共幾人?
4. 其分數在80與109之間者,佔全級百分之幾?
5. 得110分或高于110分者共幾人?
6. 得110分或高于110分者,占全級百分之幾?

第二章

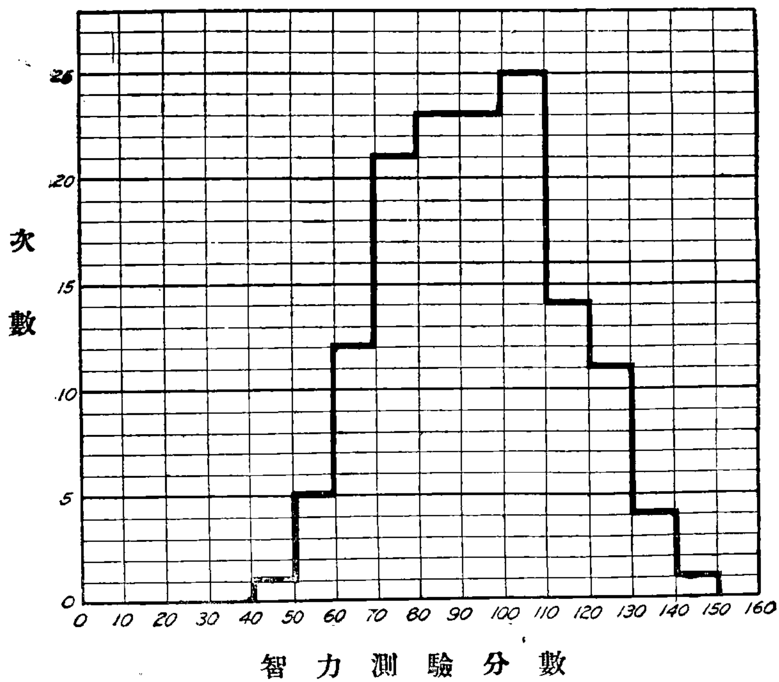
直方圖

吾人不論閱何種報告,若數字繁多者,每注意於文字方面,苟非有特殊之興味,常略而不閱。如第一表然,吾人罕有用心細閱者,蓋其意義不易探得也。事實之排列於第一圖者,較為明瞭,故探求事實於一次數表者較易,探求事實於許多毫無倫次之數目者較難。雖然,次數表所示次數大小之意義,非經精細視察,亦不甚顯明。欲求其顯明,惟有示之以圖。茲申論之。

在一個次數表上,量表之組距,常排在一縱行上,如第一圖。在繪圖時,量表之組距,則排在一橫行上,如第二圖。其法每一組距上,畫出一縱行,高與該組之次數成正比例。次數之量表,則在圖之左方。例如在第一圖上,有五個學生其分數在50與59之間。在第二圖上即用縱行之5格與50–59組距相交切處表示之。其他縱行作法相同,可與第一圖之次數表相對照。凡次數上之事實,均可用第二圖所示直方圖表出,故欲比較幾種事實時,直方圖最為適宜。

參看直方圖時,下列問題,可一一解答:

1. 組距之次數,以何者為最大?
2. 近似之全距離為何?(精密之全距離何以不能得之於直方圖?)



第二圖　直方圖之組距爲十

3. 同一次數之二組距爲何?

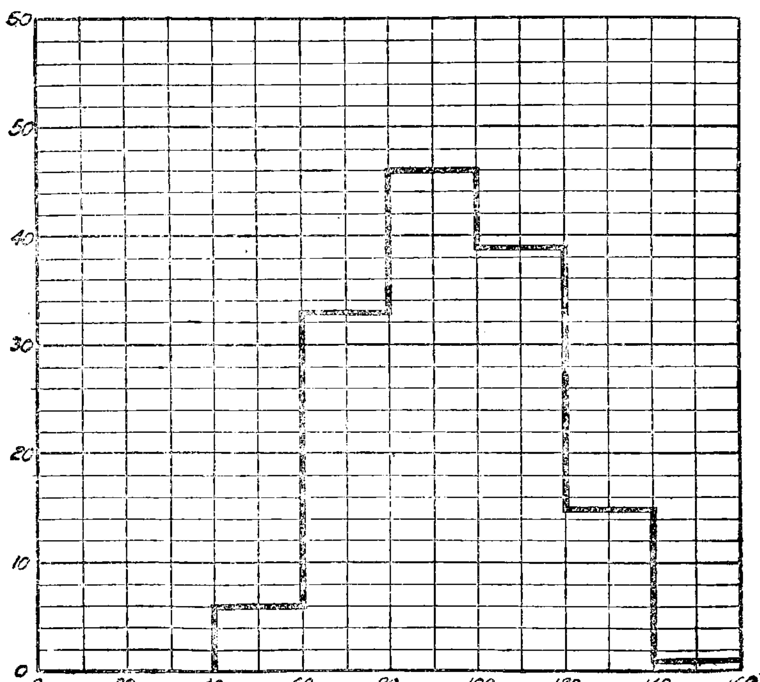
4. 有鄰接之兩組距,其一組距之次數,約兩倍於其他一組距之次數者爲何?

5. 組距 110－119 之次數爲何?

6. 得分在 90 與 109 之間者,有若干人?

組距之大小. 用組距表示事實時,其大小可以隨便。第二圖所用之組距,以 10 爲單位,與第一圖之次數表相當。但吾人若將同類事實,列入 20 單位之組距,則得第三圖。第

三圖之40–59一組,縱行之高爲6,其意即得分數在40與59之間者,爲6人。同類事實,可用第二圖證明。按第二圖,得分數在40–49之間者爲1人,得分數在50–59之間者爲5人,或得分數在40–59之間者,共爲6人。復將同類事實,用第一圖證明。關於60–79之一組,可用同一方法,將第二圖及第三圖比較。斷定組距之大小時有二點應切記者:(1)組距之最小限度,宜使同在一組距之各人,在事實上可視爲

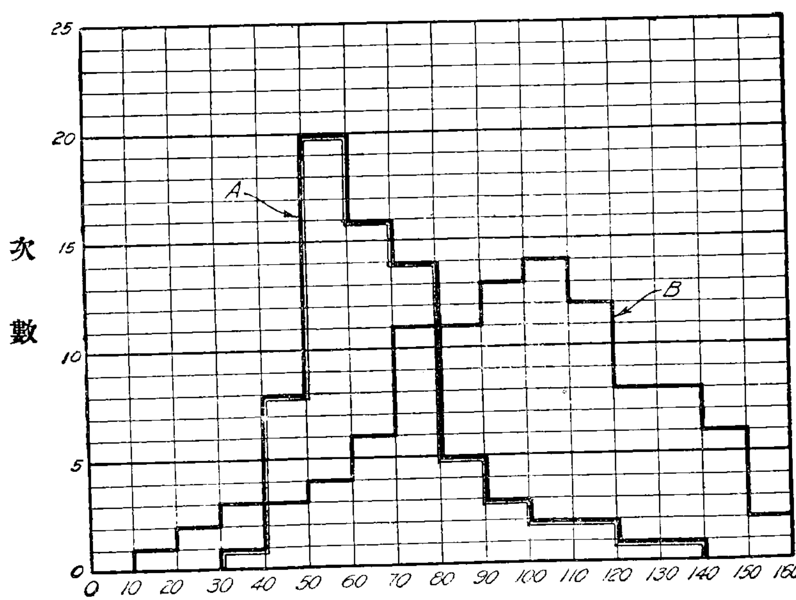


第三圖　直方圖之組距爲二十者

相等;(2)組距之最大限度,宜使次數表不致繁瑣,而使用應手。大概組距之大小,以15到25爲單位者,最爲適宜。

直方圖有一特徵,爲吾人所常用者,卽直方圖所佔之面積,與其所代表之人數成正比例。

斷定**第二圖**及**第三圖**之面積,並證明二圖之面積均爲140,此數卽爲總數,與**第一表**及**第一圖**所示者相同.惟在斷定與證明之時,須知直方圖所代表之面積,不限於某種方格紙所含之方格數目。凡底線上之一組距,應視爲一單位。縱線上之單位,卽次數也。



第四圖　重疊之直方圖

問題1. 參照第一章問題1所載駱斐耶得大學第一年生之次數表,繪三個直方圖,其組距爲10,20,與30。斷定每直方圖之面積,並證明面積與總數成正比例。每圖之標題法,宜與第四圖相同。

問題2. 第四圖代表A與B兩班智力測驗分數之直方圖。細察此圖而解答下問,並申述其理由:

1. 根據測驗,智力最高之學生,在何班?
2. 根據測驗,智力最低之學生,在何班?
3. 何班人數爲較多?(一望卽知)
4. 大約得較高之平均數者爲何班?
5. 何班之全距離爲較大?
6. A班共有多少學生?
7. B班共有多少學生?
8. 若次數表之組距加大,則組距之數目,有何變更?

第三章

次數多邊圖

次數多邊圖(Frequency Polygon)爲一圖形,與直方圖相同,而其意義亦相近。但此圖較直方圖爲佳,因其表示分配之性質,較爲明晰。用圖形表示次數時,有一重要之假設,至此不得不聲明者。卽用直方圖時,分數歸入各組。凡分數之在40與50間者,歸爲一組,而該組所含分數之數目,謂之組次數。玆爲實用起見,假設40–50組之分數,均在該組之中點,卽45。如是則得41分之人與得49分之人視爲相等。二者旣歸入一組,均應視爲與該組中點相等,卽45也。蓋組距自0而至160,以10單位爲一組距,計共有16組距,此種分法,在實際上可謂精密。故此處之假設,卽凡一組之分數,均集中於該組之中點。

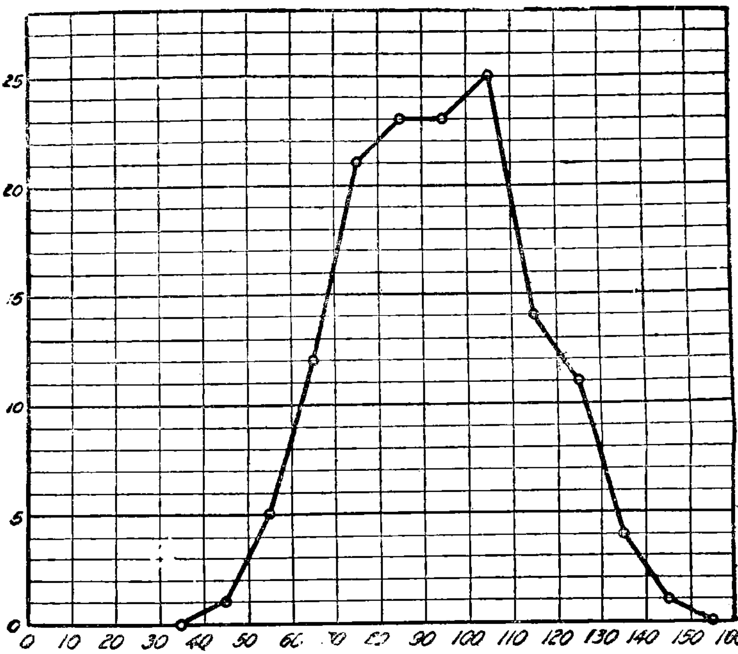
次數多邊圖之作法如下(參看**第五圖**;此圖代表之事實,與**第二圖**相同):

1. 用任何適用量表,將組距在底線上表出之,其作法與直方圖相同。

2. 次數在圖之左方,用縱線表出之,此亦與直方圖同。

3. 在每組距中點上,作一小點,其高度卽爲該組之次數。第五圖上之點,係用小圈表示。

4. 將此許多小圈,用直線連接之。



次數

智力測驗分數

第五圖　次數多邊圖

直方圖上,可重疊一個次數多邊圖,其法卽用直線連接直方圖每組橫線上之中點。

比較**第五圖**與**第二圖**,並明瞭該二圖所代表之事實;卽**第一圖**也。

問題 1.　參照第一章問題 1 所載之騎斐耶得大學第一年生之次數表,並參照第二章問題 1 表示同類事實之直方圖。試用此同類材料,作一次數多邊圖。

第四章

直線相關

本章所討論者,爲二變量相關之圖示法。但表示此種相關之方程式,暫不論及。

第六圖顯出二種變量之通常圖示法。其底線爲橫坐標(Axis of Abscissas)或稱x軸(x-axis)。此軸常用以代表測驗之分數。其左方縱線爲縱坐標(Axis of Ordinates),或稱y軸(y-axis)。此軸常用以代表組距之次數。量表之x軸作法,常自左而右,量表y軸之作法,常自下而上,如第六圖。故x與y軸,均從左下角起點。此點稱之曰原始點(Origin),亦即二量表之零點。

圖之兩軸,不可無標題。標題又必須明晰,使讀者一望而知兩軸之單位爲何,如溫度之用攝氏表,一點鐘後記得之字數,計時之用秒等。如是,則讀者不賴書本之助,而一目瞭然矣。

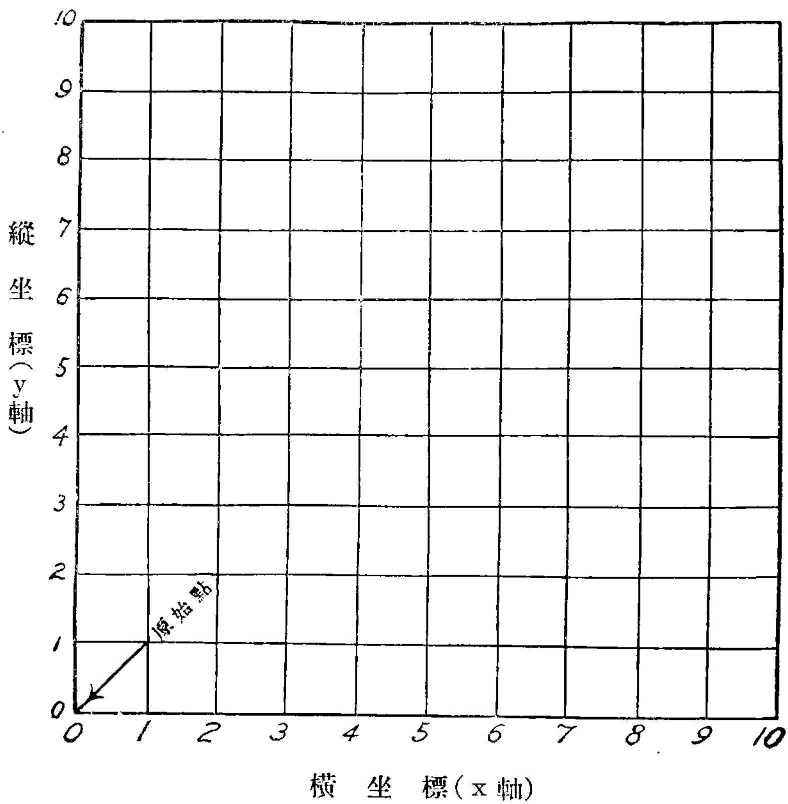
有二個變量時,常常一個變量爲已知,而同時欲求得其他一個變量。變量之已知者,謂之**自變量**(Independent Variable);變量之待求得者,謂之**倚變量**(Dependent Variable)。

獨立變量常載在x軸上;附屬變量常載在y軸上。

設有許多數目,而每個均用一常數除之。如17.6。

凡數目之經過一番計算而仍不變者謂之**常數**(constant)。

若數目太多,則冗長之除法可省,而用圖式求之。設用四個數目之除法,列之如下表。在此表內,已知數爲 x。其第二行爲商數$\frac{x}{17.6}$,即 y。證實之。

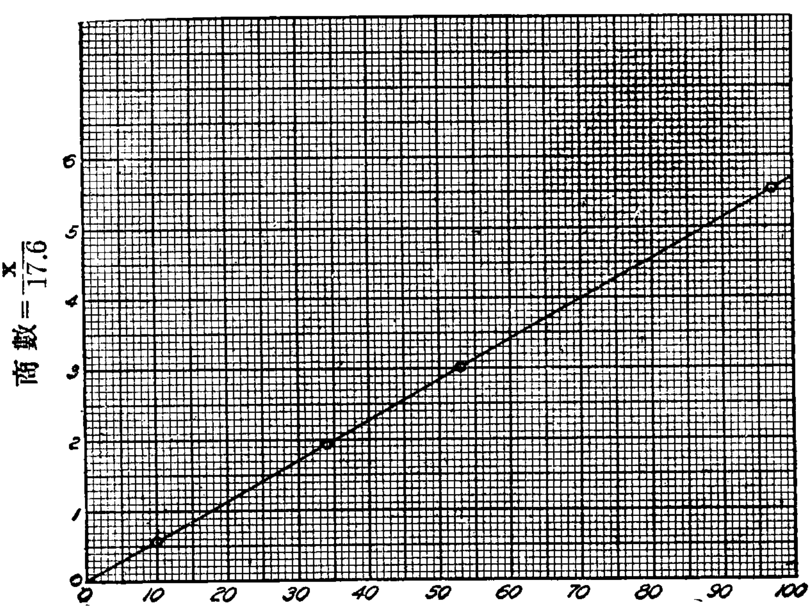


第六圖 橫坐標與縱坐標

x	y
10	.57

34	1.93
53	3.01
97	5.51

現將此四點,作在**第七圖**上。其最高點爲 x 軸上之97,此點與 y 軸上相對之點,爲 5.51。證實其他三點並用圖比較其位置。最重要者,卽此四點,均在一直線上。此直線卽從此四點連結而成,如圖。其他各數之商數,亦可視圖卽知,其便利莫甚於此。如 $\frac{78}{17.6}$,其商數爲 4.43。欲得此數,可先求 x 軸上之78,由此向上推,得4.43卽爲答數。不過用圖時,第三數祇能近似,而不能精確。此種讀圖之法,久練則熟,熟則生



第七圖　圖示乘除法

巧。今試以常數17.6除25,39,與87而求其商數。若按照此圖,以17.6除0,則其商數應爲何數?

現試繪吋與糎之相關。由一數而變爲他數時,可用下式:

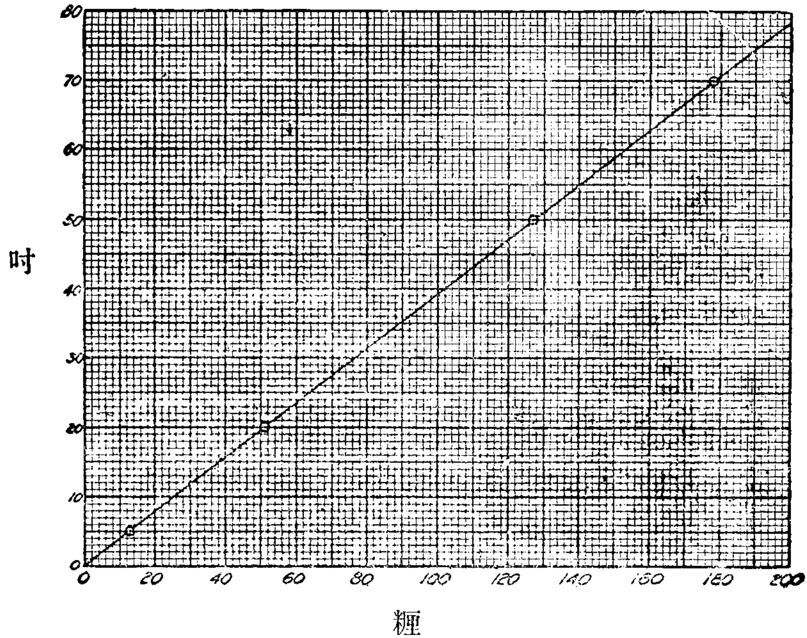
1吋＝2.54糎

任舉數例,有如下表:

y	x
吋	糎
5	12.7
20	50.8
50	127.0
70	177.8

以上四數,繪之如第八圖。一視此圖,卽知20吋爲幾糎,36吋,44.5吋等爲幾糎。若爲44.5吋,則在橫線44與45之間,須插入他數。此種工夫久習之亦不難也。同法可定若干糎數時吋數爲多少。試用圖證實100糎卽爲39吋。

學者宜知作線於方格紙上,非僅繪數線而已。圖之如第七第八者,其所表出之意義,實有非文字所能盡者。凡學習統計之學生,千萬不可無了解統計圖表之能力。蓋統計之事實,無不一用圖表,而其意義之披露,較文字之形容爲明晰。故學者宜常以觀察圖表之眼光,而想像統計上之事實。請舉例以明之。若問.30之平方爲何數,則必答曰"九百",



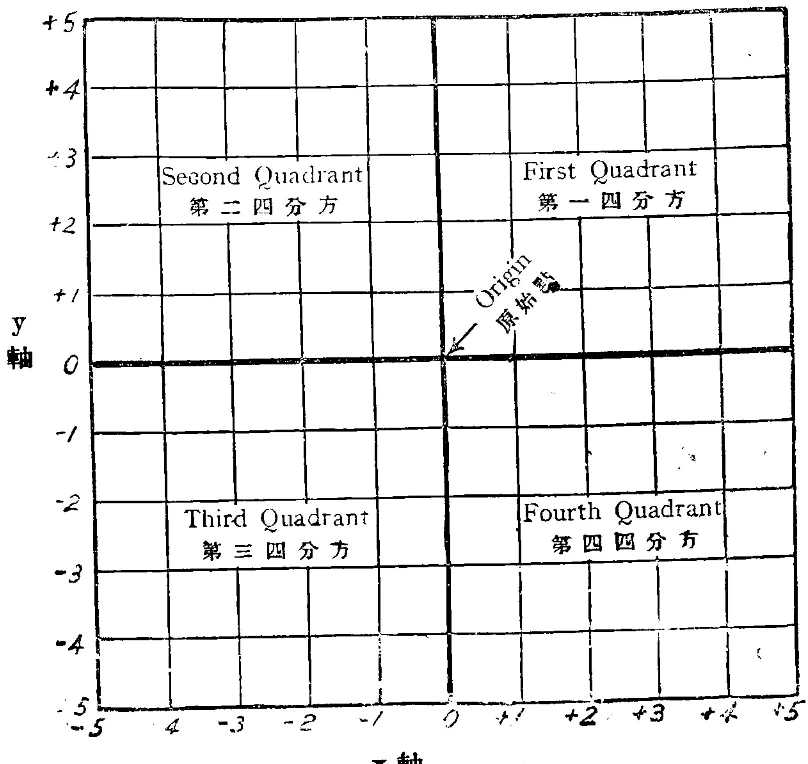
第八圖　測量單位迻譯圖

再思之,四位小數爲“.09”但不用紙筆計算,又覺不甚可靠。故心算之類此者,照下法試行,較爲簡易:將一整數,作爲直線視之,而.30爲其三分之一。.30之平方,約爲.30之三分之一,即約爲.10。如是較用算法爲易。.65之平方必在.43之鄰近或爲.44,因.65爲整數之三分之二。.65之三分之二,約在.44之鄰因.65之三分之一爲.22也。數目當以距離或空間視之。學者尤宜視圖線,次數表等爲有生氣,而寓事實於其中,若徒記方程式,則殆矣。

若兩軸有負數時,則第六圖可開展之如第九圖。比較

第六第九兩圖,**第六**圖卽爲**第九**圖之上右方。此方包入x與y軸之正數而稱之曰**第一四分方**(First Quadrant),如**第九**圖再加其他各四分方,則凡x與y之負數,均可表於此圖表示矣。

第九圖之x軸,有一量表自左而右,其原始點爲二軸最低之負數。y軸自下而上,其原始點爲最低之負數。此種圖之原始點在其中心,如圖。此中心爲二零軸之交切點。

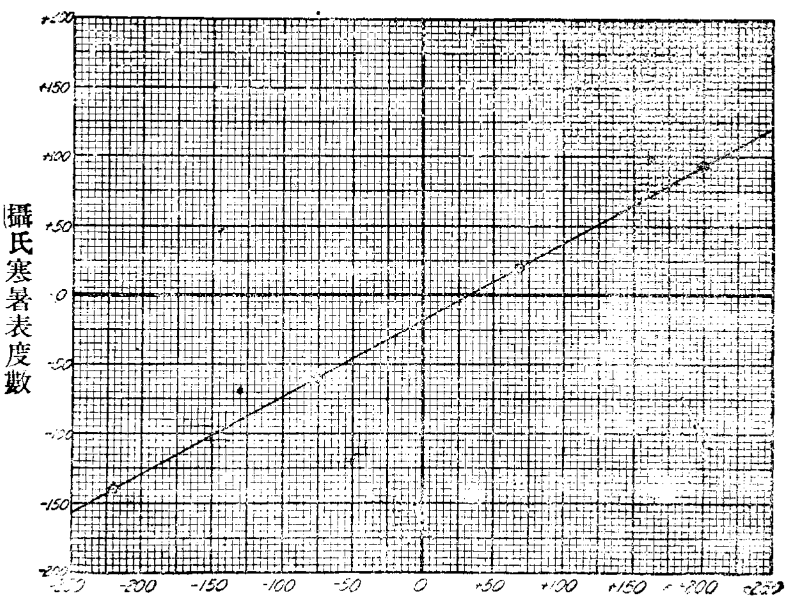


第九圖 四分方

現用四分方圖表示華氏與攝氏二寒暑表之關係,即知此表與彼表相當之圖數。第一步任取二表之對偶觀察數對如下:

x 華氏	68	–220	–76	199
y 攝氏	20	–140	–60	93

以上觀察,可繪在圖上,如第十圖。試比較上表之四個對偶觀察與第十圖用小圈表示之四點。一視此圖,即可不用計算將一種温度譯成他種。



第十圖　正負二數俱全之相關

例如華氏90度,即為攝氏30度。可用上圖證實。以上所

繪之相關,可用直線繪在方格紙上表之。此種相關謂之**直線相關**(Linear Relations)。

表示直線相關之直線,不必通過原始點。如 x 與 y 二變量同時爲零度,則此線通過原始點。若此線不通過原始點,吾人可斷定其一種變量爲零度時,其他種變量必非零度。

尙有二名詞爲繪圖時不可不知者。即一直線或一曲線交切 x 軸時,則其交切點爲 **x 交切**(x-intercept)。如第十圖,x 交切爲 32,意即 y 爲 o 時之 x 軸之價值。按第十圖之意義,即攝氏零度時,華氏爲 32。同法,**y 交切**(y-intercept)。即 x 爲 o 時之 y 軸之價值。按第十圖,y 交切約爲 −18 度。此即華氏零度時之攝氏溫度。

問題 1.　祇用觀察,不用圖式,試斷定以下各點在何四分方:

點	x	y
1.	+3.2	+6.9
2.	+5.3	−7.8
3.	−4.2	−1.7
4.	+0.3	+2.1
5.	−5.8	+1.0
6.	−3.4	0

問題 2.　下列對偶觀察,其中偶有錯誤,請繪圖以明之,並求其錯誤之所在。

x	y
7.2	0.6

4.4	10.0
9.5	2.9
2.6	12.5
8.7	4.0
1.6	14.0
7.5	5.7
7.3	11.0
6.2	7.5
1.5	1.5
5.0	9.2
5.0	15.5

問題3.　觀察第十圖,解答下問:

1. 華氏 −68 度時,攝氏幾度?
2. 若水沸時,華氏爲 212 度,問攝氏爲幾度?
3. 攝氏零度時,華氏爲幾度?
4. 華氏零度時,攝氏爲幾度?
5. 在幾度時,二氏之度數適相等?

問題4.　作一圖表示圓周與圓徑之關係。圓周之長,3.14 倍于圓徑。繪圖表示此二變量之關係時,圓徑之長,可達至 5 吋。請說明在作圖之前,吾人能斷定此線必經過原始點。表之廣袤,可用 $\frac{1}{4}$ 吋表出之。

問題5.　作一圖表示工作之時間與費用之關係。每工每點鐘需洋八角。每工材料需洋二元。用圖表示四點鐘工作之一切費用。是表上之時間,以十五分鐘爲一組距。

第五章

非直線相關

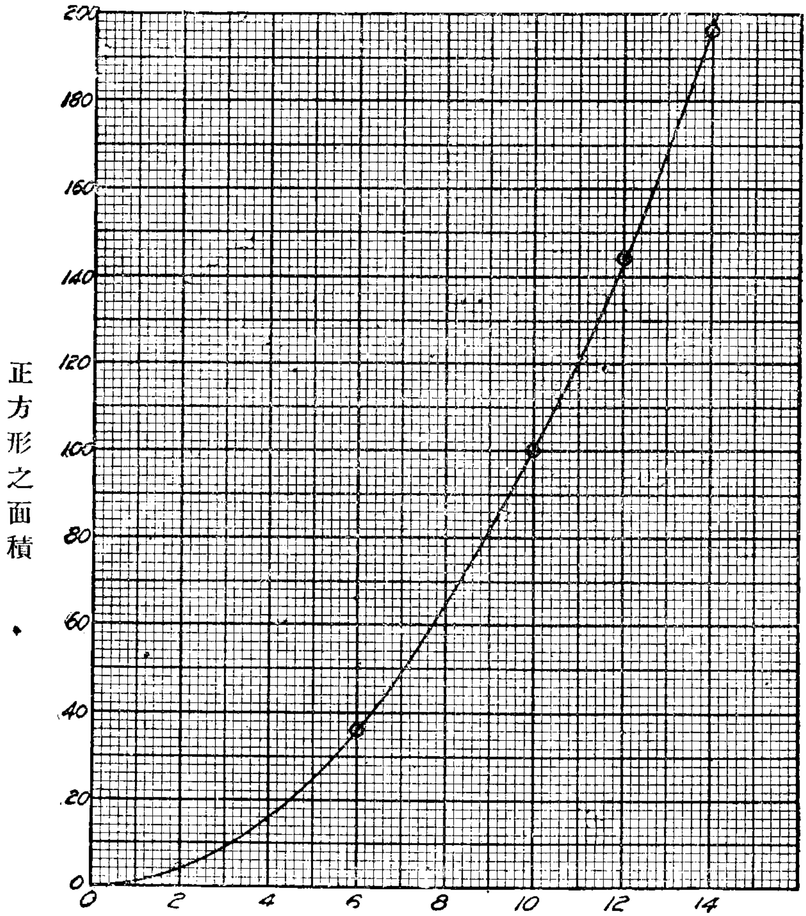
圖上之各點,有時非爲一直線而成一曲線。此種曲線之相關,謂之**非直線的**(Non-linear)。實則無論其線之爲直爲曲均稱**曲線之繪法**(Curve Plotting)。蓋"曲線"二字,已成爲圖上任何線之通稱。

試繪一正方形而求其邊與面積之相關。先表列數個觀察如下:

x	y
正方形之邊	正方形之面積
6	36
10	100
12	144
14	196

此四點之繪法,可用四小圈表示之。如**第十一圖**。此曲線必經過原始點,因零之平方仍等於零。如是則繪曲線,又多一點。惟點數愈多,則曲線愈爲精密,此固無疑義者。**第十一圖**之各點,不成一直線而成一曲線。若經過此數點畫一平勻之曲線,則其圖卽可表示一個數目與該數平方數之相關。

例如欲斷定6.8之平方時,先從**第十一圖**x軸上求出

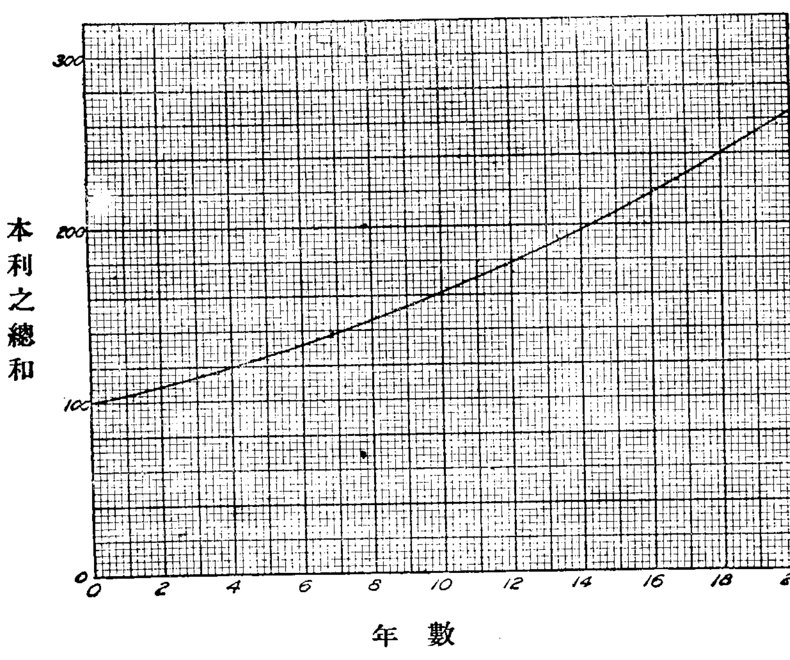


第十一圖　非直線相關

6.8 一點，再從此點向上推算，直至與曲線相交切處，此點在 y 軸上爲 46。故 46 即爲 6.8 之平方。吾人亦可根據此圖求數之平方根。例如欲求 78 之平方根時，則從 y 軸上得 78，向

右推算，至與曲線交切點爲8.8。此數約爲78之平方根。從視圖所得之結果，固不若從計算所得之精確，其優點能使人了解二變量之功用上之相關。此種了解，有時實較小數點之精確與否，尤爲重要。

第十二圖所示錢財如何可由複利而增加，亦足以表示非直線之相關。圖之x軸爲年數，y軸爲本利之總和。其曲線表示本銀百元逐年照利率五釐(5%)推算之增加。此圖示10年以後，則$100可增至$160。吾人若推求欲得一固定數目，須抽利幾年，亦可於此圖得之。例如14年後，則本利

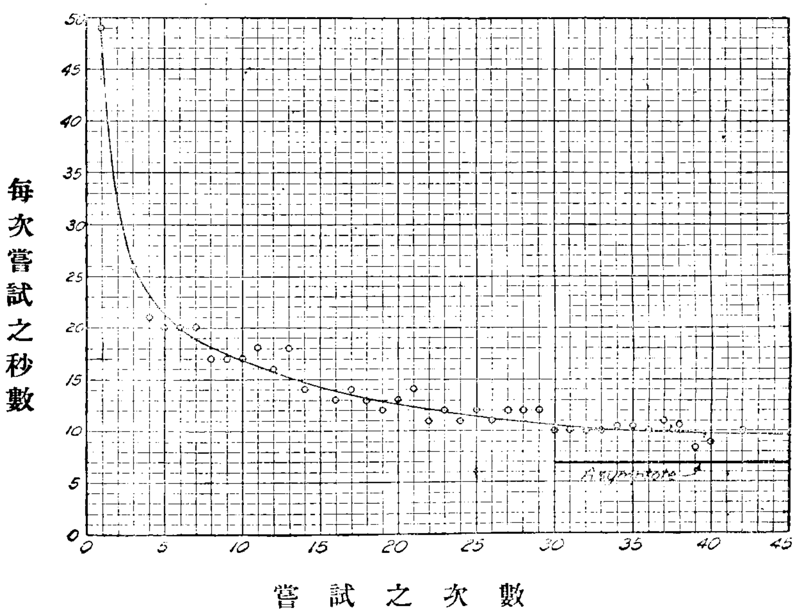


第十二圖　複利曲線

和可倍於本銀。

此種事實,由利息表求得者,固爲精確,但此圖能表示一種狀態而決非利息表所可表示者。此圖不但表示總數之逐漸增加,並表示增加率之增加。如最初四年之增加,祇$20,而第十六至第二十年之增加則爲$40。此曲線不但逐漸升高,且年數愈多,則其升高之度愈速。此種功用,決非利息表所能表出者。

第十三圖所示者爲鏡畫(mirror drawing)練習之進步。此種實驗,在心理學實驗時常用之。其內容爲鏡內反射之六角星影,使被試者看其倒影而描摹之。此試驗共做四十



第十三圖　代表實驗觀察之一曲線

二次,如圖所示。第一次嘗試費時最多。其實被試者之進步,全以其嘗試次數之多少爲準。嘗試愈多則時間愈少。

第十三圖之x軸,表示嘗試之次數。y軸表示每次嘗試之秒數。例如第一次嘗試費49秒鐘,如圖;第五次嘗試費20秒鐘;第三十次嘗試祇費10秒鐘。

全部結果用小圈繪於圖上,然後用二法完成之。(1)用短直線將各點連接之,其法與作次數多邊圖相同。(2)作一平勻曲線以代表被試者鏡畫練習進步之大概。

圖上各點所居之地位,半爲被試者增進之能力所使,半爲其他微小之影響所使,如分心,努力之分量,疲倦之情形等是。此圖之足以引起吾人注意之點,爲二變量之相關。卽練習之由嘗試次數而定者與能力之由每次所需時間而定者。若作平勻曲線,使此線上下之各點相等,則此線卽足以代表由練習而得之進步。此種平勻曲線比較上可以免除某日某時之複雜影響。第十三圖之平勻曲線,係用觀察法作成。

社會科學之事實與純粹科學之事實,以統計之方法言之,有一根本上之不同點。卽社會科學之測量,不若純粹科學之精細,且其所受各種外界之偶爾影響,亦較純粹科學爲多。因此吾人欲成立一種趨勢時,必修勻曲線以圖免除各種不相關分子之影響。

表示二變量之曲線圖旣作成,玆略作解釋如下。此圖

之曲線,一視而知為左高右低,意即謂鏡畫實驗初試之結果,較後試為多,蓋熟則生巧,所試時間,自應縮少。尚有一事,足以惹起注意者,即最初十次嘗試時間之縮減,較最後十次嘗試時間之縮減為多,此點若專看表上之材料而乏圖示之助,則不甚顯明。詳言之,在最初十次嘗試時,每次時間由49秒減至17秒或共32秒。自第30次至第40次嘗試時,十次內時間之減少,祇一秒鐘耳。繼續練習,必得進步,但進步常先速而後緩。此種狀態,閱圖即知,視表則知之較難。

第十三圖示被試時雖經四十二次之嘗試而仍未達到不可再進之境,欲達到此境,即再加練習,亦屬難能。觀乎曲線在四十二次之後,仍向下轉,即可知矣。欲知被試者熟諳此技後,鏡畫之速度如何,可將鏡取去,令其對物描寫。結果共需七分鐘故吾人可以假設被試者苟作無限期之鏡畫練習,其描摹之速度,必與無鏡時之描摹相等,即為七分鐘。故七分鐘上之粗線,可視為可近而不可達之限度,因名此線曰漸近線(Asymptote)。

問題1.　從實驗所得之事實,作一表,再作一圖與第十三圖相同。

問題2.　下表代表美國軍隊甲種(Army Alpha)智力測驗與軍隊乙種(Army Beta)智力測驗之分數。作一圖表示此種分數之相關,並修勻曲線使之平勻。

甲種	乙種
2	11
4	17

7	24
11	30
16	37
21	42
27	47
33	53
40	58
47	63
56	67
63	71
71	75
78	78
85	81
93	84
102	88
114	91
125	95
137	99
147	104
161	108

第六章

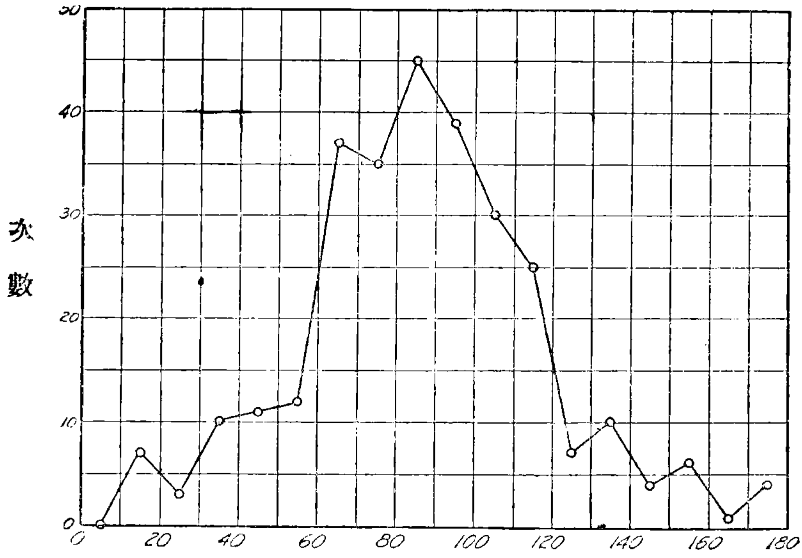
次數多邊圖之修勻法

次數多邊圖作成時,如數目不多,則曲線不能平勻。數目愈多,則曲線愈趨平勻而愈足以代表分配之眞象。照通常之例,凡次數多邊圖作成後,必加修勻,以示數目較多時應具之形狀。修勻之次數多邊圖,固不必與數目較多時之圖相脗合,然與未修勻時之圖相較,則妥善多矣。

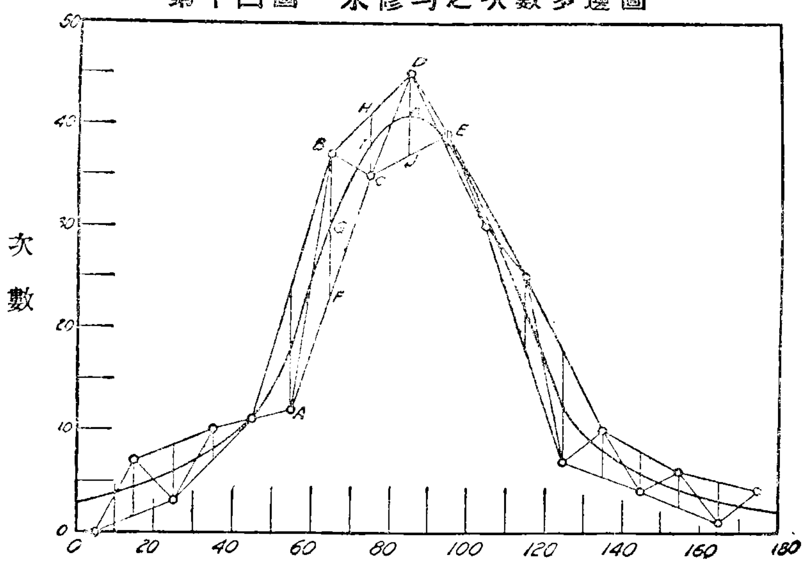
次數多邊圖上之各點,與其眞正價值之差異,或上或下,由其鄰近各點斷定之。試看**第五圖**組距 100–110 上之次數25,似乎太高,組距 90–100 上之次數23,似乎太低。若將此班學生重行同一之智力測驗,則 100–110 組距上之高峯,或不必如**第五圖**之所示者。修勻次數多邊圖之目的,蓋欲探知一切複雜變化除去後此圖應得之形狀。修勻之法甚多,玆所採取之一法,爲作圖與算術均屬便利者。

第十四圖之次數多邊圖,係直接從一次數表作成。**第十五圖**表示修勻之手續,**第十六圖**表示最後之修勻,可以作報告之用。

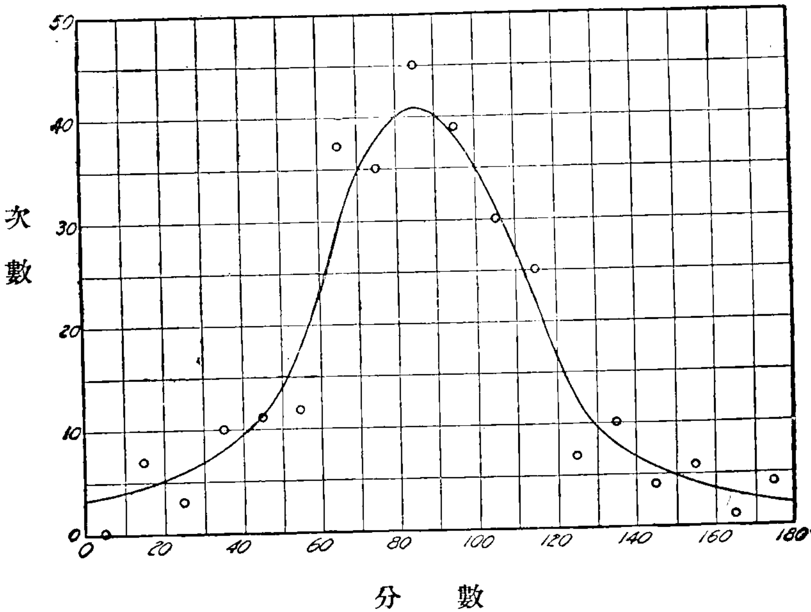
第十五圖之次數多邊圖,用直線連接小圈,此與**第十四圖**相同。爲說明修勻法之便利起見,有數小圈用字母註出。如注意C點。此點代表70–79一組之次數35。其地位均較低於鄰近二點37與45,此較低之理由,或爲偶然之原因



第十四圖　未修勻之次數多邊圖



第十五圖　次數多邊圖之修勻法



第十六圖　已修勻之次數多邊圖

所使。欲均衡C點之次數可用直線連接其鄰近二點得BD。由C點作垂直線,遇BD於H。求HC線之中點得A,其價值爲38,卽C之均衡次數。如是則C之次數,已與二鄰近次數互相適合,而免除偶然原因之影響。

欲均衡D點之次數45,可將其鄰近二點CE相連。由D點作垂直線與CE線遇於J。DJ之中點爲K,故80-89一組之均衡次數爲41。同法,可指定G點而得30,爲60-69一組之均衡次數,餘類推。經過各均衡次數作一曲線,如**第十五圖**。

若作一曲線經過均衡次數,或經過由實際觀察而得

之次數，則此曲線爲**次數曲線**(Frequency Curve)。**第十五圖**與**第十六圖**之曲線，即爲次數曲線。反之，若次數點用直線連接者，即爲次數多邊圖。

指定均衡次數，不必在度量上吹毛求疵。稍加練習，可以隨手畫之。蓋此種修勻方法，充其量亦不過求得最精確曲線之近似形耳。

由**第十五圖**改爲**第十六圖**，即次數曲線與原來用小圈表示之實得次數耳。**第十六圖**所示者，爲報告時應用之次數曲線。但次數多邊圖亦未嘗不可用，即不甚雅觀之直方圖，亦有用之者。

若用次數曲線之報告時，最好利用小圈將實得次數表出，庶讀者一閱即知原來事實與次數曲線差異之程度。**第十六圖**即表示此意。小圈示實際事實，曲線示此事實在理想中之情形。

以上修勻方法，若用算式代之，可述之如下：例如欲修勻**第十五圖** C 點之次數 35，即將 C 點之次數加倍 $(2\times35=70)$，並加上其鄰近二次數$(37+45)$，得 152；以 4 除之，得均衡次數 38。此種均衡次數，最好不必用表列出，以免讀者誤認其爲實得次數，實則非也。

修勻次數多邊圖之又一法，爲增大其組距。蓋組距增大，則其所包含之次數亦多，因之凡一切微小變化，可以免去。若次數表用數個組距大小不同之圖表示之，則其差異

立見。

問題1.　下列次數表爲一班學生之智力測驗分數。試用此項事實作三種圖:一直方圖,一次數多邊圖,一次數曲線圖。在次數曲線圖上,用小圈代表實際次數,如第十六圖。試問此曲線有無漸近線?同時並將各修勻線一一畫出。

分數	次數
22	1
21	1
20	1
19	6
18	16
17	13
16	31
15	47
14	52
13	62
12	70
11	62
10	67
9	57
8	58
7	41
6	42
5	29
4	12
3	11
2	4
1	5
0	2

問題2. 下表所載爲一班工程學生之某種測驗分數。作三個次數多邊圖,其組距爲1, 2, 與4。在每圖組距中點上作小圈。注意分數之爲5者,意卽答對五個問題或兼答第六問題之一部分。作此三圖時,用同樣之x與y量表。

討論此三個多邊圖之曲線,孰較平勻,孰較易作,但不必將曲線畫出。如是可以看出組距愈大,則曲線愈平勻而愈加繼續不斷。

此三圖之x與y量表相同,且代表一樣數目,何以其大小則不相同。

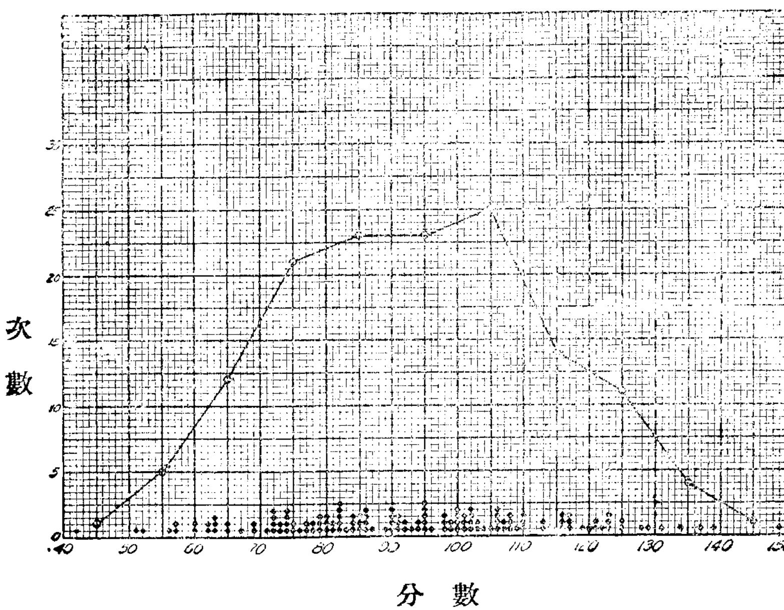
分數	次數
0	0
1	1
2	3
3	6
4	19
5	21
6	31
7	24
8	30
9	41
10	43
11	35
12	46
13	37
14	41
15	44
16	14
17	19
18	21
19	5
20	1
21	0
22	1

第七章

圖示法

次數表不過是一種方法,用之可以達到一種目的。照常例言之,次數表爲次數多邊圖之初步。有時爲節省時間起見,次數多邊圖卽可根據原有事實(raw data)作成,不必經過次數表之助。如第十七圖,全根據第一表與第一圖作成,中間並未得次數表之助。其法如下:

將原有事實上之數目,載在圖上,每讀一個數目時,在x軸上作一點。第一表之第一數爲62。在62上之第一橫線



第十七圖　圖示法

上作一點。x 軸上切不可作此點,因此軸代表次數之爲零者。第二數爲123。在 x 軸 123 上之第一橫線上作一點。以後62又發現一次。因第一橫線上已有一點,故在第二橫線上再作一點。全部數目讀完時,則**第十七圖**之各點成矣。一視此圖卽知82分共有五個,102 分共有四個,96分祇有一個,餘類推。橫線之用,能使讀者不數點數,而知次數。在實用上,各點不必如**第十七圖**所載者之大。若圖不付印,則鉛筆點已甚著明。

50	64	70	68	64	72	65	55	79	55
75	48	60	51	55	46	51	66	62	58
47	72	47	65	74	67	55	49	46	65
51	58	63	54	60	62	82	61	73	50
51	68	83	70	54	63	54	77	51	70
77	50	65	64	59	66	65	51	55	63
46	48	79	67	82	72	57	65	58	72
66	62	58	68	52	58	59	78	66	48
74	73	53	61	62	73	67	60	48	64
46	57	60	77	78	53	51	55	68	49
72	50	52	59	58	60	68	63	53	57
59	83	67	65	55	59	51	60	61	58
44	51	78	64	62	50	57	67	69	55
66	68	61	57	68	61	59	50	60	56
48	57	65	54	59	65	76	64	54	64
60	67	58	66	63	54	63	60	62	51
56	48	73	60	57	73	52	56	58	47
78	64	52	52	56	58	68	52	77	56
54	59	63	65	67	63	60	58	46	60
56	60	54	61	57	50	66	49	47	49
54	54	75	46	60	74	58	72	56	43
70	60	44	56	63	45	56	60	44	63
42	46	64	61	67	40	62	64	71	37
63	69	43	33						

第三表　某班學生之智力測驗分數

各點作成後,應計其數,與**第一表**相比較。第二步之手續爲選擇組距。**第十七圖**之組距爲十。多邊圖之量表,不必與各表之量表相同。欲定每組之次數,惟有計算該組之點數而已。**第十七圖**之 100－109 組共有25點。此組之次數卽用該組中點高度25上之小圈表示之;其餘各小圈,用同一方法作成,然後連以直線而成次數多邊圖。

此法之便利,卽一種組距不合式時,立刻可變換一組距,而另作一圖。且不必覆查原有事實。但作點時,切宜謹愼而免遺漏。

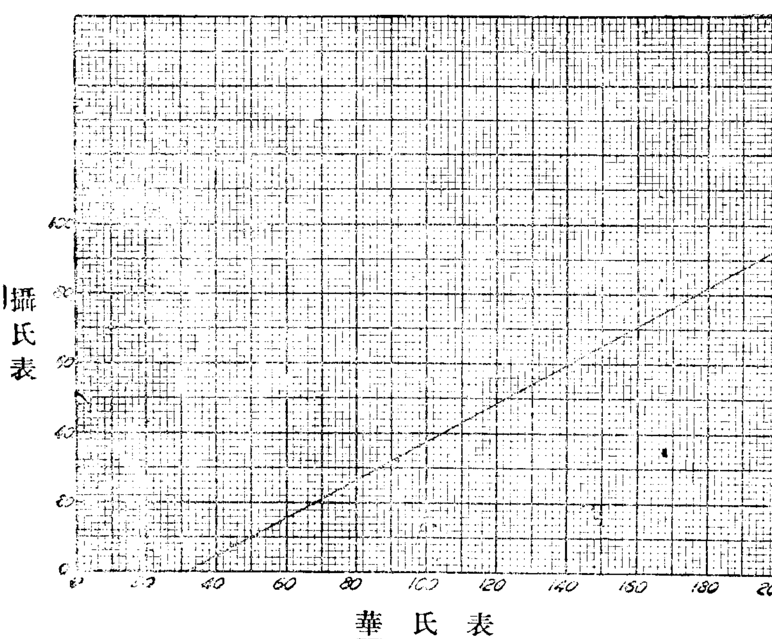
問題1.　第三表代表一班學生之智力測驗分數,將此事實載在方格紙上。以十單位爲一組,作一次數多邊圖。根據此圖作一次數表而以五單位爲一組。

第八章

經過原始點之直線方程式

任何科學之探討,均含二種或二種以上變量相關之研究。量的相關,可用不同方法表出之。

1. 二變量之相關,可用口述。例如研究攝氏與華氏寒暑表之相關,可口述之如下:"攝氏冰點爲零度,沸點爲百度。華氏冰點爲32度沸點爲 212 度。攝氏與華氏之度數成比例" 若將攝氏58度改爲華氏,而欲從口述得之,殊感不便。



第十八圖 表示一方程式之圖

2.　二變量之相關,亦可用圖示,如**第十八圖**。在此圖上,華氏與攝氏之冰點或沸點之交切點可用直線連接之。若已知攝氏爲58度,而欲譯爲華氏,閱圖卽知,卽137度,此用圖表示二變量相關之優點也。

3.　二變量之相關,亦可用方程式表示。攝氏華氏二寒暑表相關之方程式,可書之如下:

$$C = .55F - 17.7$$

在此式內,C爲攝氏之度數,F爲華氏之度數。此方程式之來歷,不久卽加討論。若攝氏58度欲譯爲華氏時,卽以58代入上式之C求F:

$$C = .55F - 17.7$$

$$58 + 17.7 = .55F$$

$$\frac{58 + 17.7}{.55} = F$$

$$F = 137$$

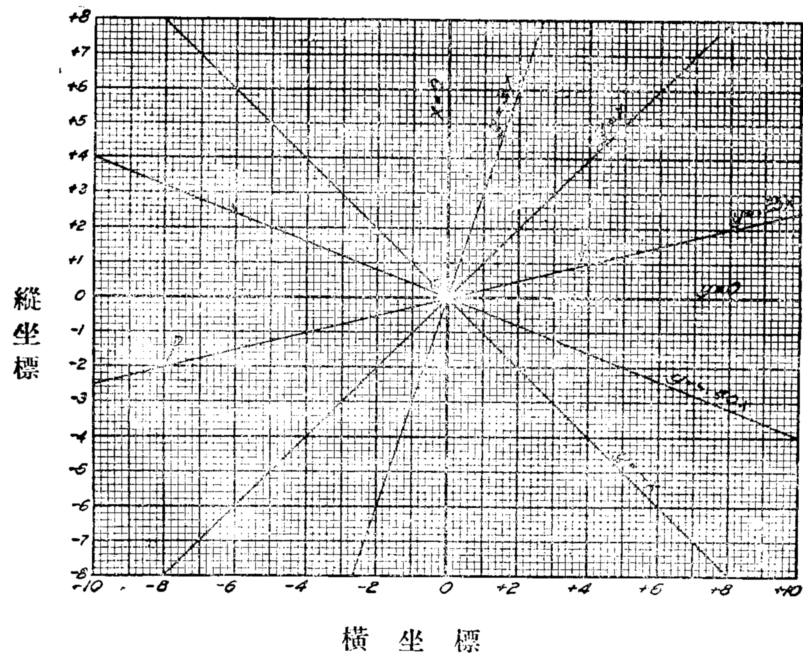
圖示與方程式所用之方法不同,而所得之結果則一。圖示法較爲敏捷而欠精細;方程式所佔之地位較少,且較爲精密。

經過原始點之直線。其普通公式爲

$$y = ax,$$

在此式內,a爲**常倍數**(Multiplying Constant)。**第十九圖**內,此種公式,共有七個。先注意代表$y = x$公式之直線。其意爲在此線上之任何點,y之價值常等於x之價值。例如$x =$

+5則 $y=+5$ 而 $(x=+5)$ $(y=+5)$ 之一點卽在線上。餘類推。



第十九圖　經過原始點之直線及其方程式

再討論代表 $y=.25x$ 方程式之線。按此公式 $x=6$ 時,則 $y=1.5$。而 $(x=6, y=1.5)$ 之一點,可從此線上求得。學者可援例尋求,無論用何數目,代入 x,卽可求得 y,或代入 y,卽可求得 x。若圖較小,而不能包括較大之數目,可將圖放大。

第十九圖之各線不同,皆由常倍數而定。常倍數亦名線之傾斜度。其名稱甚爲適當,蓋此數愈大,則其傾斜度亦愈大。試比較三線,其傾斜度爲3, 1, .25一望而知線之傾斜度爲3者,較線之爲.25者,其傾斜度爲大。此種觀察可推至

第二與第四之四分方上之二線。不過在此二線上，x與y之價值，適相反耳。

橫軸或x軸之方程式爲$y=o$，因在該線上之y常爲o。縱軸或y軸之方程式爲$x=o$，因在該線上之x常爲o。凡適合該方程式之線，稱之曰該線之**軌跡**(Locus)凡x與y之對偶價值之足以適合此方程式者，均於此方程式之軌跡上之一點表示之。所謂適合者，卽方程式中之二數，等於x與y價值之代入者。

線可根據方程式而作。方程式亦可由線而定。若一線經過原始點O，則其相當之方程式，可由此線之傾斜度而定。其法如下：作任何直角三角形ABC，以所假定之線爲弦。三角形之股AC與x軸相並行。此種三角形可作於任何線上，且不論大小。試量BC與AC。此線之傾斜度爲$\frac{BC}{AC}$。BC之距離爲1.5，AC之距離爲6。其比例爲$\frac{1.5}{6}$或.25，此卽此線之傾斜度。欲增加此傾斜度測量之精確，可將此三角形放大。如用DBE三角形，則其形較大矣。其傾斜度爲$\frac{BE}{DE}=\frac{4.5}{18}$或.25，如前。

歸宿以上各點，得結果如下：(1)圖上之線可用方程式代表之；(2)$y=ax$之方程式，可常用一經過原始點之線代表之，其傾斜度由常倍數a而定；(3)無論何線經過原始點者，均可得$y=ax$式之方程式。由此觀之，圖示法與方程式，可謂異途同歸。

問題1. 作一具有四分方之圖,如第十九圖,並作以下各線:

1. $y=+1.5x$

2. $x=-2y$

3. $x=\frac{2}{3}y$

4. $y=-.15x$

5. 作一線而包含以下各點,($x=+2$, $y=+6$);($x=o$, $y=o$)。斷定此線之方程式。

在以上之任何一方程式內,假定二變量中之一價值而解求其他變量之價值,並在圖上指定其地點。表示每次此點均在其適當之軌跡上。

問題2. 研究第十第十八兩圖,並示此二圖形狀雖異,而其所表示之關係則一。

第九章

直線之普通方程式

本章所述者,爲圖上直線之普通方程式。其式如下:

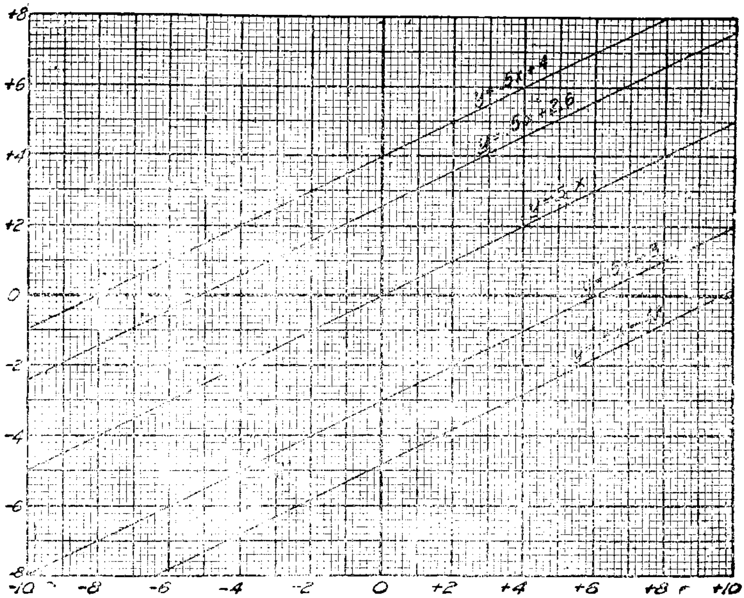
$$y=ax+b$$

在此公式內,a 與 b 爲常數, x 與 y 爲二變量。前章曾經提及 a 爲常倍數並能斷定線之傾斜度。b 爲**常加數**(Additive Constant),並能斷定此線在何處與 y 軸交切。在**第二十圖**內,共有五線,而各有其相當之方程式。此五線均各平行。且此五線之常倍數,或傾斜度 a,均相同,即 +.5。次注意五線之常加數均各不同。實則常加數均等於 y 交切。其方程式 $y=.5x+4$ 之線,其常加數爲 +4,此與該線之 y 交切 +4 相同。證以**第二十圖**上之其他各常加數與 y 交切亦相同。

細察**第二十圖**,即知凡常加數爲正號時,則線在原始點上交切 y 軸;此數爲負號時,則線在原始點下交切 y 軸。線經過原始點時,則常加數爲零度,而其方程式亦較簡明,即 $y=.5x+o$, 或 $y=.5x$, 如**第二十圖**之所示者。

在**第二十一圖**上,任意劃數線,可以察圖而定其方程式。先論 A 線。吾人已知直線之普通方程式爲 $y=ax+b$。故必先斷定方程式內二常數之價值。至於常加數即爲 y 交切。按圖 F 點在 y 量表上看爲 +2,故常數 b 爲 +2。

縱坐標



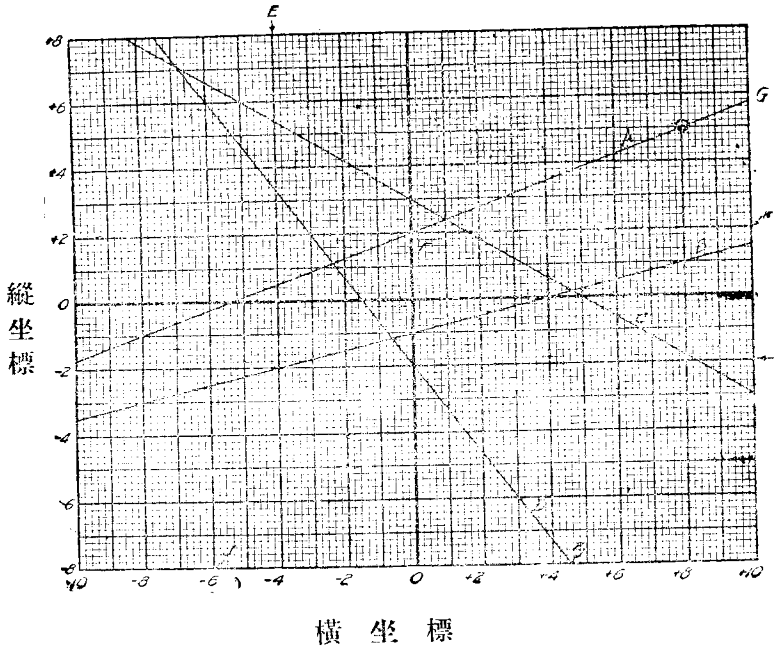
橫坐標

第二十圖 平行線與其方程式

傾斜度 a 之斷定法,前已略述。$\frac{GH}{FH}$ 之比例,按 x 與 y 量表看,即爲 $\frac{3.7}{10}$ 或 .37。A 線之全方程式爲:

$$y = .37 + 2$$

凡線向左傾斜時,則其傾斜度爲正數,凡線向右傾斜時,則其傾斜度爲負數。凡線爲地平線時,如 M 線,則其傾斜度爲零,而 x 項消除。故 M 線之方程式爲 $y = ox + b$, 或 $y = b$。常加數 b 之價值,可從圖中求得即 -2。故 M 線之方程式爲 $y = -2$。意即 y 在 M 線上之價值,均爲 -2,即視圖亦甚明



第二十一圖　直線之方程式可由觀察而得者

瞭。

關於A線之方程式,尚可用此線上之任何點或用小圈表示之一點而證實之。此點之二數爲(x=+8, y=+5)。將此數代入A線之方程式:

$$y=.37x+2$$

$$5=.37\times 8+2$$

$$5=2.96+2$$

此方程式二項數目之相等,幾等於由圖而得者。若取

非A線上之任何一點,則其方程式之二項數目,必不相等。

E線之方程式為$x=-4$,換言之,凡E線上各點之x價值,均為-4。細察第二十一圖,可證明之。

D線之方程式,亦可同樣決定。其y交切為-2,故常數b為-2。D線之傾斜度之求法如前,即該線為弦之正角三角形之勾股比例。若用IJK三角形,則其比例為$\frac{IJ}{JK}$或$\frac{14}{-10.5}$或-1.33。故傾斜度為-1.33,D線之方程式為

$$y=-1.33x-2$$

此方程式,可用前法證實,即假定x為任何價值,如-3。代入此值於方程式內而求y之價值,得$+2$。在圖上指定此點$(x=-3, y=+2)$並注意此點適在D線上。在圖上指定某點時,常先寫x值後寫y值。故書明x與y實不甚重要。通常$(x=-3, y=+2)$之一點,可用$(-3, +2)$表示之。

同法,試證實c線之方程式為$y=-.6x+3$。方程式中常數之符號,關係甚為重要,宜注意之。學者宜一視方程式之傾斜度與y交切,即能默想一直線之方程式之性質。又y交切與x交切之比例,等於傾斜度,此點亦宜注意。

根據直線而寫出其方程式,其法已經說明。根據方程式,如$y=ax+b$,亦可作直線於圖上。其法甚簡,即假定數個變量之價值,而求其他相當變量之價值。將各點作在圖上,而以直線連接之。通常二點即能成一直線,然為免除錯誤

起見,至少要用三點。

偶有一種方程式,驟視之,似非一直線之方程式,而實則可以直線之方程式視之。試舉例以明之。在下列方程式內,

$$y=2x+3+2$$

如將二加項相加,使方程式變為 $y=2x+5$, 則此式可以直線之方程式視之。又例 $4y=8x+6$,可約為 $y=2x+1.5$,而以直線之方程式視之。又 $3x=8-4y+2x$, 亦為直線之方程式。

吾人所宜注意者,即 y 為方程式之一項,而 x 項與相加項,為方程式之又一項。如此敍述,可謂顯用 y 項表示者 (explicitly in terms of y)。其法如下:

$$3x=8-4y+2x$$

$$x=8-4y$$

$$4y=-x+8$$

$$y=-.25x+2$$

最後一式,與通常直線之方程式相同,其傾斜度為 $-.25$ y 交切為 $+2$。

總結. 前二章所述原則,歸納如下:

1. 圖上之每線,均可用下列方程式代表之, $y=ax+b$。

2. 線上之傾斜度,可用常加數a表示之。若線向左傾斜,則此數爲正。若線向右傾斜,則此數爲負,若線爲地平線時,則此數爲零。此數可由y交切與x交切之比例而定,或由圖上三角形之二股而定。測量時係用x與y量表,而決非用某種方格紙上所具之長度。

3. 常加數表示y交切。倘線在原始點上與y軸交,則此數爲正。倘線在原始點下與y軸交切,則此數爲負。

4. 倘二直線之方程式具相同之常倍數,其軌跡必並行。倘二直線之方程式具相同之常加數,則其線同在y軸上之一點交切。

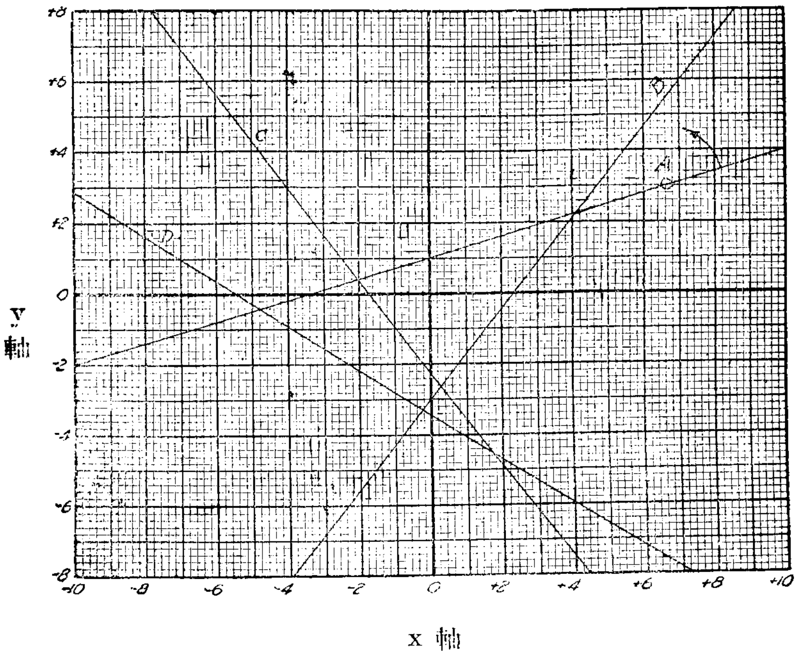
問題1. 在方格紙上,作四分方圖,並繪下列各方程式所示之線。

1. $y=.6x+3.4$
2. $y=4.2x+5$
3. $2.5x=-4.7-3y$
4. $14-9+2x-1.8y=0$
5. $(y-1.6)+(x+.7)=4$

將上列各方程式,在繪線以前,根據$y=ax+b$式重寫之。繪線以後,假定每線上之任一點,並表示其相當方程式與該點之數的價值相適合。

方程式1與3,以聯立方程式解之,並決定所得x與y二價值與二方程式相適合。表明此解答代表圖上二線之交切點。

問題2. 將第二十二圖之各線,重作於方格上,並求各線之方程式。



第二十二圖　用於第九章第二問題

若A線以小圈A處爲中點,而依箭頭之方向旋轉,則常倍數有何變更?常加數有何變更?在何種情形之下,直線之以整數爲常倍數者,可與x軸成一45度角?

問題3.　觀察下列各方程式後,答以下各問,但不宜畫圖。

1.　代表二平行線者,其方程式爲何?

2.　交切y軸之最高處者爲何線?

3.　交切x軸之最左處者爲何線?欲解答此問,將零代入y,卽以心算解求x交切。

4.　經過原始點者爲何線?

5. 向右傾斜者爲何線?

6. 若各線作成後,以何線爲最傾斜?

1. $y=4x+3$

2. $y=4x+9$

3. $y=2x+0$

4. $y=-4x-3$

5. $y=-2x+0$

第十章

算術平均數

設施行智力測驗於二班學生,而欲知何者較為聰明,則必求二班之平均數。平均數較高者,其智力亦較高。求平均數之法,即將分數相加,得總數,再以分數之總次數除之,所得商數即為平均數。為讀者便利及明晰起見,可用下式:

$$m=\frac{\Sigma S}{n},$$

在此式內,m 為分數之平均數,ΣS 為分數之總和,n 為此班之人數。統計學上之通常平均數,均稱為**算術平均數**(Arithmetic Mean)蓋別於他種平均數而言也。

試細思計算平均數之故何在,必曰欲得一數以代表全數欲得一分數以代表全部分數。但有時平均數之數目常與其所代表之各數,無一相同者。例如 2, 4, 6, 8, 之平均數為 5。吾人用之以代表其他四數,然此四數中,無一為 5 者。

凡一數用之以代表數目則此數為全數之**集中趨勢**(Central Tendency)。集中趨勢乃一單個數目或量表上之一點,而最足以代表全部數目者。其數常居於全部數目之全距離之中間。算術平均數或普通平均數,為集中趨勢之最通用者。茲將計算此數之四種方法述之。此四法大致相同。

因其所得之結果,亦無甚差異,故何時應用何法,全憑以下情形:計算機之有無;全部數目之多少;數目之大小;數目之歸類方法。算術平均數爲最通用之平均數,故亦簡稱爲**平均數**(Mean)。其他尚有各種平均數爲**倒數平均數**(Harmonic Mean)。與**幾何平均數**(Geometric Mean)。

數目未經列表時**計算平均數之法**. 第一表爲若干學生智力測驗之分數。求平均分數最便利之法,將各數用加數機 (adding machine) 加好。以分之次數 140 除總數 12,900,得 92.785,即爲平均分數。學習統計者,若有機器在旁,宜盡力用之,庶心算可以減少,而注全力於問題之解答。蓋解答問題,非機器所能爲也。

求平均數之簡便公式如下:

$$m=\frac{\Sigma X}{n}$$

在此式內, m 爲平均數, X 爲數列中之任何一數, ΣX 爲各 X 數之總和,而 n 爲各數之次數。符號 Σ 用時甚多,宜切記之。此號宜讀爲"和"。如 ΣX 爲 X 數之和。故以上公式應讀爲"平均數等於各 X 量數之和,再以各數之次數除之"

用次數表計算平均數之法. 各分數已表列後,如第四表,則用表計算平均數,較爲便利。蓋如是各個數目,不必一一讀出。其手續先預備一表,如表。第一行爲組距。第二行爲次數。第三行爲組距中點。第四行爲第二第三行之積數。

平均數可照下式決定:

$$m=\frac{\Sigma fX_m}{n},$$

在此式內,m為平均數,ΣfX_m為第四行之積數,而n為次數。注意ΣfX_m與ΣX無異。初學者常生誤會。fX_m為每組各x數之和。即某組中點,以該組之乘數乘之。故ΣfX_m為各組距X數之總和。亦即ΣX之意義。然用ΣfX而不用ΣX者,因各組之次數不同,用之以示區別也,照此方法,平均數為93.64與前所得之絕對平均數92.78,相差無幾。

分數	f	X_m	fX_m
40－49	1	45	45
50－59	5	55	275
60－69	12	65	780
70－79	21	75	1,575
80－89	23	85	1,955
90－99	23	95	2,185
100－109	25	105	2,625
110－119	14	115	1,610
120－129	11	125	1,375
130－139	4	135	540
140－149	1	145	145
	n＝140		13,110＝ΣfX_m

$$m=\frac{\Sigma fX_m}{n}$$

$$=\frac{13110}{140}$$

$$m=93.64$$

第四表　用次數表計算平均數之法

用等值量表計算平均數之法. 第四表之積數,其數目過大,若無計算機之助,則計算太繁。苟能用較少之數目,而得同一之結果,尤爲利便。在統計法上,常用**等值量表**(Equivalent Scale),即計算時以較小之數替代較大之數,俟得到結果後,再應用校正數,使等值量表仍變爲原來事實。試將**第四表**之事實,用等值量表法,求平均數(如**第五表**),如是則計算時,數目可以較小。

若不用X量表而用等值量表則平均數爲,

$$m_e=\frac{\Sigma fE}{n}=\frac{681}{140}=4.864$$

但平均數須用X量表表示,故可用以下之關係求之。

$$\begin{aligned} m &= 10\ m_e+45 \\ &= 48.64+45=93.64 \end{aligned}$$

由等值量表變爲X量表時,其不同之點有二。一觀**第五表**之第三第四縱行,即知等值量表所用之單位,較X量表所用者爲小,又等值量表從零起點,而X量表則從45起點。職是之故,所得之結果,必經變更。若數目繁冗,非用計算機不可者,則用等值量表足矣。

用假設原始點計算平均數之法. 此法最爲普通,以後計算相關時,亦用此法求平均數。用等值量表法,計算時數目縮小。但若將零數排在量表之中間,則數目更可縮小。

分數	f	X_m	E	fE
40－49	1	45	0	0
50－59	5	55	1	5
60－69	12	65	2	24
70－79	21	75	3	63
80－89	23	85	4	92
90－99	23	95	5	115
100－109	25	105	6	150
110－119	14	115	7	98
120－129	11	125	8	88
130－139	4	135	9	36
140－149	1	145	10	10
	140			681

(1) $m_e = \frac{\Sigma fE}{n}$,

在此式內，m_e 爲平均數之用等值量表表示者。

$m_e = \frac{681}{140} = 4864$

(2) $c = I \times m_e$,

在此式內，I 爲 X 量表單位之數目。

$c = 10 \times 4.864 = 48.64$

(3) $m = c + m_a$,

在此式內，m_a 爲 X 表之組距中點，此數在等值量表內，用零表示。

$m = 48.64 + 45 = 93.64$

第五表　用等值量表計算平均數之法

在 X 量表上而置等值量表之零數之一點，謂之假設原始點(Arbitrary Origin)或(Assumed Origin)。此點可任置何處而無害於精確，但以放置於 X 量表之中間爲最妥，因其能使所用之數目，縮爲最小故也。

第六表 表示用假設原始點計算平均數之步驟。其事實與上述者相同。計算法如下:

分數	f	E	fE
40—49	1	−5	−5
50—59	5	−4	−20
60—69	12	−3	−36
70—79	21	−2	−42
80—89	23	−1	−23
90—99	23	0	$-126=\Sigma fE_{負}$
100—109	25	+1	+25
110—119	14	+2	+28
120—129	11	+3	+33
130—139	4	+4	+16
140—149	1	+5	+5
	140		$+107=\Sigma fE_{正}$

(1) $\Sigma fE_{pos}+\Sigma fE_{neg}=\Sigma fE=+107-126=-19$

(2) $\frac{I.\Sigma fE}{n}=c$

$\frac{-19\times 10}{140}=-1.36$

(3) $m_a+c=m$

$95-1.36=93.64$

第六表 用假設原始點計算平均數之法

第六表之第一縱行爲組距。第二縱行爲次數。第三縱行爲等值量表。此量表之作法,卽假設零點,在X量表全距離之中間之一組上。如是使此點上下之次數大略相等。估計此零點時,祇須大致不差,無損於結果之精確也。在X量表上各組數目之大於零者,用 +1, +2, +3, 等表示之。反之,數目較小之各組,用 −1, −2, −3, 等表示之。在第六表上,假

設原始點爲95,此點爲90－99一組之中點。100－109之一組爲較大之一組,故在等值量表上,卽用＋1表示之。80－89之一組爲較小之一組,故在等值量表上,卽用－1表示之。

第四縱行含有積數fE,卽第二第三縱行之積數。正的積數,fE, $\Sigma fE_{正}$ 爲＋107。負的積數fE, $\Sigma fE_{負}$ 爲－126。總次數n,爲f縱行之和。

第二步計算ΣfE,此 $\Sigma fE_{正}$ 與 $\Sigma fE_{負}$ 之和。照本例ΣfE爲－19。旣得以上各數,卽可按照下式,計算**校正數**(Correction),c。

$$c=\frac{I\Sigma fE}{n},$$

在此式內,c爲校正數。此數爲眞正平均數,m與假設平均數,m_a之差數。

n＝總次數

I＝X量表上之組距單位。照本例此數爲10。旣決定校正數,c後,眞正平均數,可照下式求得:

$$m=m_a+c$$

計算本章章末之問題時,儘可不必參考說明之文,而依據**第四,五,六,表**所載極簡明之格式。

在未計算各種平均數之前,玆將平均數之要素,略加解釋。假設作一直方圖於硬紙板或他種實體上,並將此圖之外線剪出。若以現在討論之事實爲例,則此剪出之形,將

與第二圖中粗線所表示之形相同。現以此板之底,衡於刀鋒之上,左右移動,直至均衡,則其均衡之點,即第六表所得之絕對平均數點也。用此法以說明平均數,其觀念之了解,實勝於公式多矣。

硬紙板之相等長方形代表個人。長方形之闊,代表組距。長方形之高,代表次數一。吾人宜注意者,即每個人所居之地位,離刀鋒愈遠,則平均數在該邊所"稱"(Weigh)愈重。直方圖之小長方形所代表之個人,均有使此紙板在其所居之一邊稱重之趨勢,此趨勢謂之個人之重距(Moment)。嚴格言之,重距爲面積或體積及其與刀鋒之距相乘之積。因各個長方形均相等,故個人之重距,由其所居之地位至平均數間之距離而定。因之凡居於兩極之個人,不啻有決定平均數之權。凡在平均數上之個人,全是中立。在實際上,即量表兩極端之數個個人,對於平均數之影響極大,而中間之數個個人,對於平均數之影響極小。

既得平均數,宜一視X量表,而察其地位是否大約居中。若其地位在全距離之外者,則必有計算上之錯誤。

問題1. 參看第三表,用本章所詳四種方法求平均數。

第十一章

中數

中數(Median)爲測量集中趨勢之又一法。若有若干數目,由小而大,依次排列,則其最中間之一數,謂之中數。故在X量表上,中數前後之量數,其數目相等。算術平均數卽無此種情形。有時測得集中趨勢時中數較算術平均數爲尤佳。

試看下例:

7　14　11　2　17　1　22　13　9

將以上數目,由小而大,依次排列,則得

1　2　7　9　11　13　14　17　22

中數爲11,因在此由小而大之數目中,11爲其最中之一數。有四個數目比中數大,有四個數目比中數小。算術平均數爲10.7。以上一例,其數目爲奇數而非偶數。若數目爲偶數時,則中數在二個最中數目之中間。此層與中數定義,亦甚符合。

下列數目,

11　29　8　4　10　3　17　37　22　7

可重爲排列如下:

3 4 **7** 8 10 11 17 22 29 37

中間二數為10與11。中數在二數之中間,即10.5。算術平均數為14.8。

事實歸類時,則計算中數之法,應照**第七表**。此表之事實與**第一圖**同。

X 量表	f 次數	遞加次數	
40–49	1	1	
50–59	5	6	
60–69	12	18	
70–79	21	39	
80–89	23	62	
90–99	23	85	8個數較中數低=a 15個數較中數高=b
100–109	25	110	
110–119	14	124	
120–129	11	135	
130–139	4	139	
140–149	1	140	
	n=140		

中數$=90+\frac{8\times10}{23}=90+3.48=93.48$

第七表　中數之計算法

以下為計算中數之步驟:

1. 第一第二兩縱行,與以前次數表相同。
2. 第三縱行為遞加次數。第三縱行之任何一組距

上,如 70－79,記載該組與該組以下各組之次數之和,如 $1+5+12+21=39$。

3. 計算次數之總和爲140。求 $\frac{n}{2}$ 得70。

4. 求含有中數之組距。此組距之遞加次數含有 $\frac{n}{2}$ 即90－99之一組。

5. 將此組之次數,分爲 a 與 b 二部,使較小各組距之遞加次數加上 a 等於 $\frac{n}{2}$。按本例爲 $62+8=70$。故 $a=b$; 且 $a+b=23$, 因知 b 必爲15。

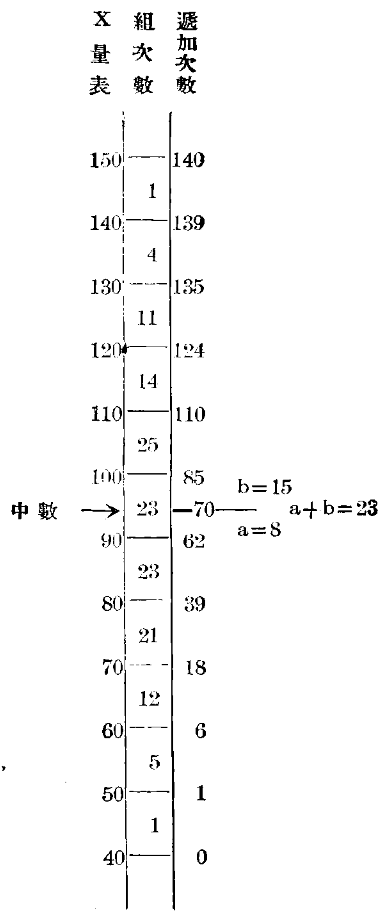
6. 設 X 表之單位爲 I。則中數等於含有中數之低限度加上 $I\frac{a}{f}$ (f 爲該組之次數)。

計算之法,甚爲簡短,一察**第七表**,即易於了解。

第二十三圖說明計算中數之法中數在90－99之一組內。該組共有23個次數。假設此23個次數平均分配於該組上。吾人已知中數爲 X 量表上之一點,其上下次數均相等。從 X 量表最低處數起,至70次爲止,而注意此點在 X 量表上之地位。此點即爲中數。數完6個次數時,在 X 量表上,適爲60。數完39個次數時,在 X 量表上,適爲80。數完62個次數時,在 X 量表上,適爲90。若直至量表之100,則次數已至85,但祗須至次數70,故中數在90－99之一組內。

當數至70個次數時,在90－99組距內之23個次數,適過去8個,假設此23個次數平均分配於組距之上,則中數在量表90以外之一組之 $\frac{8}{23}$ 距離處。每組共有10單位,故 X

量表上之中數,爲90加上$\frac{10\times 8}{23}$。若計算之法,由上而下其理論相同,結果亦相同,如**第二十三圖**。



$$\text{中數}=90+\frac{8\times 10}{23}=93.48$$

$$\text{中數}=100-\frac{15\times 10}{23}=93.48$$

第二十三圖　中數之計算法

在一直方圖上,若經過中數點,作一垂直線,則此圖分為兩部,其面積亦相等。

問題1. 計算第二表所載事實之中數。用二法,一由上算下,一由下算上。且看其答數是否相同。

第十二章

衆數

衆數(Mode)爲測量集中趨勢之第三方法。在X量表上之曲線,以衆數上之一點爲最高。若事實依組距歸類,而每類各有次數,則含有次數最多之一組之組中點,即爲衆數。此假設尙無謬誤,大致可用。故欲求衆數,無計算之必要,若欲用修勻曲線之法,而得其較精確之價値,則又是一問題。然通常此法不用。

茲將三種集中趨勢之定義與其特性總括之。

算術平均數爲通常平均數。若有一列數目,以各數目之次數,除其總和,卽爲算術平均數。作直方圖於硬紙上,將此圖剪出,則此圖之算術平均數點,可均衡於刀鋒之上。

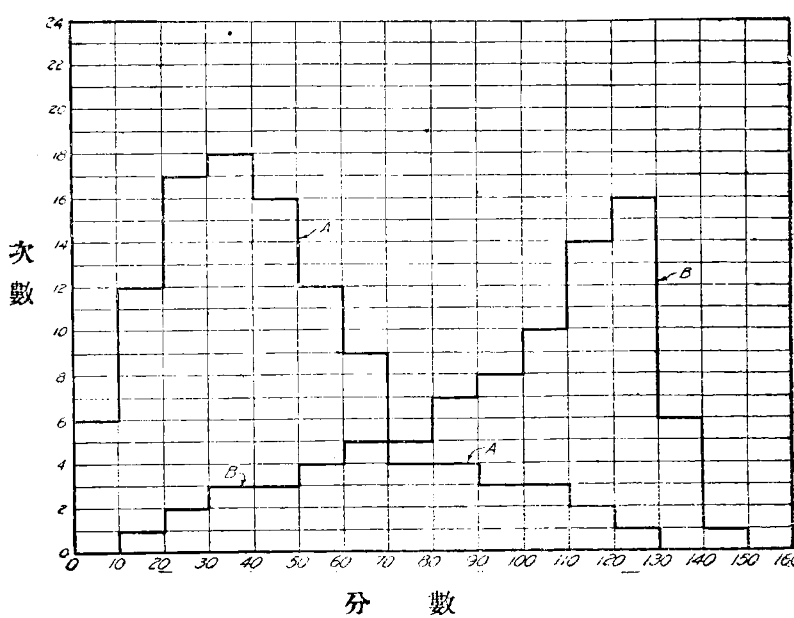
中數爲若干數目中由小而大排列時之中間一個。中數前後之數目相等。有時初學者,以中數爲全距離之中間一點,此定義實不甚精確。

衆數爲量表上一價値,在此數上之次數曲線爲最高。換言之,衆數爲X價値,在此點上之曲線,得最高之縱線。若將事實排成一直方圖,則含有最大次數之組距之中點,卽爲衆數。

凡表示集中趨勢之量數,均爲X量表上之一點。有時次數曲線受複雜影響,致生二個或二個以上之高峯。若高

峯有二個,且其成立,均非爲機遇所使者,則此曲線謂之雙峯(Bimodal)。曲線上之雙峯,或爲有意義者,抑爲機遇之變化所使者,實爲統計上應行討究之問題。

若次數曲線爲**不對稱**(Asymmetrical)時,如**第二十四圖**,此曲線稱爲**偏態**(Skewed)。若衆數兩邊之組距內次數兩兩相稱,則此曲線稱爲**對稱**(Symmetrical)。曲線可爲對稱而同時非爲常態。請看第二十四圖之A,右端特長,而B則左端特長,故A稱爲**正偏態**(Skewed Positive),而B稱爲**負偏態**(Skewed Negative)。A之所以稱爲正偏態者,以其特長之端,在X量表之上端。B之所以稱爲負偏態者,以其特長之



第二十四圖　偏態次數曲線

端,在X量表之下端。

如果分配對稱,則三種集中趨勢之量數,均在X量表之一點。

若分配不對稱,則三種集中趨勢之量數,在X量表之不同點上。然則吾人應用三種不同之集中量數,其理由蓋亦明矣。當曲線偏態時,此三種均各有其重要處也。

問題1. 求第二十四圖AB二曲線之算術平均數,中數,與衆數。作一圖,表示此三種不同之集中趨勢量數在一偏態曲線上之彼此地位,並示此三種量數地位之關係,是與其性質相符合。

第十三章

離中趨勢

茲章所論,爲統計學原則上之又一觀念。集中趨勢爲量表上足以代表全部分配之一點,前章已經討論。集中趨勢共有數種,而最通用之三種,爲算術平均數,中數,與衆數。吾人若欲簡略說明一個分配,則指定其集中趨勢足矣。若吾人討論若干人之薪金問題,則知其平均數爲 \$3000, 已大致無差。固然,此若干人之薪金有多於 \$3000, 有少於 \$3000,但其平均必爲 \$3000 無疑。欲知其詳,必知薪金之多於平均數者或少於平均數者,其分布如何,此離中趨勢法之不可或缺也。蓋離中趨勢法者,乃表示集中趨勢數前後量數差異之情形。若已知量數之集中趨勢數,而又知此數前後各數分布之情形,則吾人對於此種量數之分配,可謂略悉一二。本章及以後所討論者,爲下列四種之表示離中趨勢法:

1. **全距離**(Range)
2. **平均差**(Mean Deviation)
3. **二十五分差**(Quartile Deviation)
4. **均方差或標準差**(Standard Deviation)

全距離(Range). 試比較二列數目,並察其集中趨勢與離中趨勢。

a.	5	10	15	20	25	30	35	40	45
b.	21	22	23	24	25	26	27	28	29

二列數目之平均數各爲25;故就集中趨勢論,二列數目相同。惟第一列數目較第二列數目爲分散。第一列數目之全距離爲 45－5＝40,而第二列數目之全距離爲 29－21＝8。故全距離爲離中趨勢之一種量數。若欲形容此二列量數,而同時免除詳述之手續,則可表出之如下:

	集中趨勢	離中趨勢
a 列	平均數＝25	全距離＝40
b 列	平均數＝25	全距離＝ 8

一視上式,卽知其意義之所在。例如細察二數列,卽知數目之爲30者,均在每列平均數之上,然此數離 a 列之平均數較近,離 b 列之平均數較遠。

全距離有一極短處,卽全距離祇由一列中之最大與最小二數而定,不甚可靠。若 a 與 b 二列數目代表智力測驗分數,則其全距離全賴二極端之數目。二端中間之各數目毫無效力,請看下列:

c.	5	22	23	24	25	26	27	28	45

c 列與 a 列之全距離相同,但除極端二個數目外,其數目之分散,較 b 爲尤甚。故欲測量離中趨勢,不得不用他法。

平均差(Mean deviation).再看下列數目:

x 爲數列:	4	7	9	10	11	11	12	13	13	14	15	17	20
d 爲差數:	8	5	3	2	1	1	0	1	1	2	3	5	8

$\Sigma x=156$　　x之算術平均數$=12$

$\Sigma d=40$　　d之算術平均數$=3.08$

第一行爲一列數目,自4至20,其全距離爲16,算術平均數爲12。第二行爲每一數目與算術平均數12之相差,不計符號。故15與平均數12之差數爲3; 20之差數爲8; 9之差數爲3;而12之差數爲0。

各數之和$\Sigma x=156$,算術平均數爲$\frac{\Sigma x}{n}=12$。差數之和$\Sigma d=40$,平均差爲$\frac{\Sigma d}{n}=3.08$。平均差數爲各差數之平均數,但差數不計符號。平均差數大時,則平均數前後之數目,分散較大;平均差數小時,則平均數前後之數目分數較小。

平均差數可從各種集中趨勢算出,從算術平均數與中數算出均可。故平均差數究從何種集中趨勢數求得,務宜表出。上例之平均差數,則從算術平均數算出。若求平均差數於已經歸類之量數,則其算法如**第八表**。

1. 表列各組距,組距中點及各組距之次數。
2. 將次數行相加而求總次數n。
3. 求fx之積於下一行。

4. 將 fx 行相而求 Σfx。

5. 從下式求算術平均數。

$$算術平均數=\frac{\Sigma fx}{n}。$$

6. 在下一行上,求組距中點與算術平均數之差數 d。

7. 表列各 fd 之積。

8. 將 fd 行相加而求 Σfd

9. 從下式求平均差。

$$平均差=\frac{\Sigma fd}{n}。$$

以上算法,若用計算機,最爲便利。若數目過多而又乏計算機,則用計算平均數時所用之等值量表爲佳。

集中趨勢爲量表上之一點 (point),前已提及。離中趨勢爲量表上之一距離,爲本章所應注意者。

問題 1. 比較**第二十五圖**三分配之算術平均數與平均差。從每分配之算術平均數上,求平均差。報告時須注意:

1. 表示**第二十五圖**各分配之算術平均數,可不用計算,而一看卽知。(注意三分配圖均爲對稱)。

2. 不用計算表示 A 與 B 二多邊圖,何者之離中趨勢爲大。

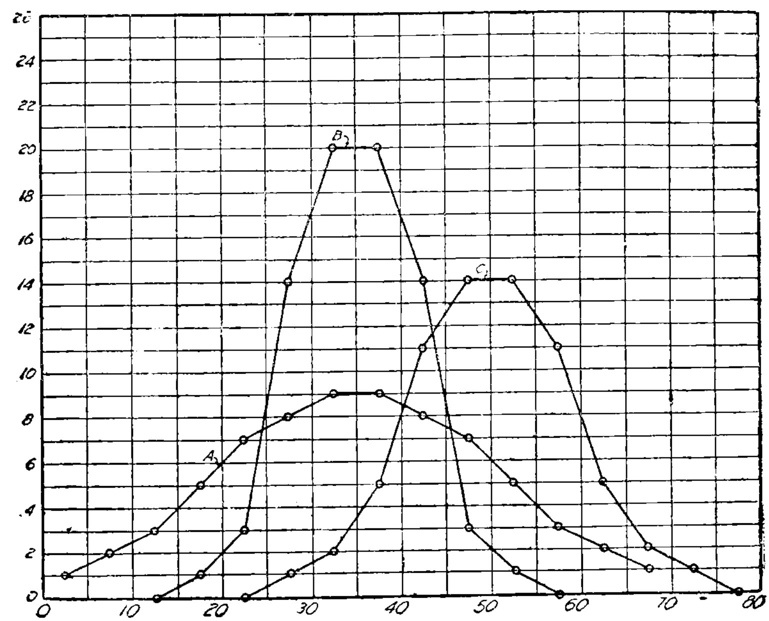
3. 不用計算表示 B 與 C 二多邊圖,何者之離中趨勢爲大。

4. 將 A, B, C, 三多邊圖之事實,作成三表,與**第八表**相同。組距與次數,可從圖中求得。計算**第二十五圖**之算術平均數與平均差。將答數

與問題1中三問之答數相比較。

5.　若將多邊圖B,向右移動而不變易其形,則此圖有何變更?

6.　試用全距離法,比較此三圖之離中趨勢。試問此三圖之次序,用此法所得者,是否與用平均差所得之次序相同?



第二十五圖. 三個多邊圖,其集中趨勢與離中趨勢均不同

7.　假定500人身長之平均數與平均差已經算出,試問增至1000人,則平均數與平均差有何變更?

問題2.　試用等值量表法,計算第八表之平均差,假定原始點,可在量表之中間。將公式寫出,用作者自已符號。表示所得結果,與第八表同。

組距	中點	次數		與平均數之差數	
	x	f	fx	d	fd
40—49	45	1	45	56.3	56.3
50—59	55	3	165	46.3	138.9
60—99	65	6	390	36.3	217.8
70—79	75	12	900	26.3	315.6
80—89	85	18	1,530	16.3	293.4
90—99	95	34	3,230	6.3	214.2
100—109	105	28	2,940	3.7	103.6
110—119	115	17	1,955	13.7	232.9
120—129	125	12	1,500	23.7	284.4
130—139	135	8	1,080	33.7	269.6
140—149	145	4	580	43.7	174.8
150—159	155	2	310	53.7	107.4
160—169	165	1	165	63.7	63.7
		n=146	Σfx=14,790		Σfd=2472.6

算術平均數$=\frac{\Sigma fx}{n}=\frac{14790}{146}=101.3.$ 平均差$=\frac{\Sigma fd}{n}=\frac{2472.6}{146}=16.94.$

第八表 平均差之計算法

第十四章

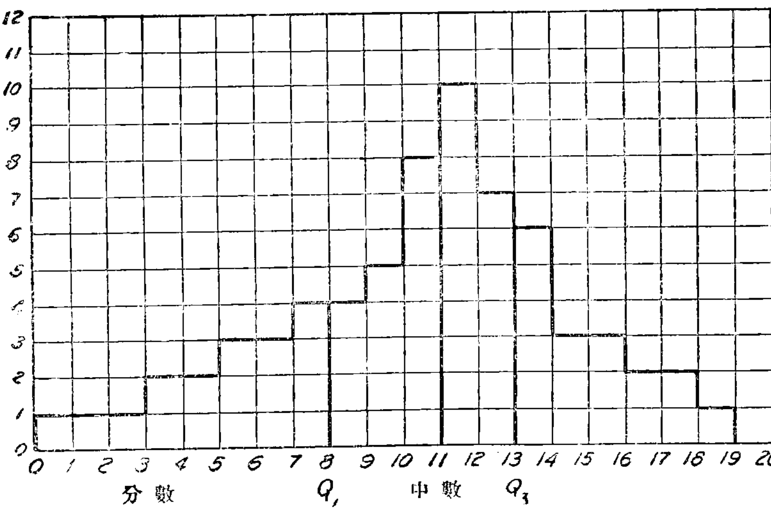
二十五分值

離中趨勢之量數,如全距離及平均差,已經討論。尚有一種,其應用較爲普通者,爲**二十五分差**(Quartile Deviation 或 Semi-interquartile Range)。吾人觀察一種次數分配而欲專注意於其離中之程度,則最捷之法,爲全距離。若全距離大,則量數之離中趨勢亦大。但用全距離而測量離中趨勢,法非盡美,因全距離祇由量數中之最大與最小二數而定。若此極端二數,一有變更,則全距離亦因之而變更矣。

第二十六圖之直方圖,共有68個數目。其中數爲11,因此數之上下,共有34個數目。其全距離爲19。此數之外,尚有一法足以測量分配中間百分之五十(34個)數目之分布情形,亦可用爲離中趨勢之量數。從中數向上數四分之一面積,17個數目,得**上二十五分值**(Upper Quartile),通用Q_3爲符號。看第二十六圖,上二十五分值,卽在X量表13處。同法,從中數向下數四分之一面積,17個數目,得**下二十五分值**(Lower Quartile),在量表8處。此數通用Q_1表出之。中數偶有用Q_2表出者。

此直方圖已分成相等四部;卽0－8,8－11,11－13,13－19。每部各有17個數目。中間之一半數目,在上二十五分值Q_3(卽X量表13處)與下二十五分值,Q_1(卽X量表8處)之間。

X 量表上之 Q_3 與 Q_1 中間 5 個單位爲二十五分距。此亦爲離中趨勢量數之一種，較之全距離爲通用，且易求而固定也。二十五分距係由全部數目而定，故不致受一二單獨量數之影響。



第二十六圖　二十五分點

由中數而至最低量數之距爲 11。由中數而至最高量數之距爲 19－11，或 8。故此直方圖向下偏或其偏爲負。此點亦可由下二十五分距（中點 － Q_1）大於上二十五分距（Q_3－ 中點）看出。故二十五分值之位置，不特足以表出離中之趨勢，且足以看出偏態之程度與方向。

但表示離中趨勢時，多不用二十五分距，而用二十五分差，即二十五分距之一半也。第二十六圖之二十五分差

爲2.5。

分配爲對稱時,則上與下二十五分距相等;Q_1與Q_3至中數之距亦相等,而中數爲全距離之中心。在此種分配時,上二十五分距下二十五分距,與二十五分差均相等。

現試另舉一例而求二十五分值之各常數。第二十六圖,不含小數,蓋以便利計算,但從實際上言,二十五分值鮮有適爲整數者。計算二十五分點之法,與計算中數之法相同。實則中點亦爲二十五分點之一種耳。

二十五分值之計算法. 排列事實,如**第二十七圖**。在第一行上,表列組距。在第二行上,表列次數。將次數行之各數相加得總次數,n 爲143。以4除之,得35.75,卽二十五分值。

試求含有此三個二十五分點之組距。其法卽從任一端數,得35.75爲第一個二十五分值。以2×35.75或71.5爲中數;以3×35.75或107.25爲又一個二十五分值。

組距之含有此三個二十五分點者,旣求得後,將次數分爲相等四部,如**第二十七圖**括弧之所示者。

求每組距內之二十五分點,如**第二十七圖**所示。計算時有一假設:卽每組距內之數目,均平均分布於此組上。故上二十五分點必近於100–109組之上部,因此組之次數,25,分爲二部,4.75屬於上二十五分,其餘20.25屬於下二十

五分。故上二十五分點爲該組上部之$\frac{4.75}{25}$。

故上二十五分點$=110-\frac{4.75\times10}{25}=110-1.9=108.1$ 其餘各二十五分點,可同理求之。

組距	次數		
x－scale	f		
140－149	1		
130－139	5		35.75
120－129	11		
110－119	14		
100－109	25	4.75 20.25	35.75
90－99	24	15.50 8.50	
80－89	23		35.75
70－79	21	4.25 16.75	
60－69	12		
50－59	5		35.75
40－49	2		

上二十五分值 $Q_3=110-\frac{4.75\times10}{25}=110-1.9=108.1$

中數 $Q_2=100-\frac{15.5\times10}{24}=100-6.5=93.5$

下二十五分值 $Q_1=80-\frac{4.25\times10}{21}=80-2.0=78.0$

第二十七圖 二十五分值之計算法

既得二十五分點,可按照下式,求二十五分差各常數:

二十五分距 $=Q_3-Q_1=108.1-78.0=30.1$

二十五分差 $=\frac{Q_3-Q_1}{2}=\frac{108.1-78.0}{2}=15.05$

上二十五分距 $=Q_3-Q_2=108.1-93.5=14.6$

下二十五分距 $=Q_2-Q_1=93.5-78.0=15.5$

茲將二十五分值特性總結如下:

1. 三個二十五分點將全分配分爲四個相等部分。每個二十五分值含有同一數目,即全分配之四分之一。三個二十五分值所分成之四個面積均相等。看**第二十七圖**。

2. 若將分配上之各個數目,另寫紙片上,而收此許多紙片,擲入帽內,則隨意拾起之數目,或在上二十五分值與下二十五分值之內,或在二者之外,其機遇相等。此理甚明,因在此上與下二十五分點內之數目,與在外之數目相等也。

3. 二十五分值各常數之任何一數,均足以表示集中或離中之趨勢。

問題1. 下表爲二團學生智力測驗分數之分配。此二團可稱之爲A與B。計算每團之中數,上二十五分距,下二十五分距,二十五分差,且用此數種常數,比較此二團之人。用此二分配作二個次數多邊圖於一張紙上,并表出二十五分值之各常數。

組距	次數 A團	次數 B團
20—29	0	0
30—39	2	0
40—49	4	2
50—59	4	4
60—69	6	8
70—79	10	12
80—89	16	16
90—99	14	8
100—109	10	6
110—119	8	4
120—129	2	4
130—139	2	4
140—149	0	4
150—159	0	2
160—169	0	2
170—179	0	2
180—185	0	0

第十五章

均方差

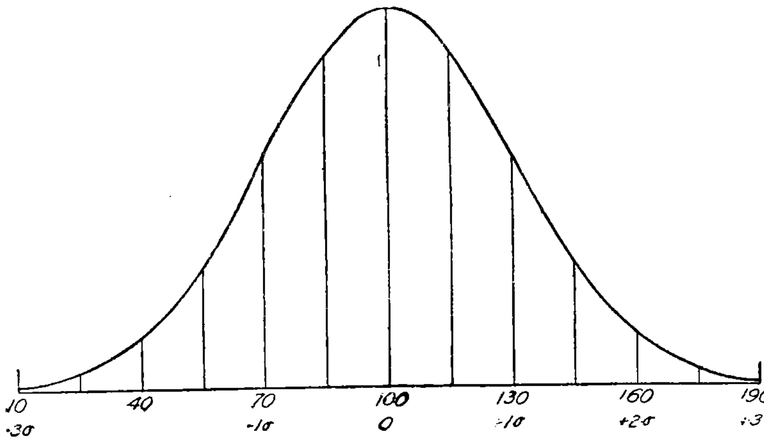
茲章所論,爲離中趨勢量數之最通用者,名曰**均方差**又稱**標準差**(Standard Deviation)。平均差爲不計符號之各差數之平均數。欲求均方差,先將差數平方之,然後再加;所得之總數,以總次數除之;將商數開方,即爲均方差。最後開方一步,甚爲重要,否則其數之大小,與原來之量數,不能相等。

均方差通用"式克碼"(Sigma),或 σ, 之符號表示,其公式如下:

$$\sigma=\sqrt{\frac{\Sigma d^2}{n}}$$

在此式內, d 代表量數與平均數之差數, Σd_2 代表此項差數之和之方。均方差可用代數式計算,此非他種差數所能及者。

第二十八圖代表一個次數曲線,其平均數爲 100, 均方差爲 30。其意即全數之三分之二均在 70 與 130 之內。均方差爲一距離,與他種差數相同。**第二十八圖**橫坐標上,繪有 X 量表及其相當之均方差,或式克碼。數目 70 可用 -1σ 表之; 130 可用 $+1\sigma$ 表之; 115 可用 $+.5\sigma$ 表之; 40 可用 -2σ 表之。餘類推。



第二十八圖　次數曲線之用均方差為底線上測量之單位者

凡一次數分配,有極可注意之二事—— 一為集中趨勢,即X量表上之一點,一為離中趨勢,即X量表上之一距離。集中趨勢表示X量表上各數之大概位置。例如有一薪水表,究其大多數集中於$100抑$200。離中趨勢表示各薪水對於平均薪水相離之遠近。若專提一個數目論,如**第二十八圖**之160者,則欲知此數與他數之關係,必先知此圖上分配之集中趨勢與離中趨勢。如160一個數目,以 $+2\sigma$ 表之,則二事可以明瞭,一則此正符號表示此數在平均線之上,二則 2σ 表示此數遠在平均數之上。對稱分配包含六個 σ,三正,三負。就理論言,平均數上下之數,可展至無窮,但在一個對稱分配上, $+3\sigma$ 與 -3σ 包括全數之百分之九十九(99%)。

若有二個次數分配而比較其離中趨勢,則均方差大者,其離中趨勢亦大,均方差小者,其離中趨勢亦小。均方差之特性,將來當在常態次數曲線內,詳加討論。

#	X	d	d^2
1	5	0.7	0.49
2	7	1.3	1.69
3	3	2.7	7.29
4	6	0.3	0.09
5	7	1.3	1.69
6	4	1.7	2.89
7	1	4.7	22.09
8	5	0.7	0.49
9	6	0.3	0.09
10	4	1.7	2.89
11	5	0.7	0.49
12	7	1.3	1.69
13	2	3.7	13.69
14	6	0.3	0.09
15	6	0.3	0.09
16	3	2.7	7.29
17	10	4.3	18.49
18	5	0.7	0.49
19	4	1.7	2.89
20	5	0.7	0.49
21	7	1.3	1.69
22	12	6.3	39.69
23	8	2.3	5.29
24	4	1.7	2.89
25	7	1.3	1.69
26	6	0.3	0.09
27	5	0.7	0.49
28	8	2.3	5.29
29	8	2.3	5.29
	$\Sigma X=166$		$\Sigma d^2=147.81$

算術平均數 $m=\frac{\Sigma X}{n}=\frac{166}{29}=5.72$

$$\sigma=\sqrt{\frac{\Sigma d^2}{n}}=\sqrt{\frac{147.81}{29}}=2.26$$

第九表　無組距時計算均方差之法

均方差在統計上之用途極廣,計算均方差之法亦甚

多。茲擇其最要者三種,述之如下:

1. 無組距時

2. 有組距及假定原始點者

3. 用原來數目表示者

三種方法,究用何種,可由個人之取捨。計算機之有無,及事實之排列法而定。

1. **無組距時計算均方差之法.** **第九表**所示者,爲計算均方差最簡單之法,毫無取巧之處。第一行所列者爲次第,共有29個。第二行所列者爲分數。其和爲 $\Sigma X=166$。用29除166得5.7.爲算術平均數。第三行爲各數與此平均數之差而不計及符號。第四行爲各差數之平方其和爲 $\Sigma d_2=147.81$。其餘公式之應如何代入,詳**第九表**。上述算法,若次數太多,或數目太大,則計算時殊不便利。

2. **有組距及假定原始點計算均方差之法.** **第十表**表示計算均方差之又一法。此法之應用最爲普通。第一行爲組距。第二行爲其相當次數。f行之和,爲次數總和, n,卽140。

第二步爲假定一原始點,大約在分配之中間。若有一組照吾人之臆度,與平均數最爲切近者,則該組可用爲假定原始點,而以零表之。其實原始點可在任何處,卽在量表之外,亦無不可,但此點與平均數最近時,則計算最爲簡短。按**第十表**以90－99之一組爲零,在d行上,卽以o表之。至

於 d 行上之其他位置,均用 1, 2, 3, 4,等填入,而列於零組距之兩面。故差數均用組距表示,而非用原來單位表示。組距之高於零組距者,以正差數表之,組距之低於零組距者,以負差數表之.看**第十圖** d 行。

fd 行爲 f 行與 d 行之積。fd 行負項之和爲 $\Sigma fd_{負}$,卽 −126。fd 行正項之和爲 $\Sigma fd_{正}$,卽 107。此二項之差爲 Σfd,卽 −190。

fd^2 行爲 d 行與 fd 行之積。此行之和爲 Σfd^2, 卽 589。用假定或估計原始點時,則計算時必用校正數 c。其求法與用法詳**第十表**。

組距	f	d	fd	fd²
40−49	1	−5	−5	25
50−59	5	−4	−20	80
60−69	12	−3	−36	108
70−79	21	−2	−42	84
80−89	21	−1	−23	23
			−126	
90−99	23	0		
100−109	25	+1	+25	25
110−119	14	+2	+28	56
120−129	11	+3	+33	99
130−139	4	+4	+16	64
140−149	1	+5	+5	25
	140		+107	589

$n = 140$
$\Sigma fd+ = 107$
$\Sigma fd- = 126$
$\Sigma fd = -19$
$\Sigma fd^2 = 589$

$$c = \frac{\Sigma fd}{n} = \frac{-19}{140} = -0.136$$

$c^2 = 0.018$

$$\sigma = \sqrt{\frac{\Sigma fd^2}{n} - c^2}\ 組距$$
$$= \sqrt{\frac{589}{140} - 0.02}\ 組距$$
$$= 2.04\ 組距$$
$$= 20.4\ 單位$$

第十表. 有組距及假定原始點時計算均方差之法

本例之均方差爲2.04組距,但每組距共有十單位,故均方差爲2.04**組距**(class intervals)或20.4**量表單位**(scale units)

#	X	X^2
1	5	25
2	7	49
3	3	9
4	6	36
5	7	49
6	4	16
7	1	1
8	5	25
9	6	36
10	4	16
11	5	25
12	7	49
13	2	4
14	6	36
15	6	36
16	3	9
17	10	100
18	5	25
19	4	16
20	5	25
21	7	49
22	12	144
23	8	64
24	4	16
25	7	49
26	6	36
27	5	25
28	8	64
29	8	64
	166	1098

$$\Sigma X = 166$$

$$n = 29$$

$$m = \frac{\Sigma X}{n} = \frac{166}{29} = 5.72$$

$$m^2 = 32.7$$

$$\Sigma X^2 = 1098$$

$$\sigma = \sqrt{\frac{\Sigma X^2}{n} - m^2}$$

$$= \sqrt{\frac{1098}{29} - 32.7}$$

$$= \sqrt{5.14}$$

$$\sigma = 2.26$$

第十一表. 用原來數目表示時,計算均方差之法

3. **用原來數目表示時計算均方差之法.** 第十一表所載者爲用原來數目表示時計算均方差之法,如是則校正數與假定原始點,均可不用。此法偶視之,似甚簡短,其實祇數目少時,用之始便利。且用此法計算時,應求小數點多位,始可較爲精確。

第十一表之第一行爲次第。第二行爲分數。其和爲166。第三行爲分數之平方。其和爲1098。其餘計算法,詳第十表。

大約第二種計算均方差之法,最爲節省工夫。

第三種方法所得結果與第二種方法相同。證明如下:

任何數目,X與平均數m之差數d,爲

$$d = X - m$$

方之

$$d^2 = X^2 - 2mX + m^2$$

其和爲

$$\Sigma d^2 = \Sigma X^2 - 2m\Sigma X + nm^2$$

以n除之

$$\frac{\Sigma d^2}{n} = \frac{\Sigma X^2}{n} - 2m\frac{\Sigma X}{n} + m^2$$

$$= \frac{\Sigma X^2}{n} - 2m^2 + m^2$$

$$= \frac{\Sigma X^2}{n} - m^2$$

故

$$\sigma=\sqrt{\frac{\Sigma d^2}{n}}=\sqrt{\frac{\Sigma X^2}{n}-m^2}$$

題問 1. 計算第二十五圖三曲線之均方差。

問題 2. 某分配有以下常數:m=76.4;σ=14.36。問橫坐標上−1.62σ處之量數X爲何數?

第十六章

百分等級

在統計學上,常常比較各個人之等級。例如一組共有十七人,則最高者可予以"十七"之等級,而最低者可予以"一"之等級。

若僅知某甲在某組之等級爲27,而不知該組之人數,則某甲等級之爲高,爲低,或居中,無從懸測。若在29人中,某甲之等級爲27,則其等級比較爲甚高。若在1,000人中,某甲之等級爲27,則其等級比較爲甚低。故以絕對等級表示一人在某組之位置,則該組之人數,亦必說明。所謂絕對等級者,卽將一組之人,由高而低,依次排列後,得最大之數目者,卽爲等級最高之人。如是,則絕對等級之表示,與尋常稱"第一"者不同。用統計語言,"一"之意爲最低數目,適與通常所用者相反。

若欲表示一人之等級,而同時不用某人所屬一組之人數,可用**百分等級**(Percentile Rank)表示某人之絕對等級。其意若曰有百人於此,則此人之等級爲何?設有五十人,則中間之一人,當爲25,但其百分等級應爲50。百分等級所表示者,卽某組中在此指定等級以下之百分人數。若某人之百分等級爲72,則此人超過百分之72人數,而爲百分之28人數所超過。

中數當爲百分之50。上二十五分值爲百分之75,而下二十五分值爲百分之25。

若數人而有同樣之分數,則其等級之規定,應有一種假設。試考以下十人之分數:

個人:	A	B	C	D	E	F	G	H	I	J
分數:	16	24	24	35	41	56	56	56	72	83

以上之分數,由小而大,順次排列。若以絕對等級表示之,則排列有如下者:

個人:	A	B	C	D	E	F	G	H	I	J
分數:	16	24	24	35	41	56	56	56	72	83
絕對等級:	1	2	3	4	5	6	7	8	9	10

但B與C之人數相同,故欲求公平,則其等級亦應相同。補救之法,予每人以相同之等級,卽折中等級是也。茲將絕對等級,修正如下:

個人:	A	B	C	D	E	F	G	H	I	J
分數:	16	24	24	35	41	56	56	56	72	83
絕對等級:	1	$2\frac{1}{2}$	$2\frac{1}{2}$	4	5	7	7	7	9	10

B與C之分數相同,各爲24,故其絕對等級,各爲$2\frac{1}{2}$;F, G,與H之分數相同,各爲56,故其絕對等級各爲7。等級總數爲十,無人得6或8之等級,因三人共佔6, 7, 8三等級,平均之,得7,而6與8可省去矣。

百分等級,亦應用此理。設有百人,由低而高,依次排列,

最高二十人之分數相同。若一人之百分爲81,而又一人之百分爲100,殊欠公允,因此二人之分數,原來相等。在此情形之下,此二人之百分等級,應各爲90,蓋此數爲最高二十人分數之中數。凡分配之中部或低部,苟有相同之分數者,亦用同一方法解決之。如是,則雖有數人均得同一之最高分數,而無人能得100之百分等級。

第十二表爲表示忒克塞斯(Texas)大學文科第一年生智力測驗分數之百分等級計算法。前二行表示組距,第一行爲最低限度,第二行爲最高限度。第三行爲組距中點。第四行爲次數。次數行之總和爲n,在本例爲860。

第二步爲表出三個百分行;下百分,上百分,中百分。因組距低限度之百分等級,決不若組距高限度之大。中百分表示全組之百分等級。此點論理,若不能採取,則百分等級之計算,將限於無窮之矛盾矣。

計算之第一步爲定率,此率爲總數之倒數。按例,總數n爲860。率爲$\frac{1}{n}$或$\frac{1}{860}$即.001162。此可由除法求得,或用巴羅氏乘除表(Barlow's Tables)求得,或用他法求得。

每一組距代表一百分距離。第五行爲每組百分距之低限度,第六行爲每組百分距之高限度。每組百分距之最高限度與較高一組百分距之最低限度相等。此節看表即知。每組個人應得之百分等級爲該組百分距之中點。詳第六行。

分數			次數	百分值		
從	至	中點		下	上	中
1	2	3	4	5	6	7
0	9	5	0	......	......	......
10	19	15	0	......	......	......
20	29	25	6	.000	.006	.003
30	39	35	19	.006	.029	.017
40	49	45	56	.029	.094	.061
50	59	55	94	.094	.203	.148
60	69	65	154	.203	.382	.292
70	79	75	161	.382	.569	.475
80	89	85	143	.569	.735	.652
90	99	95	95	.735	.845	.790
100	109	105	65	.845	.921	.883
110	119	115	33	.921	.959	.940
120	129	125	16	.959	.978	.968
130	139	135	8	.978	.987	.982
140	149	145	7	.987	.995	.991
150	159	155	3	.995	1.000	.997
160	169	165	0	......	......	......
	..................總數		860	......	......	......
率=1/n=1/860=0.001162						

第十二表　百分等級之計算法

百分率求得後,可用計算機計算。按**第十二表**之例,將6加6次(因該組之次數爲6);得數.006,爲20—29一組之最高百分等級。最低組距之最低百分等級爲零;最高組距之最高百分等級爲1.00。將百分率加19次,則得數爲.029。此不過遞加的加法耳。組距20-29之最高限度.006與組距30-39之最低限度.006相等。將百分率共加56次則得數爲.094,卽下一組距之最高百分值。

全行加完後,其得數爲1.00,等於百分率共加n次(.001162×860),**卽爲最高組距之最高限度。計算百分等級**

時,若用勃羅氏(Burroughs)加數機,甚為便利。

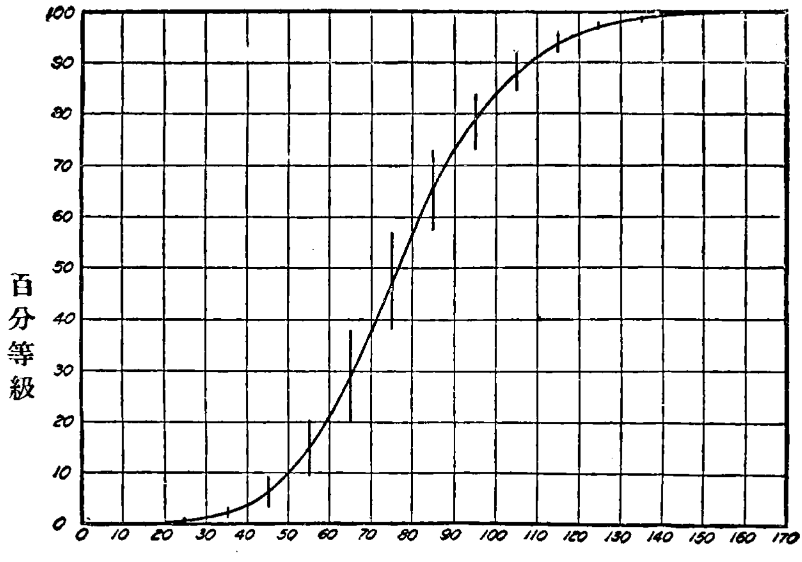
百分等級之計算法,即將百分率按照次數行遞加。在未遞加某組之次數以前,所得之和為該組之最低百分等級。在已遞加某組之次數之後,所得之和為該組之最高百分等級。

若得加數機之助,則**第二十二表**所示百分等級之計算,祇費三四分鐘,雖書記亦能為之。且此法可自動的校對其錯誤,因其最後之和,必為整數也。

百分曲線為表示任何變量(如分數)與其百分等級相關之曲線。根據此曲線,則某分數應得之百分等級為何,一視即知。所謂等級者,即該曲線所代表之分配之等級。**百分曲線實為一遞加的次數曲線,其次數之各縱線為總次數n之分數。**

第二十九圖之百分曲線,代表**第十二表**之次數及其百分值。底線代表測驗分數,縱線代表百分等級。例如,表上所載,80–89一組之百分等級為.65。按圖亦然,因80–89一組中點上之高處亦為.65。

80–89一組之最低百分等級與最高百分等級為.569與.735。此距離在圖上亦用短垂直線指出。故每組之百分等級,直可由圖查出。此垂直線非畫在組距之任一端,但畫在其中點之上。此短垂直線係根據**第十二表**之第5與第6二行而作。此曲線係連接此短垂直線之中點而成。故實



第二十九圖　百分曲線(與第二十表對照)

際上可以不必計算百分中點,因按圖觀察,即能畫得十分精確,而適於應用。苟慣爲此事,則表中之第三,第六,第七各行,皆可免去,而用圖決定。

就實際而言,若數目不多,則經過垂直線之曲線,不能平匀。故只好就其大概,而作此平匀曲線,稱之曰**已修匀百分曲線**(Smoothed Percentile Curve)。將百分多邊圖修匀爲百分曲線圖之理由,與將次數多邊圖修匀次數曲線圖之理由相同。但百分曲線圖修匀後,短垂線必須插入,以其能代表實際觀察也。且讀者一視曲線與垂直線中點相差之

距,可知曲線修勻之程序。百分多邊圖較相當之次數多邊圖爲繼續不斷。卽次數多邊圖極不規則時,若改爲百分多邊圖,則不然矣。

茲將百分曲線之特性總結如下:

1. 百分曲線對於底線或零點以及100等級爲**漸近的**(Asymptotic)。就理論而言,百分曲線終不致遠到零度與百度。故每人之等級,决不致低至零度高至百度。此事吾人常疏忽而不注意。

2. 中數可由百分曲線求得,因此數適爲50百分値處。按本例,50百分値處之分數爲77分,卽中數也。照此求法,可以節省不少工夫。

3. 上二十五分値,亦可用同法求得,因此數適在75百分値處。按本例,75百分値爲92分。下二十五分値爲25百分値,卽63。故百分値作好後,則各二十五分値,可以不必計算,而直由圖中求得。旣得各二十五分値,則二十五分距或二十五分差,亦不難求得。

4. 衆數之定義爲曲線最高點下底線上之一點。故衆數爲分配上次數最多之數。若在百分曲線上,則底線上之一點與曲線傾斜度最大之一點相對者,爲衆數。百分曲線,旣爲一遞加曲線,則量數中次數之最大者,當與傾斜度最大之處相對。此種用圖斷定之法,祗能得其大概,然爲實用起見,已覺得十分精確,且可以減少計算之工夫.

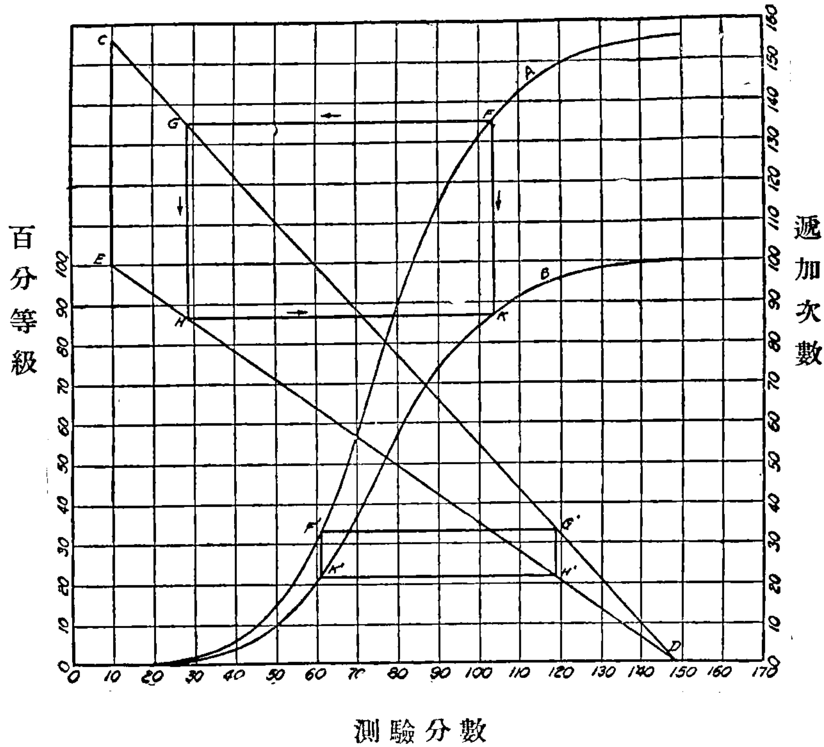
5. 若比較兩種分配,則察閱其百分曲線,可得許多事實。例如二分配全距離之比較,即爲其百分曲線全距離之比較。二分配離中趨勢之比較,即爲其百分曲線傾斜度之比較。百分曲線之傾斜度愈大,則其離中趨勢愈小。二分配之中數,亦可用圖比較。

作百分曲線時,**第十二表**所示之一切計算,未必完全需要。該表所以如此詳盡者,不過藉此以示計算之步驟耳。實際上,祗第一,二,四及第五各行,務必算出。但計算縮短時,須注意第五行之百分,並非劃歸於其相當之組距。乃爲其相當組距之最低百分。某組之最高百分,即第五行所載該組相當百分之高一個。此層祗須多加練習,即無困難發生矣。

百分等級之用圖計算法.[1] 若無數目之計算,亦可求百分等級。**第三十圖**即示此法。假定有一次數表,其分數依組距排列,並有相當之次數。設此分配之總數爲156。每組爲十單位自零以至170。遞加次數可用每組中點上之垂直線表之。垂直線之長,即爲每組次數之由一圖之右方量表而計算者。經過此垂直線,作平勻曲線A,此即爲遞加次數曲線。此種垂直線,在圖上未經表示,所以免除混亂。遞加次數曲線A可用以斷定本組無論何數之絕對等級。但吾人

[1] 此節在指定功課時,可由學生任意閱讀。

所需要者爲同樣曲線而可以斷定百分等級者。從圖之左方作任何量表代表百分等級自零度以至100。



第三十圖.　用圖計算百分等級法

作圖之法,指定任何點C,其高與曲線A之頂相並。指定任何點E,正在C之下面與圖左百分量表100處,從C與E兩點作二直線,與底線上任何點D相遇。

在曲線A上,假定任何點F。由F作平行線與上對角線相交於G點。由G作垂直線與下對角線相交於H點。從

H作一平行線與F點之垂直線相交於K。此K點為百分等級之一點,其價值與絕對等級之F點相當。同法從曲線A作同樣之點若干。在圖之下面,曲線A上之F′點,亦同樣指定。

經過此種點,作一平勻曲線,卽得百分等級曲線B。此法若用畫板,行之極易。作原來曲線A時,此線必經過各垂直線之中點,蓋此中點代表組距次數。

作此種曲線時,有一勘誤之法,卽曲線A之高度,必等於分配之總數156。

若有計算機時,則圖示法可以不用。最好仍用數目的次數,其法已詳**第十二表**。

表列教育及心理事實時,若用百分等級,則百分曲線之特性,務須明瞭。若慣用之則百分等級之臆測法,儘可棄之。蓋此種任意指定之百分等級,常有一種錯誤。此錯誤之來,由於百分之斷定以一組距之最高限度或最低限度為準。若用**第十二表**之算法,則此項錯誤,可以免矣。

問題1. 假定有二行數目,第一行為一班學生之考試分數。第二行為其百分等級。假定分數之分配為常態,則估計其均方差之法為何?

問題2. 用**第三,四,九及十四各表**,作一百分等級表。

問題3. 用**第三,四,九及十四各表**,作百分曲線。

問題4. 假定有一千學生之人名及其考試之百分等級。用手作一略圖,以示此百分等級曲線之形狀。又在同一圖之底線上,作一略圖,以示五百人之百分等級,此五百人係由一千中乘機而取出者。

問題5. 假定某班共有十個學生,且各人之分數均不相同。其百分等級應爲何? 此十人中,無一人能有0或100之等級,此不可不注意者。

問題6. 用手作二個百分曲線于一圖上,代表A與B二分配。此二分配之平均數及總次數相同,但A之離中趨勢較B爲大。二曲線須分別標明。

問題7. 用手作二個百分曲線于一圖上,代表A與B二分配。此二分配之平均數及均方差均相同。但A之總次數較B爲大。二曲線須分別標明。

問題8. 用手作二個百分曲線于一圖上,代表A與B二分配此二分配之均方差及總次數均相同。但A之平均數較B爲大。二曲線須分別標明。

問題9. 下表代表烏海烏省立大學(Ohio State University)第一級生750人在1913年之身長。試就各高度,計算其百分等級。試問5呎10吋之學生之百分等級爲何?試問此處百分等級之意義爲何?

高度(用吋表示者)	次數
61	2
62	10
63	11
64	38
65	57
66	93
67	106
68	126

69	109
70	87
71	75
72	23
73	9
74	4
	750

第十七章

二項展開式

人類智力與人事測量之次數分配,多少適合於所謂**機率曲線**(Probability Curve)者。欲用統計方法處理此種近似對稱形之機率曲線。應先研討此曲線之特性。

錯例　用各種次序排列許多事物,謂之**錯列**(Permutations)。試論a與b二事物。此二事物,可有二種排列法,ab與ba。此即謂之錯列。

再論三事物,a,b與c。若將三事物歸類,二種一次則可以排成六個錯列,即,

ab	ba	ca
ac	bc	cb

此處每種事物,與他一種接連組合一次。若將三事物歸類,三種一次,則亦有六種,即

abc	bac	cab
acb	bca	cba

再論四事物,a,b,c,d。若將四事物歸類,二種一次,則共有十二個錯列,即,

ab	ba	ca	da
ac	bc	cb	db
ad	bd	cd	dc

若將四事物歸類,三種一次,則共有二十四個錯列,卽,

abc	abd	acd	bcd
acb	adb	adc	bdc
bac	bad	cad	cbd
bca	bda	cda	cdb
cab	dab	dac	dbc
cba	dba	dca	dcb

同法,若將四事物歸類,四種一次,則共有二十四個錯列。n事物每次取r種,則其可能之錯列數目,可從下面公式預定之:

$$P_{nr}=n(n-1)(n-2)(n-3)\cdots\cdots(n-r+1)$$

以上各例,均可用此公式證實。故四事物每次取三次,其錯列之數目爲 $4(4-1)(4-2)=24$

組合. 組合(Combinations)與錯列不同。如abc組合,若將其次序調換,可得若干錯誤列,如abc, acb, bac, bca, cab, cba,以上六個錯列,祇能當作一個組合。故無論如何,錯列之數,常較組合之數爲多。

試論 a 與 b 二事物。若每次取二個,則組合之數目,祗爲一,卽 ab。若有 a, b, c 三事物,每次取二個,則共有三個可能之組合,卽 ab, ac, bc。若三事物,每次取三個,則祗有一個可能之組合,卽 abc。

再論四事物, a, b, c, d。若將此四事物歸類,每次取二個,則共有六個可能之組合,卽 ab, ac, ad, bc, bd, cd。若將此四事物歸類,每次取三個,則共有四個可能之組合,卽 abc, abd, acd, bcd。若將四事物歸類,每次四個,則祗有一組合。

n 事物每次取 r 種,則共可能之組合數目,可由以下公式預定之:

$$C_{nr}=\frac{n(n-1)(n-2)(n-3)\cdots\cdots\cdots\cdots(n-r+1)}{r!}$$

在此式內 r! 代表 r 之階乘積(Factorial r) 或 $(1\times2\times3\times4\times5\cdots\cdots\cdots r)$。故四事物每次取三個,則共組合之數目爲

$$C_{4\cdot3}=\frac{4(4-1)(4-2)}{1\cdot2\cdot3}=\frac{4\cdot3\cdot2}{1\cdot2\cdot3}=4.$$

再討論下列各例,則初學者對於組合與錯列之了解,益加透徹。假定購置二種物品。旣購之後,則全部購備爲二物品之一組合。但購辦之次序,或是不同。此購辦之次序,或爲 ab, 或爲 ba。此卽所謂錯列也。故二物品共有二錯列,而祗有一組合。

設購辦二種物品,而從三種物品中選出,共有多少組

合?組合之數目,爲每次購辦二種物品時購辦之數目。若三物品爲 a, b, c, 又若購辦祇限二種物品,則共有三種可能之購辦,卽, ab, ac, bc。用統計語言,三物每次取二個,共有三組合,每次購辦之物,其購辦之次序或不同。用統計語言,三物每次取二個,共有六個錯列。

一事之機率(Probability of Single Event). 設袋中藏白球五十,黑球五十。任意從袋中取一球,則得白球之機率爲 $\frac{1}{2}$,得黑球之機率,亦爲 $\frac{1}{2}$。二機率相加,等於整數。其關係如下:

$$P_B + P_W = 1$$

若袋中藏有16個白球,20個黑球,24個青球,28個紅球,計袋中共有88個球。現在若從袋中任意取一球,則得白球之機率爲 $\frac{16}{88}$,或 .182。得青球之機率爲 $\frac{24}{88}$,或 .273。其餘各色球之機率,可由以下方程式表出之:

$$P_W + P_B + P_G + P_R = 1$$

或　　$$.182 + .227 + .273 + .318 = 1$$

機率 (Probability) 之意義,爲取得任何一色之可能數目與總可能數目之比例。因袋中祇有四色,故此四色之機率之和必等於整數。

若吾人之目的,在從袋中取得白球而不願取得他色

之球,則取得白球之機率爲成功之機率,就本例而言,此項機率爲.182。失敗之機率,爲取得他色球之機率,此機率卽爲其他各機率之和,在本例爲.818。故就本例言,成功之機率,約爲一與五之比。此一與五之比,不致有十分變更,有之,則得與不得之機遇,大約各得其半。

成功之機率用P標記,失敗之機率,用Q標記,故

$$P+Q=1$$

複事之機率(Probability of Compound Event)　假設袋中藏白球五十黑球五十,若任意從袋中抽出二球,其錯列如下:

1. 白,白
2. 白,黑
3. 黑,白
4. 黑,黑

以上機率,代表三個組合,卽,

1. 二白球
2. 二黑球
3. 一白球,一黑球

複事之機率所涉及之問題,爲連抽二白球之機率,連抽二次所得之一白球與一黑球,連抽五次所得之四白球等。

連取二白球之機率爲何?第一次抽出一白球之機率爲.5第二次抽出一白球之機率亦爲.5,不過吾人必有一假設,卽第二次未抽前,第一次之白球,必須送回原袋。本問

題所討論者,爲決定連抽二白球之機率。在第一次抽出時,一白球衹由袋中一半之球取出。故第一抽中,有一半不能實現。而第二抽必爲一黑球,蓋袋中一半爲黑球也。故連抽二次均爲白球之機率,衹有四分之一。連抽二白球之機率,爲二抽之機率之積,卽 $.5\times.5=.25$。

若袋中藏白球20,黑球80,則抽一白球之機率爲 .2。連抽二白球之機率爲 $.2\times.2=.04$。連抽三白球之機率爲 $(2)^3=.008$;換言之,千抽之中,衹有八次成功。連抽三黑球之機率爲 $(.8)^3=.512$。

二項展開式(Binomial Expansion). 二項展開式各項對於複事之機率,有何意義,茲說明如下:

$$(P+Q)^n=P^n+\frac{n}{1}P^{n-1}Q+\frac{n(n-1)}{1\cdot 2}P^{n-2}Q^2$$

$$+\frac{n(n-1)(n-2)}{1\cdot 2\cdot 3}P^{n-3}Q^3$$

$$+\frac{n(n-1)(n-2)(n-3)}{1\cdot 2\cdot 3\cdot 4}P^{n-4}Q^4+\cdots\cdots\cdots$$

第一,二項寫出後,其餘各項,亦可推理寫出。上式之符號如下:

P=成功之機率

Q=失敗之機率

n=事件之數目

茲應用此方程式於複事之機率,例如投擲四錢,則此

各項之意義如下:第一項代表四錢完全正面之機率;第二項代表四個錢中有三個正面之機率;第三項代表四個錢中有二個正面之機率;第四項代表四個錢中有一個正面之機率;第五項代表四個完全反面之機率。此五個機率之和,等於確數或整數。

茲解釋此展開式之五項與投擲四錢之關係。此五項代表四個正面,三個正面,二個正面,一個正面,與四個反面之機率。其和爲整數。

第一項(The First Term)。若一錢投擲一次,則正面之機率爲$\frac{1}{2}$,反面之機率爲$\frac{1}{2}$。四錢連擲四次而完全正面之機率爲$\frac{1}{2}\times\frac{1}{2}\times\frac{1}{2}\times\frac{1}{2}=\left(\frac{1}{2}\right)^4=\frac{1}{16}$。此原則在討論複事之機率時,已經成立。因吾人既以正面爲成功,反面爲失敗,則此機率之標記爲P^n,其$P=\frac{1}{2}$而$n=4$,得$\frac{1}{16}$。故投擲四錢十六次中必有一次爲完全正面。此爲二項展開式第一項之意義。

第二項(The Second Term)先三次投擲時,完全正面之機率爲$P^3=\left(\frac{1}{2}\right)^3=\frac{1}{8}$,第四次投擲時,完全反面之機率爲$Q=\frac{1}{2}$。故先三次投擲時爲正面,第四次投擲時爲反面之機率爲$P^3Q=\frac{1}{8}\times\frac{1}{2}=\frac{1}{16}$。但吾人所欲知者,爲四擲中之

任何三擲爲正面之機率。意卽反面之錢或爲第一錢或爲第二,或爲第三,或爲第四錢。故"四擲中之三擲爲正面"較之"先三擲爲正面與第四擲爲反面"其次數應有四倍之多。四擲中之任何三擲爲正面之可能,有如下式:

H	H	H	T
H	H	T	H
H	T	H	H
T	H	H	H

吾人已知先三擲時爲正面與第四擲時爲反面之機率爲$\frac{1}{16}$。先二擲爲正面之機率爲$\left(\frac{1}{2}\right)^2=\frac{1}{4}$。故得先二擲爲正面與第三擲爲反面之機率爲$P^2Q=\frac{1}{8}$得最後一擲正面之機率爲$\frac{1}{2}$。故得先二擲爲正面,第三擲爲反面,與第四擲爲正面之機率,爲

$$P^2QP=\frac{1}{4}\times\frac{1}{2}\times\frac{1}{2}=\frac{1}{16}$$

同法,可示第一擲爲正面,第二擲爲反面最後二擲爲正面之機率爲$\frac{1}{16}$,又第一擲爲反面,最後三擲爲正面之機率爲$\frac{1}{16}$。此四機率之和,$\frac{1}{16}+\frac{1}{16}+\frac{1}{16}+\frac{1}{16}=\frac{1}{4}$爲四擲中任

何三擲爲正面之機率,用二項展開式表之,其式如下:

$$nP^{n-1}Q=4\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)=4\cdot\frac{1}{16}=\frac{4}{16}=\frac{1}{4}$$

第三項 (The Third Term) 先二擲爲正面之機率爲 $P^2=\left(\frac{1}{2}\right)^2=\frac{1}{4}$。先二擲爲正面而後二擲爲反面之機率爲此二機率之積,卽,$P^{n-2}Q^2$。此僅說明第三項一部分之所以然。但吾人所需要者,爲四擲中之任何二擲爲正面之機率。此機率較$\frac{1}{16}$爲大,因二個正面在以下任何一組合中,均可發現:1−2, 1−3, 1−4, 2−3, 2−4, 3−4。此數代表投擲。故四擲中之任何二擲爲正面之機率爲$\frac{6}{16}$。因數6由係數$\frac{n(n-1)}{1\times2}$,而定,此式之$n=4$。此因數爲組合之數。此組合爲四物排列每次取二個,故二項展開式之第三項代表四擲中任何二擲爲正面之機率,其 $n=4$, $P=\frac{1}{2}$,$Q=\frac{1}{2}$。

第四項 (The Fourth Term) 第一擲爲正面之機率爲 $P^{n-3}=\frac{1}{2}$。後三擲爲三面之機率爲 $Q^3=\left(\frac{1}{2}\right)^2=\frac{1}{8}$。故第一擲爲正面而後三擲爲反面之機率,爲

$$P^{n-3}Q^3=\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)^3=\left(\frac{1}{2}\right)^4=\frac{1}{16}。$$

但吾人所要者爲四擲中之任何一擲爲正面之機率。意卽爲三個正面可在四擲中之任何一組合內發現。現每次取

三個,四物組合之數爲

$$\frac{n(n-1)(n-2)}{1\cdot2\cdot3}=\frac{4\cdot3\cdot2}{1\cdot2\cdot3}=4.$$

故四擲中之任何一擲爲正面之發現,其次數之多,四倍於四擲中之第一擲爲正面之發現。四擲中之任何一擲爲正面之機率,可用二項展開式之第四項表之。此式爲

$$\frac{n(n-1)(n-2)}{1\cdot2\cdot3}\cdot P^{n-3}Q^3=4\cdot\left(\frac{1}{2}\right)\cdot\left(\frac{1}{2}\right)^3=\frac{4}{16}=\frac{1}{4}。$$

第五項(The Fifth Term)在四擲中四個反面之機率爲 $Q^4=\left(\frac{1}{2}\right)^4=\frac{1}{16}$。二項展開式第五項之係數爲

$$\frac{n(n-1)(n-2)(n-3)}{1\cdot2\cdot3\cdot4}=\frac{4\cdot3\cdot2\cdot1}{1\cdot2\cdot3\cdot4}=1$$

此係數代表四物每次取四個時之組合之數。此數適爲 1。此與常識相脗合,蓋四物每次取四個,祇有一組合也。故四擲四個反面之機率爲 $\frac{1}{16}$。五項相加,適等於一,此不可不注意者。其方程式之用數表者,可寫之如下:

$$(P+Q)^n=\frac{1}{16}+\frac{4}{16}+\frac{6}{16}+\frac{4}{16}+\frac{1}{16}=1$$

此爲四擲中　4　3　2　1　0　個正面之機率。

六擲. 下列六擲之結果,足以說明二項展開式在投擲較多時之應用。此展開式尙須二項,此二項可以照樣多

寫。此六擲之可能性,可用下表示之。

1. 六個正面,無反面
2. 五個正面,一個反面
3. 四個正面,二個反面
4. 三個正面,三個反面
5. 二個正面,四個反面
6. 一個正面,五個反面
7. 無正面,六個反面

從此可知若有六擲,必有七項始足表出正面與反面之組合。此二項展開式之七項如下(與前面五項之展開式比較):

$$(P+Q)^n = P^n + \frac{n}{1}P^{n-1}Q + \frac{n(n-1)}{1\cdot 2}P^{n-2}Q^2$$
$$+\frac{n(n-1)(n-2)}{1\cdot 2\cdot 3}P^{n-3}Q^3$$
$$+\frac{n(n-1)(n-2)(n-3)}{1\cdot 2\cdot 3\cdot 4}P^{n-4}Q^4$$
$$+\frac{n(n-1)(n-2)(n-3)(n-4)}{1\cdot 2\cdot 3\cdot 4\cdot 5}P^{n-5}Q^5$$
$$+\frac{n(n-1)(n-2)(n-3)(n-4)(n-5)}{1\cdot 2\cdot 3\cdot 4\cdot 5\cdot 6}P^{n-6}Q^6+\cdots$$

以上各項之數的估值,詳**第十三表**。各機率之和,等於整數,請注意。初學者若欲研究二項展開式,**第十三表**極為有用。

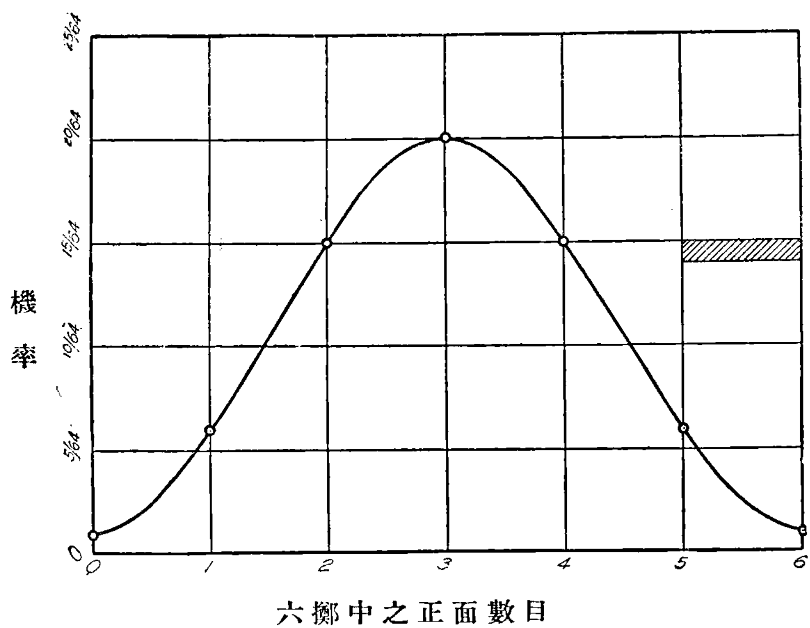
各項之次序	代數符號 係數	P	Q	係數值	P之因數	Q之因數	機率	能得機率之解釋
1		P^n		1	$\times\frac{1}{64}$		$=\frac{1}{64}=0.016$	六擲中六個正面
2	$\frac{n}{1}$	$\times P^{n-1}$	$\times Q$	6	$\times\frac{1}{32}$	$\times\frac{1}{2}$	$=\frac{6}{64}=0.094$	六擲中五個正面
3	$\frac{n(n-1)}{1\cdot2}$	$\times P^{n-2}$	$\times Q^2$	15	$\times\frac{1}{16}$	$\times\frac{1}{4}$	$=\frac{15}{64}=0.234$	六擲中四個正面
4	$\frac{n(n-1)(n-2)}{1\cdot2\cdot3}$	$\times P^{n-3}$	$\times Q^3$	20	$\times\frac{1}{8}$	$\times\frac{1}{8}$	$=\frac{20}{64}=0.312$	六擲中三個正面
5	$\frac{n(n-1)(n-2)(n-3)}{1\cdot2\cdot3\cdot4}$	$\times P^{n-4}$	$\times Q^4$	15	$\times\frac{1}{4}$	$\times\frac{1}{16}$	$=\frac{15}{64}=0.234$	六擲中二個正面
6	$\frac{n(n-1)(n-2)(n-3)(n-4)}{1\cdot2\cdot3\cdot4\cdot5}$	$\times P^{n-5}$	$\times Q^5$	6	$\times\frac{1}{2}$	$\times\frac{1}{32}$	$=\frac{6}{64}=0.094$	六擲中一個正面
7	$\frac{n(n-1)(n-2)(n-3)(n-4)(n-5)}{1\cdot2\cdot3\cdot4\cdot5\cdot6}$	$\times P^{n-6}$	$\times Q^6$	1	$\times 1$	$\times\frac{1}{64}$	$=\frac{1}{64}=0.016$	六擲中無正面

第十三表. 二項展開式之解釋

請注意:六擲皆爲正面之機率與六擲皆爲反面之機率相等。同樣,六擲中四個正面之機率與六擲中四個反面(二正面)之機率相等。此種關係,學者當能洞悉其代數符號及其理由。

第三十一圖係用圖式表此六事之機率。請注意其合乎常態圖之處,中央高起,兩端漸低。加斜線之小長方形代表 $\frac{1}{64}$,一視 x 與 y 軸可知也。若用測面器(Planimeter)求曲線下之面積,則其數爲一。

此爲理論的機率面。以前所作之曲線,係根據實得之



第三十一圖　六擲之機率

事實,此與理論之曲線,不無差異之處。現將曲線之爲實得事實所成者與曲線之爲機遇所成者作一比較。此種比較,在統計工作上;頗有價值。

參閱上表之六擲。請注意:六擲中得二個正面之機率爲$\frac{15}{64}$或.234。若六個錢共擲一千次,則其中必有234次爲二個正面。

問題1. 用一錢共擲八次,則得一個正面,二個正面,三個正面等之機率爲何,試用二項展開式法計算之。此法共用展開式九項。從計算結果,作一理論的機率曲線。若擲一個錢共648次,每次共八擲,則八次中有五個正面者,共幾次?

問題2. 若有人出洋十元與你賭洋一元,謂你十擲中不能得五次或五次以上之正面。試問賭百次後,你之勝負如何?(請看書中關於六擲之表)

問題3. 試說明以下說理之誤。"從一副紙牌抽出一張紅牌之機率爲$\frac{1}{2}$,倘'滑稽者'(Joker)一牌取出後,則抽出一張黑牌之機率爲$\frac{1}{2}$,故連抽十次時,則抽五張紅牌之機率爲$\frac{1}{2}$"

設每一牌抽出後,仍還原副牌中,試問十抽中,抽出五張紅牌之機率爲何?

問題4. 設一袋藏200個球,其分配如下:40紅球,50青球,50藍球,40黑球20白球。指定紅,青,藍球爲有顏色者,則六抽中中抽出一個,二個,三個等顏色球之機率爲何,試計算之。仿照書中六擲之計算式,將計算排列。請注意P與Q之價值,與書中所載者並不相等。將機率用圖示之。此曲線與平常之機率曲線,有何不同?在作曲線以前,如何用邏輯之

思考,而能知此圖與平常機率曲線之不同。

問題 5.　設有十牌,五紅五黑。將十牌亂排,抽出五張,每抽一張,不得放回。試計算抽出五張紅牌之機率。此機率必少于$\left(\frac{1}{2}\right)^5$。其理由安在?

第十八章

機率曲線

曲線之用二項展開式連續各項代表者,**謂之機率曲線**(Probability Curve),或**常態曲線**(Normal Curve)。此曲線可用方程式述之如下:

$$y=y_0e^{\frac{-x^2}{2\sigma^2}}, \qquad (1)$$

在此式內,

x =量數與平均數之差數之用均方差(或標準差)表示者,

y =縱線,並代表所希冀之次數,

σ =分配之均方差之為組距者,

e =常數2.718,以**納比侖**(Napierian)對數底數名,

y_0=平均縱數。

根據此方程式,則知任何一差數後,即可決定其相當縱線。平均數上之縱線,可用下式表出之

$$Y_0=\frac{n}{\sigma\sqrt{2\pi}}=\frac{n}{2\cdot5066\sigma}, \qquad (2)$$

在此式內,

n =分配之總次數,

σ =分配之均方差之為組距者,

π =常數3.1416。

根據此方程式,可以決定平均數上之次數,此次數可用實得之次數及均方差之爲組距者表出之。

故機率曲線較詳之方程式爲

$$y=\frac{n}{\sigma\sqrt{2\pi}}e^{\frac{-x^2}{2\sigma^2}} \qquad (3)$$

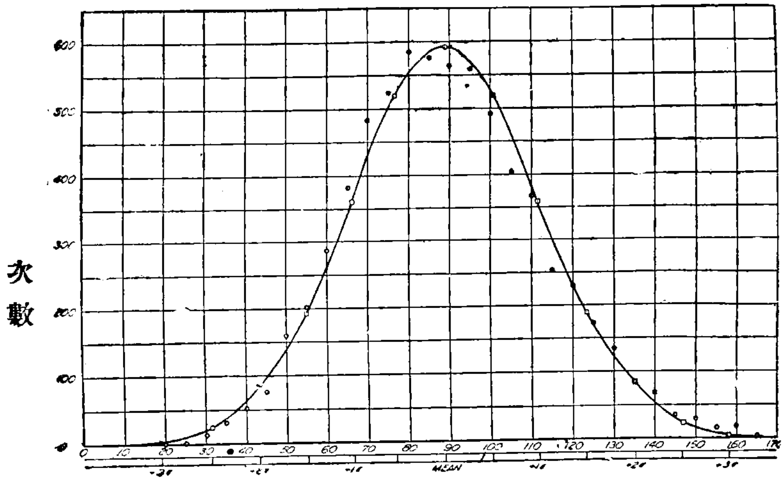
根據此式,可將理論曲線重疊於實得曲線之上,而作比較。

欲達此點,必用方程式(2),而求平均數上之次數。均方差與次數總數,可用常法求之。用方程式(2)求得平均數上之次數後,可從附錄之**第十九表**,而求得曲線他部之縱線。例如,從此表,可知在1σ處之縱線爲.606。倘平均數上之縱線爲100,則在1σ處之縱線爲60.6。此曲線爲對稱的σ之正值與其負值相等。故在$-.74\sigma$處機率曲線之縱線,爲平均縱線之.760。倘平均縱線爲100,則在$-.74\sigma$處之縱線爲76。

設繪一次數分配圖,如**第三十二圖**。將機率曲線重疊其上,藉覘實際的分配與理論的曲線,相去多少。圖上之小圈代表實際次數而與**第十四表**所載之次數相當。重疊之手續如下:

1. 計算均方差如**第十四表**。因σ照d行計算,σ係用組距表示。故$\sigma=4.61$組距或23.05分數單位,因每組共五單位。

2. 計算眞正算術平均數如**第十四表**。此數爲88.9單位。σ量表之原始點,卽在此點,如**第三十二圖**。



第三十二圖　常態曲線與次數多邊圖重疊

3. 計算機率曲線之平均縱線，其均方差為 4.61，而次數之總數為6806。

$$y_0=\frac{n}{2.5066\sigma}=\frac{6806}{2.5066\times4.61}=590$$

平均數上之次數為 590，而用組距表示者。

4. 求 $+.5\sigma$，$+1\sigma$，$+1.5\sigma$，各值在x－量表上之位置。其算法如下。眞正算術平均數為88.90。因 $\sigma=23.05$，故 $\frac{1}{2}\sigma=11.52$。

算術平均數＝88.90　　　　算術平均數＝88.90

$$+.5\sigma=\frac{+11.52}{100.42}\qquad\qquad\frac{-11.52}{77.38}=-.5\sigma$$

$$+1\sigma=\frac{+11.52}{111.94}\qquad\qquad\frac{-11.52}{65.86}=-1\sigma$$

$$+1.5\,\sigma=\frac{+11.52}{123.46}\qquad\frac{-11.52}{54.34}=-1.5\,\sigma$$

$$+\ 2\,\sigma=\frac{+11.52}{134.98}\qquad\frac{-11.52}{42.82}=-2\,\sigma$$

$$+2.5\,\sigma=\frac{+11.52}{146.50}\qquad\frac{-11.52}{31.30}=-2.5\,\sigma$$

$$+\ 3\,\sigma=\frac{+11.52}{158.02}\qquad\frac{-11.52}{19.78}=-3\,\sigma$$

X	f	d	fd	fd²
20	2	13	26	338
25	2	12	24	288
30	14	11	154	1,694
35	30	10	300	3,000
40	53	9	477	4,293
45	76	8	608	4,864
50	162	7	1,134	7,938
55	202	6	1,212	7,272
60	287	5	1,435	7,175
65	378	4	1,512	6,048
70	483	3	1,449	4,347
75	524	2	1,048	2,096
80	580	1	580	580
85	575	0	9,959	
90	564	1	564	564
95	561	2	1,122	2,244
100	492	3	1,476	4,428
105	405	4	1,620	6,480
110	368	5	1,840	9,200
115	256	6	1,536	9,216
120	233	7	1,631	11,417
125	173	8	1,384	11,072
130	137	9	1,233	11,097
135	82	10	820	8,200
140	69	11	759	8,349
145	35	12	420	5,040
150	31	13	403	5,239
155	14	14	196	2,744
160	16	15	240	3,600
165	2	16	32	512
	6806		15,276	

$n=6,806$

$\Sigma fd+=15,276$

$\Sigma fd-=9,959$

$\Sigma fd=5,317$

$\Sigma fd^2=149,335$

$c=\frac{\Sigma fd}{n}=.781=$ 之校正數

$c^2=.61$

σ 之計算

$\sigma=\sqrt{\frac{\Sigma fd^2}{n}-c^2}=$ 組距

$=\sqrt{\frac{149335}{6806}-.61}$

$c=4.61$

平均數之計算

$c=\frac{I.\ \Sigma fd}{n}=\frac{5\times5317}{6806}=+3.9$

$c=m$ 之校正數

$m=m_a+c=85+3.9=88.9$

第十四表　次數表上之平均數與均方差之計算法.

此分配代表6806個工程學生之智力測驗分數。對照

第三十二圖而證實各點之地位。

5. 計算各點上之縱線,其法如下:

σ	$\frac{y}{y_0}$	y
0	1.000	590
$+.5\sigma$	.882	520
$+1\sigma$	.606	358
$+1.5\sigma$	.324	191
$+2\sigma$	.135	80
$+2.5\sigma$	.044	26
$+3\sigma$	.011	6

$\frac{y}{y_0}$縱行下各分數,可從**第十九表**得到。y縱行下各次數為平均縱線590之分數,從**第三十二圖**可以看出各次數y之正負兩面相同,蓋機率曲線為對稱的。

試取**第三十二圖**機率曲線上之縱線,而作比較。一視而知6806測驗分數之實際分配與機率曲線相差之程度。測驗分數略略正的偏態,與常態面稍有不同。

根據**第十九表**,吾人若已知某差數之數,卽可求得與此數相當之縱線,又吾人若已知平均數上縱線之長,則與其他縱線相當之差數,亦可求得。例如某縱線之長,適為平均縱線之長之一半,則與某縱線相當之差數為多少。法卽參閱**第十九表**上之.50處,可知凡縱線之長,適為平均縱

線之一半,其相當之差數爲 1.18σ 與 -1.18σ。

問題1. 按照以下各條,作三個機率曲線于一張紙上。在每組距中點,上指定其縱線。

曲線	σ	平均數	n	組距
A	15	50	400	10
B	15	50	800	10
C	15	50	1200	10

問題2. 按照以下各條,作三個機率曲線于一張紙上,在每半個式格碼(σ)處,用小圈指出縱線。

曲線	σ	平均數	n
A	10	60	400
B	20	60	800
C	30	60	1200

問題3. 按照以下各條,作三個機率曲線于一張紙上,在每半個式格碼(σ)處,用小圈指出縱線。

曲線	σ	平均數	n
A	10	60	1000
B	20	60	1000
C	30	10	1000

問題4. 有一分配,其常數如下: m=56.3; n=4320; σ=11.62; 假定此分配爲常態。1. 試問平均縱線爲何? 2. 在何X-量數處,其縱線爲平均縱線之三分之一? 3. X-量數爲41.5時,其縱線之值爲何?

問題5. 將第十六章問題9之事實(大學一年級生之身長)畫于一方格紙上,並用小圈指出此事實之縱線。重疊一常態曲線于其上,而用圖比較此實際分配與機率曲線。在底線上,指出平均數.5σ, +1σ, +1.5σ, +2σ, +2.5σ,之正負各地位,與第三十二圖同。比較實際分配與機率曲線。

第十九章

次數面之面積

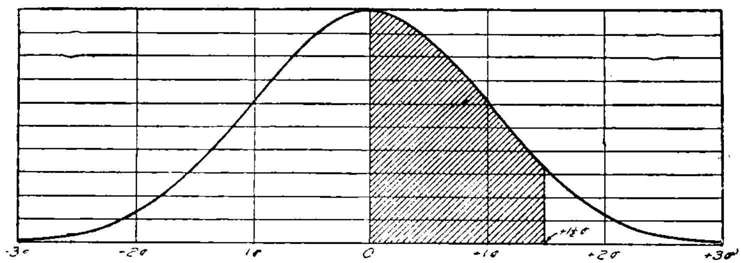
次數面之面積之大小與其所代表之次數之大小成正比例。若機率曲線下之面積爲一,則 X－量表任何二點間之數與分配之總數,可用比例寫出。**第二十表**所載者,卽爲此種比例。

假設欲求平均數與 $+1.5\sigma$ 二點間之面積爲全面積之幾分之幾(參閱**第三十三圖**之第一式)。參閱**第十二表**,得分數 .4332,意卽在平均數與 $+1.5\sigma$ 之間之面積爲全面積之 43%。若總數爲 500,則平均數與 $+1.5\sigma$ 間之數爲 $500\times.43$ 或 215。同類之問題,可用同樣之方法求之。

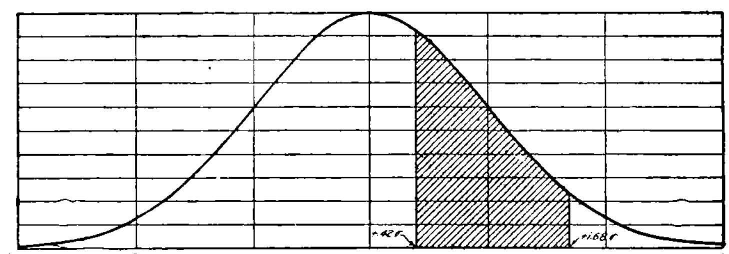
設有一常態分配,其總數爲 460,試問 $+42\sigma$ 與 1.68σ 間之數爲多少(參閱**第三十三圖**第二式)。參閱**第二十表**得 16.28%。同法,可求全數之 45.35% 乃在平均數與 $+1.68\sigma$ 之間。求 45.35 與 －16.28 之差,得 29.07,此 29.07% 卽爲 $+.42\sigma$ 與 $+1.68\sigma$ 間所佔全面積之百分數。因 n 爲 460,故本問題之答數爲 $460\times29.07=134$,此卽 $+.42\sigma$ 與 1.68σ 間所含之數也。

設有一常態分配,其數爲 280,試問平均數與 $-.65\sigma$ 間之數爲多少(參閱**第三十三圖**第三式)。參閱**第二十表**,得 24.22%。因圖爲常態,故平均數與 $-.65\sigma$ 間之數,與平均數與 $-.65\sigma$ 之數相等因 n 爲 280,故本問題之答數爲 $280\times$

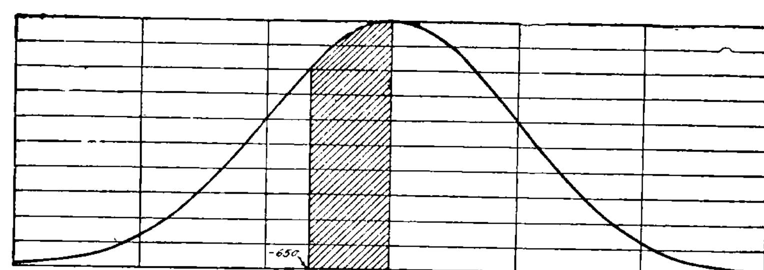
第三十三圖之第一式



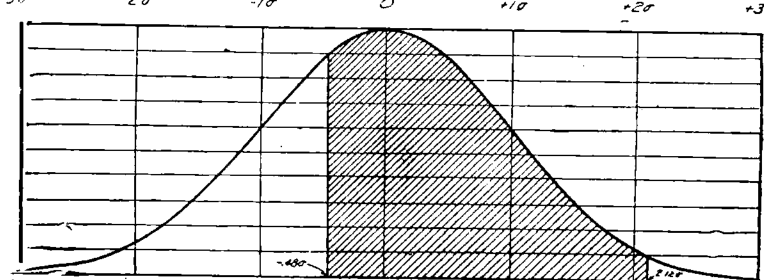
第三十三圖之第二式



第三十三圖之第三式



第三十三圖之第四式



第三十三圖　次數圖之面積

24.22＝67.8。

再求 $-.48\sigma$ 與 $+2.12\sigma$ 間之數（參閱**第三十三圖**第四式）。此圖可分爲二部，界點在平均數上。參閱**第二十表**，知平均數下面（或負面）之黑斜線部全面積之18.44%，而平均數上面（或正面）之黑斜線部爲全面積之48.30%將二個百分數相加，卽18.44＋48.30＝66.74%卽 $-.48\sigma$ 與 2.12σ 間所佔全面積之百分數也。

問題 1.　作一常態曲線于底線上，自 $-\sigma$ 至 $+3\sigma$。將底線分作五段。求每段內之百分數。

問題 2.　作一常態曲線，將底線分作五段，使每段內之百分數相等。

問題 3.　在一常態面上，求 $+3\sigma$ 與 -3σ 間之百分數。

問題 4.　有一常態面，其常數如下：$m=76.34$；$\sigma=2.43$（組距）；$I=10$；$n=785$。試問X－値62與94間之數爲多少。（I爲X－量表之組距單位數）62與94，必用均方差表出。

第二十章

量數之蛻變

等級與地位之區別,在統計學上,極為重要。若研究每組中各人身材之分配,可用二法表出之;一為絕對或百分等級,一為實在尺寸之用差數(如均方差)表示者。前者謂之**等級**(Ranks),後者謂之**地位**(Standing)。

等級與地位之分,可用例說明之。設有百人於此,由短而長,依次排列,並指定等級,則任何接近二人,其相差之數,若用等級表示,莫不相等。但用身材表示,則任何二人之相差,決不能相等。蓋兩端二人之相差,比較的常大,而中間二人之相差,比較的常小。苟用實在尺寸表示中間百分之40至60之相差,決無極端80至 100 相差之大。此由於中間之人數,常較兩端之人數為多也。

吾人若將一種量數,變為他種,則此手續謂之**量數之蛻變**(Transmutation of Measures)。若事實用百分等級式者,則此事實之分配,或為常態,或為偏態,不得而知。苟將一列等級,用地位表示,則分配宜假定其為常態者。

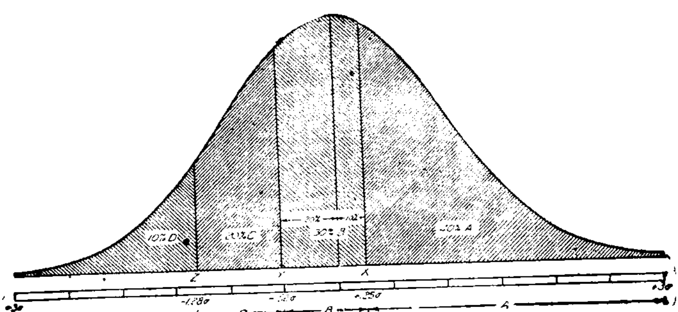
設二評判員,各評同一組人,而將每人所得之二種評判併合。苟二評判員用同一量表,則其評判可以併合。但有時一個評判員常較其他評判員為寬,或其判別之法,較為細密。設二評判員均用A, B, C, D,為評判之符號,以A為最

高,以D爲最低。若一評判員所給之A,較其他一評判員所給之A多一倍,則二評判員之A之價值,不得謂之相等。現有一人,由第一評判員處得C,而由第二較寬之評判員得B,則統計上之問題,爲如何能根據此二種評判,而與此人以一個單獨之分數。

假設能力之分配,根據常態機率曲線。又假設每組中共有百人。設二評判員所給之等第如下:

	第一評判員	第二評判員
等第A	40	20
等第B	30	30
等第C	20	30
等第D	10	20
	n=100	n=100

一視而知第一評判員之給分數,較第二評判員爲寬。



第三十四圖　量數之蛻變

量數之蛻變,手續極簡,如**第三十四圖**,代表第一評判員所給等第之分配。此圖應有常態次數面之略形,並有垂直線以指示各等第之各組。共有四組。等第A佔全數之百分之四十,故在圖上,幾佔全次數面之一半。其次百分之三

等第	全組之百分數	百分限度	平均數	相當式格碼地位
A	40	+50 +10	+30%	+ .84 σ
B	30	+10 −20	− 5%	− .13 σ
C	20	−20 −40	−30%	− .84 σ
D	10 n =100	−40 −50	−45%	−1.65 σ

第十五表　量數之蛻變

十為等第B,此段較小,而佔平均數上之位置。其他各段,照樣表出。從此圖可知A組之中間一人與全組之中數相去百分之三十。此與全組之中數相去百分之三十之點,其式克碼之值為.84 σ,視**第二十表**卽知。**第十五表**是為第一評判員而作者。但每一評判員均得一表。二評判員之字母等第,若用式克碼表示,有如下式。

字母等第之用式克碼表示者

	第一評判員	第二評判員
等第A	.84 σ	1.28 σ
等第B	−.13 σ	.39 σ
等第C	−.84 σ	−.39 σ
等第D	−1.65 σ	−1.28 σ

現在若欲合併二評判員之估計,可將字母等第改爲數字而平均之。若一組之某人,第一評判員予以等第A而第二評判員予以等第C,則其平均地位爲

$$\frac{+.84-.39}{2}=+.22$$

此爲合併二種不同評判之最良方法。但有二假設:一爲能力之分配是常態的;一爲吾人對此二評判員具同一之信仰。此二假設或可實現或可不實現;但無論如何,吾人得此二假設後,可將二種不同之評判,合而爲一。

此方法不特可以將二種不同之評判併合,并能將評判員所定之標準,用客觀式表出。此方法可適用於以下之事實:教師之學業的分數,等級量表上評判;能力估計用之以作斷定智力測驗之預占値之標準者。無論何人爲評判員,常有估計過高之趨勢。此種估計過高,本無大礙,惟在統計法上,殊感不便。一組之人,不能人人均在平均之上。近有報紙,載軍隊不識字之人頗多,大書而特書曰某團兵士之

程度,有一半在該團平均之下!

問題1. 三評判員估計之分配如下。求每評判員之每等第的式格碼(σ)地位,藉為合併估計之用。用表表示此種相當地位。設有一人,第一評判員予以B,第二評判員予以A,第三評判員予以D,試問其式克碼之地位為何? 用此種蛻變,試問最大之式克碼地位為何? 最小之式克碼地位為何?

	第一評判員	第二評判員	第三評判員
等第A	20	12	35
等第B	22	22	22
等第C	19	32	20
等第D	24	27	17
等第E	21	13	12
	106	106	106

問題2. 作一表用二直行表出分配常態時之百分等級與式克碼地位之關係。百分等級行之各數,假定為 0, 5, 10, 15, 20, 等,而同時從附錄表中求其相當式克碼之地位。

作一同樣表及同樣縱標 (Leading)。在式克碼行下,填入 -3σ, -2.5σ, -2.0σ, -1.5σ, 等,而從附錄中,求其相當之百分等級。

圖示等級(ranks)與地位(standing)之區別。

第二十一章

機誤

統計量數之可靠性. 機誤之計算,極其通行,幾乎人人用之。教育上之統計事業,視爲科學的,尊嚴的,與可信的,亦以其計算機誤(Probable Error)之故。然則何者爲機誤,何者非機誤何者爲機誤所能表示者,何者爲非機誤所能表示者;何種可靠性可由機誤而得,何種決非由機誤而得。若能了解機誤之根本假設,則可善用機誤,而不致貽機誤萬能之誚矣。

設欲知芝加哥城小學六年級生之平均年齡。其理想與完美之解答方法,卽將該城小學六年級之學生,一一列舉。然欲搜羅全體事實,毫無缺漏,實非易易。故勢不得不從事於較少而同時較爲代表的事實。若以本例而言,或可取小學六年級生千人,而求其平均年齡。所得之數,或可代表全體。此種局部事實而非全體事實之研究,統計學上,謂之**取樣**(Sampling)。

倘所得之樣爲一千個六年級兒童,則其平均年齡,應視爲可靠,蓋一千非小數也。若爲一百兒童,則其平均年齡,自不若一千之可靠。若更縮小而爲十個兒童,則其平均年齡,比較的愈不可靠,而愈不足以代表該城六年級兒童之年齡矣。更縮小而爲三四個兒童,則其平均年齡,可謂完全

不能代表該城六年級學生之年齡,吾人對之,亦不能有信用。然則數目愈大,則其所得之平均愈可靠,而吾人信用亦愈增,明矣。精密言之,平均數之可靠性,乃取樣數目之函數(Function)也。

若已求得一百兒童之年齡,並求得其平均,吾人苟欲用此平均而證明某事,自不能不知此平均數可靠之程度。用統計術語說,現在問題是:若吾人重取一百兒童,又取一百兒童等,而求其平均年齡,則第一次一百兒童之平均年齡,與第二次,第三次者相差如何。數次一百兒童所得之平均年齡,必不能絕對相等,可斷言也。雖曰他種分子純粹,而機遇(chance)一事,卽足以發生微小之變動。

此種各組每百兒童平均年齡上之微小變動,比較爲極小。倘每組之人數愈小,則各組兒童年齡之變動愈大。最後,若每組之兒童,縮小而爲一個,則各組年齡之相差,卽六年級各兒童年齡之相差。應用算術平均數之機誤之目的,卽表示若用不同各組而求其平均數時,其變動之量爲何。

不可靠性之各種理由　在教育與社會科學之統計上,有幾種錯誤,不可不知。一種算術平均數上之錯誤,是由機遇而來。若數目大時,此種錯誤,不甚重要。設一百兒童之平均年齡爲11.86歲,則其他一百兒童之平均年齡,相差必不甚多,或其所差之數,祇小數點之第一位或第二位耳。

尙有一種錯誤,在統計學上,較爲重要,卽錯誤之由於

樣本之偏重而不足以代表吾人所欲測量之事物者。如在一特別學校或城之一區而求小學六年級生之年齡,則城之他區之學生,代表社會之不同階級者,定要表示年齡;身材,身重,智力,及其他之差異。若此種所取之樣,不足十分代表統計上所需之材料,則其所得之錯誤,較由機遇所得者爲尤甚。宜注意者,卽機誤所斷定者,爲變動之全由於機遇者,若夫材料之選擇,苟有偏重,而不足以代表統計上之需要者,則機誤無從斷定其變動之程度。但此種因材料偏重而生之變動,關係甚大,不可不注意。例如六年級學生,依字母次序分班,研究者隨機的從表上取出最前一百名。偶視之,此種手續,可謂隨機選擇,蓋因字母而排列之姓名,其先後固與智力之優劣無關。但某民族(譯者按美國由許多民族組合而成)之人,其姓名或多歸於字母之某部,則此種姓名之選擇法,雖曰隨機,而或有偏重某民族之弊,如是,則其智力,決不能代表全體而不免發生變動矣。此種變動,及變動之類此者,決非機誤所能斷定。故研究機誤及他種斷定可靠性之量數時,須知機誤等所能斷定之差異,祇限於選擇法相同之材料,且此種差異,較之其他各種分子所生之差異,實不甚重要。

機誤之一個試驗. 設有一千張卡片,每張各寫一個數字。數字之排列,自零至24。又設其算術平均數爲12,均方差爲4。將卡片作爲一堆,而任意取出二十張。其平均數必

近於12,蓋12爲全堆之平均數,但亦決不能適爲12,因取出之數祗二十耳。循機遇之理,此平均數必有變動,或在11.5與13.2之間,或與12相鄰近。

機誤之目的在指示以後所取每二十張之平均數與第一次所取二十張之平均數相差之程度爲何。平均數應有之機遇的差異,可用機誤表出之,其式如下:

$$\text{算術平均數之機誤} = \frac{.67\sigma}{\sqrt{n}},$$

在此式內,σ爲任何分配之均方差,n爲次數之總和,卽二十。

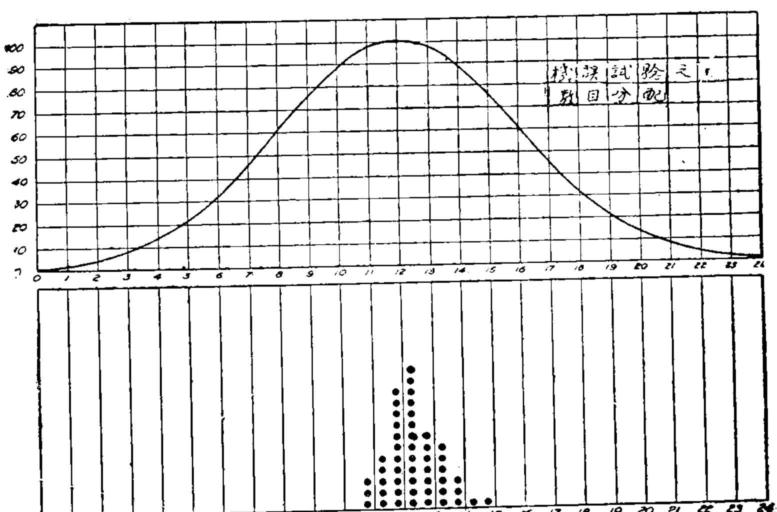
上述之例,按實際結果第一次二十個數目之平均數爲12.4,均方差爲3.21,其所得之機誤爲.48。本例之平均數若已得機誤,可用下法寫出:

平均數 $=12.40\pm.48$。 其各種解釋,略作討論如下。

以上公式之宜注意者,卽n爲該式之分母。此與以前之敍述相脗合,卽數目愈大,則平均數之機遇的變動愈小。倘此堆卡片之數字排列,自零至一百萬,則每二十張卡片之平均數之變動,自當較數字排列之自零至十者爲大。此與公式亦相脗合,卽均方差適爲分子。倘本例數目之差數變大,則平均數之機遇的變動亦大。

平均數之機誤或機遇的變動,與均方差成正比例與次數之數目成反比例。

第三十五圖爲一千數目之次數分配,其平均數爲12,均方差爲4。全部距離自零至24。由一千內,取出若干二十個數目,然後計算其各種平均數。此若干平均數用小點在圖之下面畫出。從一千數目內,取出五十樣本,每種樣本,各二十個數目。每次二十張取出後,仍放回原堆,故每取二十張時,均由一千張內取出。此五十小點表示各樣本之平均數之分布,不若單獨數目之分布。各平均數全距離在10與15之間,而各個原來數目之分布,在0與24之間。換言之,若有數張卡片,其數目屬於低限度者,如0, 1, 2,復有數張卡片,其數目屬於高限度者,如22, 23, 24,則此樣本之二十數



平均數之分配

第三十五圖　機誤之試驗

目,其平均數決不近於兩端之數。平均數之差異,不若單獨數目之差異之大,且所抽每組之數目愈大,則其平均數之差異愈小。

第十六表爲一實地試驗之記錄,有一千數目,共抽出五十樣本,每樣共二十數。第一行爲樣本之抽出次序。第二行爲樣本之平均數。各平均差異之程度,閱第二行,卽可瞭然。

二十個數之樣本	樣本之平均數	樣本之均方差	樣本之機誤	眞正平均數12是否在P.E.限度內?	二十個數之樣本	樣本之均平數	樣本之均方差	樣本之機誤	眞正平均數12是否在P.E.限度內?
1	12.40	3.21	.48	是	26	12.60	4.49	.66	是
2	13.15	4.26	.64	否	27	10.50	3.49	.53	否
3	14.0[illegible]	3.97	.60	否	28	12.05	3.49	.51	是
4	11.95	3.00	.45	是	29	11.55	3.81	.58	是
5	12.40	3.62	.55	是	30	12.25	4.06	.61	是
6	12.65	3.16	.48	否	31	12.25	3.74	.56	是
7	12.35	3.20	.48	是	32	11.05	2.91	.44	否
8	13.40	2.74	.41	否	33	12.55	3.41	.51	否
9	11.40	3.29	.50	否	34	11.35	3.80	.57	否
10	12.50	2.92	.44	否	35	11.65	3.23	.49	是
11	11.65	3.65	.55	是	36	12.40	3.06	.46	是
12	11.60	3.71	.56	是	37	13.60	3.04	.46	否
13	11.05	4.93	.74	否	38	11.45	4.48	.68	是
14	11.70	4.16	.63	是	39	12.05	5.37	.81	是
15	13.65	4.20	.63	否	40	11.50	3.19	.48	否
16	13.95	3.73	.56	否	41	12.80	6.44	.97	是
17	13.15	4.48	.68	否	42	10.55	3.38	.51	否
18	11.85	2.85	.43	是	43	13.00	3.66	.55	否
19	11.60	4.14	.63	是	44	12.55	4.21	.63	是
20	13.05	3.71	.56	否	45	12.15	4.04	.61	是
21	13.35	4.28	.65	否	46	12.15	4.76	.72	是
22	12.95	3.77	.57	否	47	12 05	4.15	.62	是
23	11.60	2.44	.37	否	48	11.95	3.67	.55	是
24	14.75	4.32	.65	否	49	12.45	4.13	.62	是
25	10.70	3.86	.58	否	50	12.40	3.90	.59	是

第十六表. 機誤之試驗的研究

第三行爲每二十個數目之樣本之均方差,其數大約在 4 左右,此亦爲大數之均方差。於此可從經驗證明,若從大數之中抽出一樣本,則此樣本之均方差,與大數之均方差相差不遠。苟吾人一視**第三十五圖**之縱線,卽知此層之不足爲奇。在 12 處之縱線爲 100,其意卽大數中共有 100 個十二,足供各樣本之抽取。在 8 處之縱線爲 61,其意卽在大數中共 61 個八。然則從大數中抽一卡片,則抽十二之機遇爲 $\frac{100}{1000}$ 或 $\frac{1}{10}$,蓋大數中共有一千卡片也。抽一個八之機遇爲 $\frac{61}{1000}$。總之,各樣本之平均數之分配,其形式與大數之分配曲線相似,祇以樣本之數目較小,故其分配面亦較低耳。樣本平均數之面積,祇見全數面積之 $\frac{20}{1000}$。從樣本抽出十二或八之機率,與從大數中抽十二或八之機率相同。故曰,樣本之差數,與大數之差數相同或相似。樣本之縱線常較短,以其數目較小也。

第四行爲各樣本之平均數之機誤。在第一樣本,其平均數爲 12.40,均方差爲 3.21,平均數之機誤爲 .48。故平均數之寫法爲 12.40±.48, 此足示機誤之全距離爲 11.92 至 12.88。

機誤之解釋. 機誤之意義如何?茲解釋之。上述之平均數及機誤,有解釋爲眞正平均數必在 11.92 與 12.88 之間者,誤矣。此大堆卡片中之眞正平均數爲 12,此固由吾人先行布置,而使之然者。就本樣本言,眞正平均數確在機誤

限度11.92與12.88之間。故**第十六表**第五行下作「是」字樣。但第二樣本之平均數爲13.15機誤爲.64。其機誤之限度爲12.51至13.79。由此觀之,則本樣本之平均數13.15。及本樣本平均數之機誤限度,均較一大數之眞正平均數爲大。

現在所欲研究之問題是"眞正平均數究在機誤限度之內否"?答曰,26次"是"24次"否"。然則大數之眞正平均數,或在樣本之機誤限度內,或在其外,蓋亦明矣。實則眞正平均數在任何樣本之機誤限度內者,其機遇相等。但吾人不可注意者,卽眞正平均數在任何樣本之機誤限度外者,其機遇亦相等。以上所言,祗限於平均數之變動之由純粹機遇所致者,苟樣本作無意之偏重,不足以代表全體數目,則機誤失其效力矣。

設測量十歲兒童之身材,而得二十個十歲兒童之記錄。計算此組兒童之平均身材,並求其機誤。倘機誤之數目小,則此十歲兒童之平均身材,可視爲正確。但吾人須知十歲兒童眞正高度之在機誤限度之外者,其機遇亦相等。

實際情形與上例相同,其不同者,爲實際施行時吾人不知**大數**(Universe)之確數:一城或一區所有之十歲兒童。在上舉卡片之例,此大數卽爲一大堆卡片。實行測驗時,吾人無法探知任何事物之全部數量。所知者僅此大數之樣本耳,此樣本之數,或爲二十,或爲一百,或爲數千,依照情形,而大小其數。吾人常從事於樣本之抽取,如水菓,水門泥,或

穀粒之估價,亦取其樣本也。此樣本若由滿車或滿袋隨機或任意取出,則此樣本多少是爲代表數目。此樣本與彼樣本,因機遇之差異,不能相同;但此種差異之程度,可由機誤公式斷定,不過吾人之估價,必爲數量而後可。但樣本之抽取,不免有有意與無意之偏重,如車中之物暴露於熱氣,日光,與潮濕等,卽不足以代表全車之物。此爲無意之偏重。反之每籃水菓,上面者常較好,不足以代表全籃,商人常以之引購客,而購客亦知其用意之所在,此爲有意之偏重。此種有意與無意之偏重,在教育與社會科學之統計工作上,亦有之。

機誤之所示者,乃從任取之樣本,而決定全部數目之可靠性。但機誤所示之變動,乃變動之純粹由於機遇者;且此種變動,較之由有意與無意而得之變動,可爲不甚重要。重要之變動之可靠性,機誤法無從斷定。

苟有一千卡片,而欲從其中任何二十張之平均數,決定其全部數目之特性,則此平均數之可靠性,關係甚重,不可不審愼考慮。每樣本共二十張,共五十樣本,但吾人不知所抽出者,爲何種樣本。或所得之樣本,其平均數之機誤限度,在眞正平均數之外。吾人並不知眞正平均數爲十二,因實際上祗知樣本,而不知全部數目也。吾人所能爲者,卽求某樣本之平均數,計算此平均數之機誤,然後知眞正平均數在機誤限度以外之機遇爲相等。至於其他複雜分子,使

樣本不足以代表全部數目者,不能因計算機誤而可以免除,惟有假設所得之樣本,純爲隨機者耳。

機誤之定義. 設有二十張卡片之樣本,而求其平均數與機誤,試問若再有二十張卡片之樣本,對於相當事實,有所知否?例如第二樣本之平均數,是否在第一樣本之機誤限度內?吾人對此能否斷言之曰不能。若閱**第十六表**,可知樣本中有一半之平均數,在已知機誤限度之內,其他一半之平均數,則在已知機誤限度之外。故吾人可斷言曰,苟抽出第二樣本,且其一切情形,與抽去第一樣本時相同,則第二樣本之平均數目,在第一樣本平均數之機誤限度之外者,其機遇相等,此原則閱表自明。

平均數之機誤(用P. E.或E表示)爲一對差數;一在平均數之上,一在平均數之下;眞正平均數在此二限度之內者,其機遇相等。此機誤之定義,有一假設,即各種差異,皆由機遇所致。

利用下表,可以確定眞正平均數所在之特別限度。此限度若用±E,±2 E,等表示者,可書之如下:

E－機遇相等

2 E－機遇爲4.5與1

3 E－機遇爲21與1

4 E－機遇爲142與1

5 E－機遇爲1310與1

6 E－機遇爲 19,200 與 1

7 E－機遇爲 420,000 與 1

8 E－機遇爲 17,000,000 與 1

9 E－機遇爲約 1,000,000,000 與 1[1]

若應用以上機率於**第十六表**第一樣本之平均數及機誤上,則平均數爲 12.40,機誤爲 .48,二倍機誤,其限度爲 11.44 至 13.36。意卽眞正平均數在 11.44 與 13.36 之間者,其機遇爲 4.5 與 1。若限度展開,則眞正平均數在此限度以內之機遇亦較大,此固不言而喻。

機誤與二十五分差或中數差之區別. **第三十五圖**共有二種次數分配,一爲原來量數之分配,一爲樣本平均數之分配。第二種分配用小黑點表示。每一小黑點,代表二十個數目之平均數。此分配之分布,較原來量數之分布爲小。每一樣本之平均數機誤,與各平均數之二十五分差相同。在此機誤限度內,有一半平均數,在此之外,亦有一半,此與任何分配之二十五分差之情形相似。

就統計學言,機誤與二十五分差二名詞相同,其數値亦相等,但二者之用途不一。若吾人硏究若干數量而欲知其離中趨勢,則求其二十五分差。若單就一樣本,而欲預料其離中趨勢,則求其機誤,其意則一也。有一個二十數之樣,

[1] C. B. Davenport, "Statistical Method," p. 14.

求其平均數,再求其機誤,則此機誤之意義,爲:若將其餘樣本,一併加入而各求其平均數,則此平均數之分布(或離中趨勢)應爲如何?設實際的從事取來之各樣本,並計算每樣本之平均數。設從事於五十或一百樣本,則此各樣本之各平均數,可用次數分配圖表示之。若從而求其機誤,則此機誤爲平均數分配之二十五分差。此二種觀念實相同。但其用法則不同。**有一實際分配而量其離中趨勢,所得者謂之二十五分差。若僅估計其離中趨勢,則謂之機誤。**二者均從最適宜之平均數兩邊展開,二者均含分配中量數之一半。

機誤公式之應用. 在某種情形之下,以下問題,可以發生:眞正平均數不小於12之機率爲何?設用第一樣本之平均數12.40,又使吾人不知眞正平均數之價值,亦可估計眞正平均數低至12之機率。理論平均數12,比實得平均數小 .40, 然則此無從探知之全體數量之平均數,較實得平均數小 .40之機率爲何?實得平均數(12.40)之機誤爲 .48, 故 .40可用$\frac{.40}{.48}$之分數表出之,卽$\frac{.40}{.48}=.83$ E。設作一次數曲線,以12.40爲平均數,以 .48 爲二十五分差,則加多樣本後,各平均數之分配形狀,可以想見。各平均數之平均數,固無法探知,但本樣本之平均數,可用作各平均數之平均數,因此爲惟一已知之數。本樣本所得之平均數,在眞正平均數上之機遇與在眞正平均數下之機遇爲相等。此點前已說明,茲不再述。

許多樣本之平均數作爲次數分配後,本樣本之平均數可在分配量表上之任何一點。眞正平均數不小於12之機率,可申說如下。試問在平均數之分配上,在12以上之面積,佔全面積百分之幾?橫坐標上之12,可用－.83 E表之。均方差與機誤之關係爲 P.E.＝.67 σ。故12在「式克碼」量表上爲－.56 σ。在平均縱線與－.56 σ 之間之面積,佔全面積之百分之二十一(21%)。故眞正平均數不比12小之機率爲.71即眞正平均數小於12之機率爲.29。但此不過估計而已,其實眞正平均數,吾人已知其爲12。

尙有一種問題,須用機誤解釋者:眞正平均數與實得平均數之差不出.3之外者,其機率爲何?吾人所知之唯一平均數爲12.40。現在吾人所研究之問題,可重述如下:無從探知之全體量數之平均數,在12.40±.30或在12.10與12.70之間之機率爲何?再假設各平均數作爲分配如前。又假設此分配之平均數爲12.40,因此爲惟一已知之數。此平均數之機誤爲.48,其全距爲－.62 E至＋.62 E,因.62爲.48之.3。改爲均方差,卽－.42 σ 至＋.42 σ。按表,在此限度內之面積,爲全體之33%。故眞正平均數在12.1與12.7之間之機率爲.33。其實眞正平均數爲12,但吾人苟根據一個樣本,固不知此數也。

機誤觀念尙有一種應用,可述之如下:欲知眞正平均數在固定限度內之機遇爲10與1者,則某樣本平均數前

後應有之距離爲多少?再用第一次二十張卡片爲例。該例之平均數爲12.40。再假定其他平均數之分配均環繞此平均數。此平均數之機誤爲.48,此卽爲其他各平均數分配後之二十五分差。玆所欲探求者,卽以平均數爲中點,從其前後展開,須到如何程度,始將面積之$\frac{10}{11}$包入,而祗餘下$\frac{1}{11}$,在此限度之外。$\frac{10}{11}$之分數爲.91,半在平均數之上,半在平均數之下。按表,分配之45%在平均縱線與1.65 σ 或2.46 E 之間。因機誤爲.48故此限度爲實得平均數12.40上下之2.46×.48或1.18。故應求之限度爲11.22至13.58。吾人可以確說眞正平均數在限度之間者,其機遇爲10與1。但切勿忘記:由實得之一樣本而言,眞正平均數在上述限度之外者,仍有$\frac{1}{10}$之機遇,而此十中之一之機遇,或絕無而僅有。

可靠性之量數之用均方差表示者. 除二十五分差外,均方差亦可作爲差異量數。二十五分差可用爲實得差異之量數,亦可作爲估計理想差異之用。從量表中點兩端展開,均方差包括全面積三分之二,二十五分差包括全面積之一半,故用均方差爲可靠性之量數時,其在實得平均數兩旁之距離較大。12.40±.72之解釋,若.72爲均方差時,卽眞正平均數在此限度之內者,其機遇爲二與一之比。用均方差表示可靠性之量數時,其算法與解釋,與機誤相似,其不同者,卽均方差包括全面積三分之二,機誤包括全面積二分之一。

其他統計常數之機誤. 機誤亦可表示其他統計常數之可靠性,如中數,均方差,及闢而生(Pearson)之相關係數。其解釋相同。若計算闢而生相關係數時,可表之如下: +.62±.05,眞正相關係數在.57與.67之間者,其機遇相等。

計算機誤時,有一假設,即分配必爲對稱或常態。若分配不對稱,則機誤失其通常意義。若分配爲鐘形時,或非極偏態時,則機誤對於各常數之應用,可保無虞。

機誤公式.

算術平均數:　$$P.E.=.67449\frac{\sigma}{\sqrt{n}},$$

在此式內,σ代表本樣本之均方差,n爲其次數之總和。

均方差:　$$P.E.=.67449\frac{\sigma}{\sqrt{2n}},$$

在此式內,σ代表本樣本之均方差,n爲其次數之總和。

闢而生相關係數(看第二十三章):

$$P.E.=.67449\frac{1-r^2}{\sqrt{n}},$$

在此式內,r代表相關係數,n代表次數之總和。

可靠性之用均方差表示者,可用以下關係求得: $P.E.=.67449\,\sigma$。

結論. 二十五分差可用作任何分配之差異量數。若所研究者非爲原來量數,而爲各樣本之平均數,則其差異

亦可用二十五分差表示之。通常祇用一樣本,此樣本或為二十個數目,或為二千個數目,不定。從此樣本,求一平均。欲知此一樣本之可靠性,假設取得許多樣本,並求其平均數。若每一樣本之數目極多,則各平均數之變動極小。若每樣本之數目,密集於其平均數,則各樣本之平均數,變動極微,否則,變動較大。每樣本數目之分散與密集,及每樣本之次數數目,二者可以決定該樣本之平均數之可靠性。吾人苟多得樣本,並計算其平均數而作為分配,則吾人可用此許多平均數之二十五分差而量其分布之程度。若不取得許多樣本,並求其平均數之分布,祇用一樣本以斷其他平均數之變動,則吾人不啻用一理想之差數,而視之為機誤。凡二十五分差之非由實際得到,但由估計或預占而得到者,謂之機誤。

問題 1. 某樣本之算術平均數為 64.78;其均方差為 6.34 組距;每組共 .5 單位。本樣本之次數為 100。

1. 求算術平均數之機誤;求均方差之機誤。

2. 求均方差之均方差。

3. 眞正平均數大於 68 之機率為何?小於 64.78 之機率為何?在 64 與 65 之間之機率為何?

4. 欲知眞正平均數在某限度之機遇為五與一之比,試問實得平均數兩旁之距離為何?

5. 眞正平均數與實得平均數之差,不出 .3 之外者,其機遇為何?

6. 倘取 150 個樣本,計算各平均數,試問此 150 個平均數分配之

二十五分差爲何？又問若共有300個樣本，每個之數目與第一樣本同，此300個樣本平均數之二十五分差爲何？

7.　假設平均數爲66。作一個150平均數之分配。在圖上指出平均數之地位。

問題2.　何謂均方差之均方差，試說明之。如欲證驗此定義，應用何種實驗手續？

問題3.　假設一班學生共500人，從中隨機取出50人此50人身高之平均數及此平均數之機誤已經求得。又假設此一班之最高50人之平均身高及其機誤，亦已算出。試問二機誤之中，以何者爲小？其理由如何？

問題4.　設有甲乙二城。甲城之人口爲1,000,000；乙城之人口爲1,000。從每城取一500人之樣本，而求其平均身高，平均年齡，及其他品性。試問二平均數中，何者比較的足以代表某城人口之可靠的指數？

500個人樣本之平均數，是否受其原來全部人口多少之影響？

此樣本平均數之機誤，是否受其原來全部人口多少之影響？

樣本代表全數之一半者，是否較樣本代表全數之$\frac{1}{2000}$者爲可靠？二者之可靠性何以不同？

若此500人之樣本包括全數，則此樣本平均數之可靠性，是否增高？其理由可用機誤公式說明否？

若將全部數目縮小而與樣本之數目相等，則機誤之意義，有何缺點？

第二十二章

相關表

變量之相關. 探求變量之相關,為一種科學問題。倘其相關得到,則問題之陳述,亦能較為清晰。然必有二種變量,互相影響,互成因果。苟變量之性質不明,或測量之單位不詳,則統計之功效全失。故解決一科學問題之主要點,卽為探求其變量相關也。

在科學初步時期,探討家之工作,全屬形容性質。卽將新探之物,詳細說明,然久而久之,此探討家必變其態度,易其方法。進而尋求新事物之相關,或新事物與舊事物之相關。考古家發現埃及古墓,內藏木乃伊及其他古物,此為其科學工作之第一步。苟欲對於科學有所貢獻,此考古家必求所得新事實之相關或新事物與歷史所載事實之相關.

物理學家若徒研究聲浪穿過各種物質之速度,亦不過形容其事耳。若進而求此聲浪與物質內容,如密度,化學成分,沸點,傳電力,等之相關,則其研究較為高深而多興趣如是則此物理學家或可發現此物質內之某分子與其傳電力,毫無關係,而他分子與其傳電力,關係極大。吾人之所以能預占與支配宇宙間自然界之一切現象者,皆由此種探求相關之功也。

相關統計所研究者,為變量之相關。前此所討論者,均

爲一種量數之特性,如集中趨勢與離中趨勢。今後所討論者,爲數量變量之相關。

欲了解相關表之理,可舉例說明之。如決定建築材料之堅固,爲一種科學問題。屋樑之中部加重則彎曲,吾人卽可研究加重與彎曲之度。此處之二變量,一爲載重,以磅表之;一爲彎曲,以長表之。其相關成正比例。因一個變量漸加,其他變量亦漸加。又屋樑之形狀,亦可斷定彎曲之度。若以樑之厚薄與其彎曲爲二變量,則厚薄彎曲之增減,其相關成反比例。凡一問題,用科學方式表出,卽包含二三種變量,吾人之目的,卽斷定此變量相關之性質。

吾人若考察氣壓與氣候之關係,則氣壓爲一變量,沉澱(表示晴風)爲又一變量。故無論解答何種問題,一用變量表示,則問題之思考與解決方法均明晰矣。

變量(Variable)爲含有一列價值之物,如年薪,高度,樑曲之爲糎,氣壓之用水銀表示,薪金之用銀元表示。變量可歸入組距,十或二十組距不等。有時祇有二組,設有千人,則在平均數上爲一組,在平均數下者,又爲一組。一爲高,一爲低。凡品質歸入二組時,仍可以變量視之,不過此變量祇有二組耳。

試討論科學上一問題:蚊與瘧之關係。蚊瘧爲二變量,均可以數表之。與蚊接觸之次數,亦可以求得。設有二羣人,一羣爲蚊所咬,一羣則非。此種試驗,可決定蚊與瘧之關係,

不過每種變量,祇有一組,而無繼續不斷之數耳。化學上用定性方法鑑別金屬,亦可以統計法助之。如某樣本中所含金屬之分量然,此樣本中金屬之「有」與「無」,即爲二變量也。

自變量與倚變量. 在任何種實際情形上,若用科學方法定有自變量與倚變量之分。如用日光,潮濕,風吹速度,及沉澱等,研究氣壓與天時之關係然,必有一已知之變量,用之以占其他不知之變量。已知之變量,謂之自變量(Independent Variable)。其他待占之變量,謂之倚變量(Dependent Variable),因其量必倚賴已知之變量也。預占天時,則氣壓,爲自變量,而所得之天時結果,謂之倚變量。

若研究屋樑之形狀與其彎曲,則形狀爲自變量,至於彎曲,則由形狀而定,謂之倚變量。彎曲之性質與程度,全由屋樑之形狀而定。

若以入學考試與在校學業相比較,則入學考試之成績,謂之自變量;在校學業,謂之倚變量。

正相關與負相關. 若有二變量,其一變量漸增時,其他一變量,亦隨之漸增,此種相關,謂之正相關。反之,若一變量漸增時,其他一變量,反因之漸減,則此種相關,謂之負相關。若有一羣學生,而施以算術測驗。苟測驗分數用時間表示,而題目爲一定數目者,則時間愈長,算術之能力愈低。故算術能力與時間爲負相關。反之,若測驗分數爲十分鐘所解答之題,則能力愈高所解答之題愈多,而題數與能力成

正相關。

相關之程度. 二變量可成正相關,但其相關同時不必完全。當一變量增加,其他一變量固可有增加之趨勢,但其相關不必個個一律完全。如體高與體重然。大概論之,體高者固比體低者較重。體高與體重成正相關,但其相關,不能完全。若比較二對相關之數,而均爲不完全之正相關者,則一對之相關必較其他一對爲密切。相關爲統計學上重要之一部。玆舉數例如下,以示完全正相關至完全負相關之程度。

鋼條之重與其容量之相關爲完全正相關。體高與體重亦爲正相關,但不若鋼條之重量與容量相關之完全。故體高與體重之相關,爲正而非完全者。相關之爲零者,或完全不相關,如一人之薪金與其所穿之鞋之大小,卽有相關,亦不過甚微小之正相關耳。物之正方重,與其價值,大約爲完全不相關,或零相關。工作之時間,與工作之精巧,爲不完全負相關。完全負相關之例,如溫度不變時,壓力與氣量之相關。一變量漸增時,其他一變量漸減。

社會科學上之相關. 純粹科學之相關,常包數量;凡物理學家,天文學家,工程學家,莫不知之甚詳。若物學理家,欲研求許多事實,則將此事實於圖上而求其關係。在生物學與社會科學上,此種定量方法,應用不多。蓋其事實多不便於數量之測量也。

科學日精,數量之工作漸多。如生物心理各科學,莫不利用定量之法。久而久之,一切社會科學,亦可用數量研究,如生物學然。

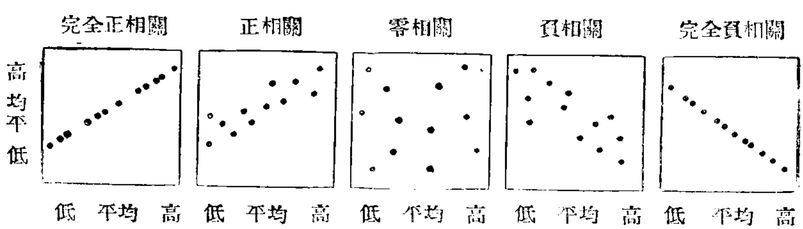
凡科學所研究者,常爲變量。但科學中,有能得精確之測量者,有不可得者。如物理學家,可將不相關之分子除去,而得較精確之測量。然後求相關之有無與性質。在生物學與社會科學上,此種除去不相關分子之手續不易。例如經濟學家研究人口稠密與城市屋租之問題。欲得屋租之精確調查,及每方居民之實得數目,固非難事,但其工作,決不能如物理學家之在一實驗室然,除去不相關之分子,而得其所欲得者。凡分子之與屋租有關者,如交通,地勢,商業中心,公園鄰近,城區聲名,房屋年代等,經濟學家,不能支配自如,而適其意。必也就事論事,其複雜情形,彼此互相影響之處,實難條分縷析。根據事實定其趨勢,固不難爲,若用圖表示,則事實分散,決不能如純粹科學之精粹不雜。

統計上相關方法,卽用以幫助解決此種分散變量之相關者。純粹科學所用之方法與此略有不同,但就數學之理而言,則不甚異。物理學家用方程式而形容其事實,此方程式之常數,可由最小方法決定之。經濟學家及心理學家用迴歸方程及相關表而形容其事實。其所用之方法,雖有形式上之不同,而理論上,則相脗合。

分布圖. 分布圖乃用圖形表示相關之方法。一察此

圖,不特相關之有無可知,即相關之程度,亦可了然。

第三十六圖共五分布圖,表示各種不同程度之相關。第一部表示二變量之完全正相關,如鋼條之量與重是也。每一分子,在圖上用一點表之。此分子或爲個人,而吾人研究其高度,體重,智力,薪水。此分子或爲物體,而吾人研究其壓力容量等。若研究人口稠密與屋租,則分子爲城區或房屋,或每區或屋,必有二變量而成一偶數。每一分子,在圖上,用一點表之。



第三十六圖　正負分布圖

圖上每點,表示二事,自點至底線上垂直線之距離,表示一變量;自點至左縱線地平線之距離,表示又一變量。有時測量之數,不自零度始。

在第三十六圖之第一部上,x一變量,在底線上,自左向右,讀由小而大。y一變量,在縱線上,自下而上讀,由小而大。x一變量漸大時,y一變量亦隨之漸大。故謂之完全正相關。

用圖之第二部,表示十二對觀察之分布圖。其相關爲

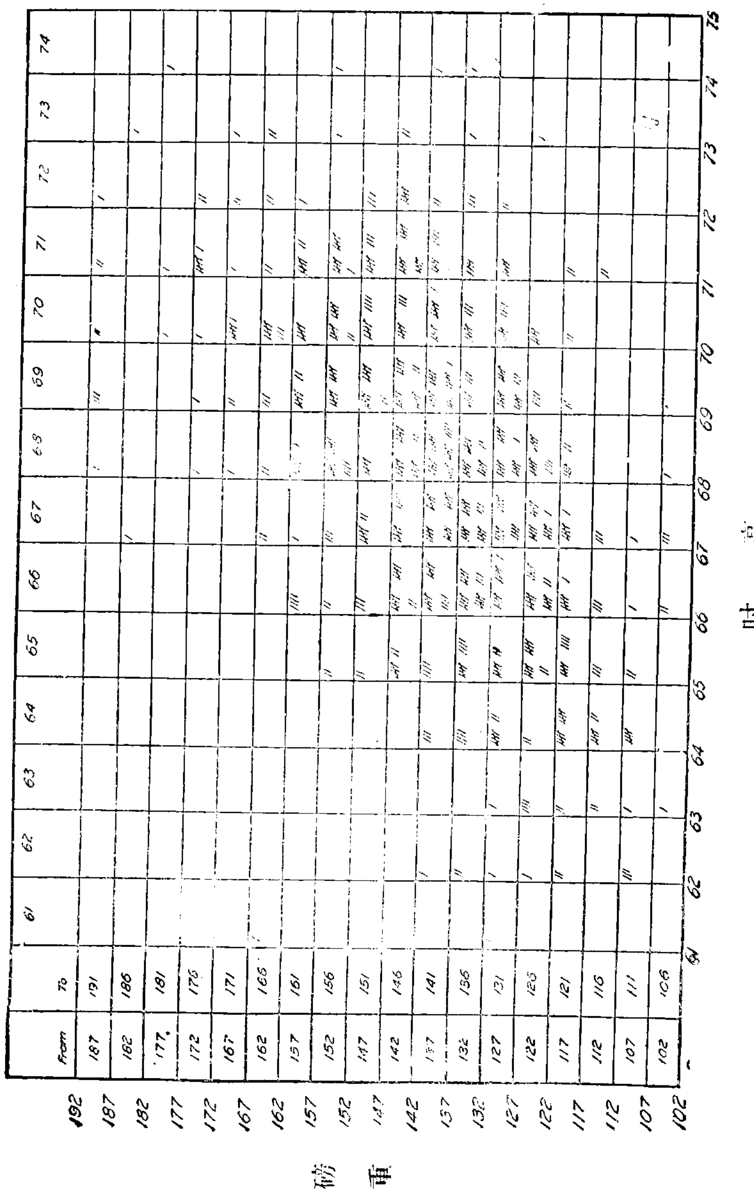
正,而不完全。一視此圖,卽知當x—變量漸增時,y—變量亦隨之漸增,但y值之增,不必與x值之增在每觀察上,成同一嚴密之正比例。如在三個觀察上,x之最低值,共有不同之相當y值三個。又x之最高值,共有不同之相當y值二個。此足見其相關之不絕對互依與精確。但六個最高x—值相當之平均y—值,比六個最低x—值相當之平均y—值爲高。

第三部表示二變量之不相關或零相關。此處六個最高x—值相當之平均y—值與六個最低x—值相當之平均y—值相等。一變量變時,則其他一變量毫不受影響,此之謂不相關或零相關。

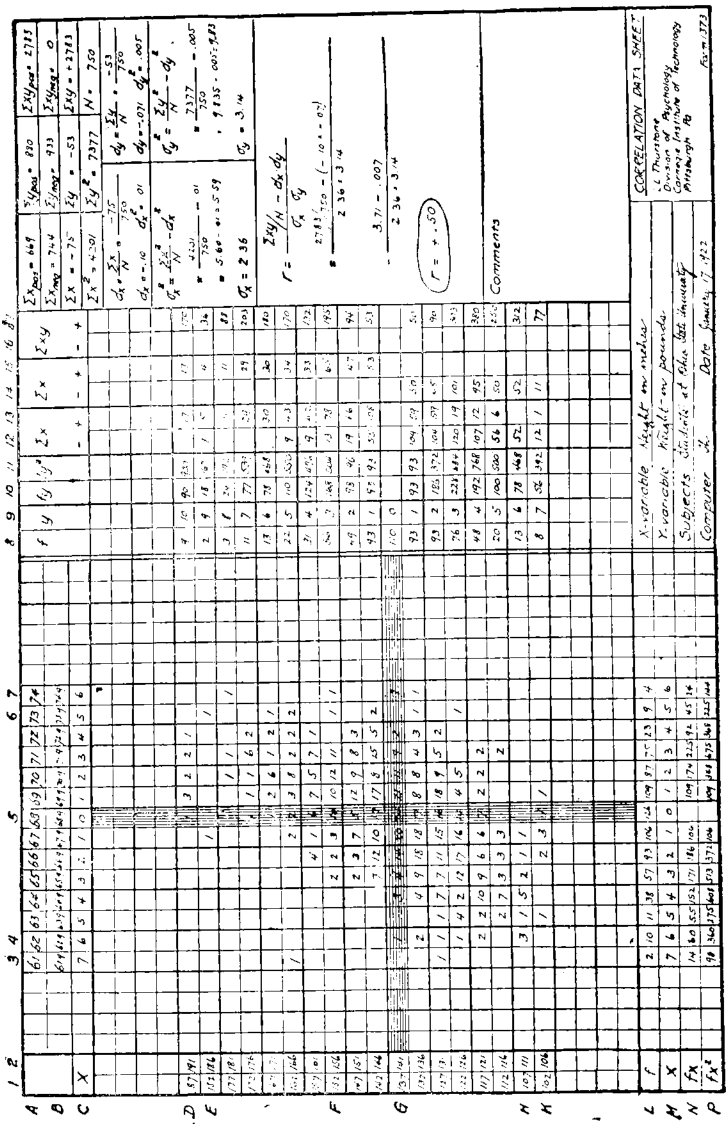
第四部表示二變量之負相關而不完全者。當x值增加時,y值反因之而減,故曰負。但六個最高x—值相當之平均y—值,較六個最低x—值相當之平均y—值爲低。例如每點鐘解答算題之數及每題所需之時,當問題之數漸增時,每題所需之時間漸減。此之謂負相關。第四部之分布圖,爲負相關之不完全者。

第五部表示二變量之完全負相關。當一變量增加時,其他一變量隨之減少,且成一定之反比例,一變量之量表,與他一量表,或大小不同,但二者成完全反比例則一也。

研究單獨變量時,此變量可歸入組距,而便於列表。在作分布圖時,可用同一方法。第三十七圖代表烏海烏(Ohio)



第三十七圖　體高體重之分布圖



第三十八圖 相關紙❶

❶ 此紙可由下列地址購得:
C. H. Stoelting & Co., 3047 Carroll Ave., Chicago, Ill., U. S. A.

省立大學750學生體高與體重之相關分布圖。此統計圖之每人,在圖上方格內,用一符號(1)作記。每符號代表每學生之二事實。其體重用磅在圖之左縱線表示。其體高用吋在圖之底線上表示。

第三十八圖示同一事實,而用相關表出。在一分布圖上,每一分子,均用一符號表示;在一相關圖上,每方格內用一數目,載其次數,此爲分布圖與相關表之不同處。否則二者皆同。但分布圖實爲相關表之初步。

吾人若閱**第三十八圖**,見體高組距漸高時,則其組之平均體重亦漸高。圖之上左角與下右角,比較空虛,而圖之下左角與上右角比較充實。此蓋由排列法使然。大約事實愈密集於對角線上,則相關愈高

若二變量不相關,則圖形完全不同。例如將吾人薪水之高低,與鞋之大小表列,則必乏顯著之相關。高薪與小鞋,或低薪與大鞋,均可發現於圖上。其結果圖之中央與四角,均充滿事實,而無餘空。既乏對角線形之存在,焉有相關之可言。

爲簡捷便利起見,若祇作分布圖或相關表而不求其相關係數或迴歸係數,亦可知二種變量之相關趨勢。從相關表上,可得作圖,與曲線及平均數等之條件。但較詳情形,則非相關表所供給矣。

問題1.　作一圖明示體高61, 62, 63等吋者之平均體重(參看第

三十七圖)。

問題2.　作一圖學生體重 105, 110, 115 等磅者之平均體高。試問與汝同一體重之學生,其平均體高如何。又問與汝同一體高之學生,其體重爲何?

問題3.　作兩圖,一則從已知各學生之體重而求其平均體高,一則從已知各學生之體高,而求其平均體重。

第二十三章

關而生相關係數

相關係數(Correlation Coefficient)爲一指示二變量相關程度之數。可自 +1 而至 −1。相關爲完全正時,則相關係數爲 +1。相關爲完全負時,則相關係數爲 −1。二變量不相關時,則係數爲零。其他係數之値,均在以上數種係數之間。若係數爲 +.8,則分布圖上各點,頗密集於對角線上;若係數爲 +.3,則分布圖上各點,散集於對角線上,但相關仍存在。體高與體重之相關爲 +.5。[1]然則相關係數爲一種用數字形容分布圖之方法明矣。不過分布圖所載之事實,較一單獨數目爲多耳。所研究之相關數量極多時,相關係數,可視爲一種簡筆或指數,如是凡諳統計者,一望而知其意義之所在。分布圖與相關係數比較,分布圖固較完美,因從分布圖可得相關係數,而相關係數決不能逞分布圖之內容也。然相關係數,簡明不繁,自亦有其長處在。

關而生相關係數之求法,常照下列公式:

$$r=\frac{\Sigma xy}{n\cdot\sigma_x\cdot\sigma_y},$$

在此式內,

---

[1] 初學者常稱相關係數爲百分,誤矣。+.5 之相關,應讀爲"十分之五",或"五十"。此非百分,乃一純數也。

x ＝X一值與X之平均數之差數

y ＝Y一值與Y之平均數之差數

n ＝分布圖上之量數總數

σ_x＝X之均方差(或標準差)

σ_y＝Y之均方差(或標準差)

r ＝隅而生之相關係數

欲從上式計算相關係數,必知xy差數之積,量數之總數,與二個均方差。

初學者最好用五十個量數左右,從眞正平均數上,求各差數,再求均方差,最後求相關係數。此法較用假設平均數法,可多得練習。但在實際工作時,則以用已經備好之相關紙爲是。用相關紙,則計算可較易,且各數目均有一定地方。

不用差數法,亦可計算相關係數。[1]此方法當含極大數目,有時不易對付。但可不用原來X與Y數目,而用他數代之,較爲便利。公式如下:

$$r=\frac{\Sigma(XY)-n\cdot m_x\cdot m_y}{\sqrt{\Sigma(X^2)-n\cdot m_x^2}\cdot\sqrt{\Sigma(Y^2)-n\cdot m_y^2}},$$

[1] 欲知詳細情形,可參看 L. L. Thurstone, "A Method of Calculating the Pearson Correlation Coefficient whithout the Use of Deviations," Psychological Bulletin, January 1917。此公式有時亦稱爲艾爾公式(Ayre's formula)。艾氏發表此公式,乃在作者數年之後。

在此式內,

m_x=x－值之平均數

m_y=y－值之平均數

現將分子之有影響於相關係數者,加以討論。若觀察不能精確,則相關係數自較觀察精確者爲小。

有時二變量之相關,因有第三變量在,此變量若不加以限制,或使之不變,則將大有影響於相關係數。例如作一相關表,而求兒童體高與算術測驗分數之相關,則必爲低的正相關。相關之能爲正者,因有年齡分子,未嘗計及,但年齡與體高確有關係。苟再作一相關表,而求體高與算術測驗分數之相關,但兒童之年齡均相等,則其相關係數必爲零。何以故,因在第一次時,年齡爲第三變量,加入而影響相關;在第二次時,年齡已受限制,而使之不變。

影響相關係數之又一分子,爲迴歸方程之直線性。若分布圖上之線或對角線趨勢爲直線者,或不甚直,而實際上可以直線視之者,則相關係數比圖上之線爲曲線時之相關係數爲大。若分布圖之線爲曲線者,則應求**曲線相關係數**(Eta Coefficient),[1]或**相關比例**,但此法易受分組之影響,此其缺點也。

尚有一分子,足以使相關係數變低者爲變量全距離

---

[1] 欲知曲線相關係數,或相關比例,參閱 (G. U. Yule, "An Introduction to the Theory of Statistics," Chapter X.

之縮小。設作兒童體高與年齡之相關表,而年齡自零至二十。如是,則其相關,必甚明著平均言之,凡兒童之年較長者,體亦較高。再設兒童年齡之限度為10至12歲。如是,則十歲以下及十二歲以上之兒童,皆不在表上;而身材與年齡之相關,亦因之變為甚低。若此二年分為二十四組距,每月一組,再求年齡與其相當體高之相關,其相關亦低。故學者宜注意及此。又如大學學生智力測驗之成績與學業成績之相關,較之中學生與小學生學業成績之相關必低。何以故,蓋大學學生為精選人材,其智力之全距離縮小,而屬於量表之最高一部。倘施行智力測驗於一萬普通人民(如施行於軍隊),而斷定此輩在大學之能力,則相關係數必高,因智力量表之低端,均為不及格或不識字者。

量數總數之大小,亦可斷定相關係數之可靠與否。苟數目極小,如二十或三十,而係數又極低,則此相關之程度,殊不可靠。數目小時,固可斷定相關之有無,但欲求相關之確度,則非有數百或數千之數不可。相關係數之可靠程度,可以其機誤為標準。

迴歸線. 設一人之體高為70吋,今欲根據**第三十八**圖而求其體重。估計最善之方法,即求凡體高為70吋者之平均體重。參看原圖,在體高為70吋之縱行上,求其平均體重。共87人,其平均體重為150磅。故曰,凡體高為70吋者,其體重之最可能數為150磅。倘每縱行上,均求其平均數,作

一線,連接各平均點,則此線或其相等之最適合直線,稱之曰**體重在體高上之迴歸**(Regression of Weighton Wtature),或曰 y 在 x 上之迴歸。此種圖或此種線,能使吾人根據體高與體重之關係,可由已知之體高,而預占其體重,不致大誤。

反言之,吾人已知凡體高70吋者,其體重大約必為150磅。若一人之體重為 150 磅,試問其體高為多少?若曰70吋,則未必合理。正當之手續,是從凡體重為 150 磅者之中,求其平均體高。其法可從**第三十八圖**横行之體重為 150 處,而求其平均體高。按此行共49人,其平均體高為69吋。由此可知體高為70吋者,其平均體重固為150磅,但體重150磅者,其平均體高,不必為70吋。蓋一則吾人求**縱行** (Vertical Column)上之平均,一則吾人求**横行**(Horizontal Column)上之平均。

各横線上之平均點,可用線連之,此線稱之曰**體高在體重上之迴歸**(Regression of Height on Weight),或曰 x 在 y 上之迴歸。學者宜注意所用者為何迴歸線不可錯亂。從已知之 x,而求平均之 y 值,但此平均之 y 值為已知者。則其相當之平均 x 值,不必為原來之 x 值。

迴歸方程如下:

y 在 x 上之迴歸(從已知之 x 值而預占 y 值)

$$y = r\frac{\sigma_y}{\sigma_x}x$$

x 在 y 上迴歸(從已知之 y 值而預占 x 值)

$$x=r\frac{\sigma_x}{\sigma_y}y,$$

在上二式內, x 與 y 均爲量數與平均數之差數。

平常可根據實得量數 x 與 y 之值,而作預占,不必用此量數與平均數之差數也。下式於實際上極有用:

y 在 x 上之迴歸(從已知之 x 值,而預占 y 值)

$$Y=r\frac{\sigma_y}{\sigma_x}X-\left[r\frac{\sigma_y}{\sigma_x}m_x-m_y\right]$$

X 在 Y 上之迴歸(從已知之 y 值而預占 x 值)

$$X=r\frac{\sigma_x}{\sigma_y}Y-\left[r\frac{\sigma_x}{\sigma_y}m_y-m_x\right]$$

在上二式內, X 與 Y 均爲二變量之實在量數,而非量數與平均數之差數。

問題 1. 作一相關表,而求二變量之相關係數。再作一相關表,其中之一變量祗用五分之一之數。求其相關係數,且示此數比較爲小,其故安在?

問題 2. 以下問題,可爲討論迴歸線之助。設十二歲兒童之平均智力年齡爲十二。試問智力年齡十二之兒童,何以不必爲十二歲?

問題 3. 作一表載體重與其相當體高。決定汝之體重。此爲體高在體重上之迴歸,抑爲體重在體高上之迴歸?若有一人不知迴歸線之意義者,則其繆誤之推論如何。

第二十四章

闢而生相關係數之計算法❶

計算相關,若不嫻熟,常有翻閱教科書之必要。若事實紙上,印有算法說明,則閱書之手續,可以免去。余嘗見巧於計算相關者,常有將其計算,散布於若干紙上,而乏一定之步驟。計算而乏次序,則查對錯誤,幾不可能。

余所用之相關紙,較信紙大一倍,可以摺好,與信紙及他種報告,一同保存。紙上有一空地,可作計算時起稿之用。各種項目,如次數,總數,方數等,均在紙上,有一定地位。查對計算經過,極為利便。節省工夫,莫甚於此。

事實張之應用說明. **第三十八圖**縱行上,有時用數字表出,以利說明。但原來事實張上,並無此種數字。橫行上,用大寫字母表出。縱行與橫行,以後統稱行列。紙上每方格則用縱行數與橫行字母之交切點表之。故 C–3 表示張上之一方格,其中之數字為 7。此張之事實,則從惠司得(Carl J. West) 統計教科書❷上採取

❶ 此章係由(Journal of Educational Research, June, 1922),重印,並得其編輯部之許可。

❷ Carl J. West, "Introduction to Mathematical Statistics" Columbus, Ohio: R. G. Adams and Company, 1918. Page 67.

1. x－變量,可採用任何組距。在本事實張上,組距爲1。

2. 在橫行A上,記載各組距(61,62,等)。排列之法,數目小者在左,大者在右。

3. 橫行B載組距之高限度。故橫行A載低限度,橫行B載高限度。

4. 擇一與y適宜之組距。在本例爲五。縱行1載組距之低限度,縱行2載組距之高限度。將組距之最低限度數與最高限度數,一併登載,可免混淆。排列之法,數目小者在下,大者在上。

5. 將次數登載於方格內,卽在D與K之間,及3與7之間。

6. 在橫行L上,載各橫行之總次數。

7. 在縱行8上,載各橫行之總次數。

8. 將L橫行上之次數相加,而事實張之上右角上N處寫(N=　)。在本例N=750。此爲總次數。

9. 將8縱行上之次數相加,而與L橫行之總數相對驗二總數應相等。

10. 擇任何一組爲x－變量之假設原始點。在本例爲縱行5。此原始點可在任何地方,但以愈近平均數者爲愈好。觀察卽知,不必將平均數算出。用藍色鉛筆將縱行5畫出,須此行與圖上之他行,顯然不同。

11.y－ 變量之作法相同。用橫行G為原始點，亦用藍色鉛筆畫出。現相關圖，可照縱橫藍色線分為四個分方。但此四個係根據假設平均數而作，並非根據眞正平均數而作，學者宜注意及之。

12. 在M－5上記載零點。零之左右記載法，參看橫行M。橫行C與橫行M作法相同。

13. 在橫行N上，記載fx積。此積為橫行N上二行之積。

14. 在橫行P上，記載fx^2積。此積為橫行P上二行之積。

15. 在G－9處，記載零。零之上下，均記載1,2,3,4各數如縱行9。

16. 在縱行10上，記載fy積。此積為縱行10左面二縱行之積。

17. 在縱行11上，記載fy^2積。計算甚易，一視即知。

18. 在橫行D上，假設原始點之右之數為3,2,2,1,等。在此數目之上，橫行C上所載者為1,2,3,4。橫行D上之每數，用相當C行之數乘之，總之如下：

$$(3\times1)+(2\times2)+(2\times3)+(1\times4)=17$$

此數17，可載在D－13上。假設原始點(縱行5)以右所得之和，均載在縱行13上。原始點以左所得之和，均載在縱行12上。筆難盡述，察圖自明。

其餘原始點左右每橫行上之數,均照此算法。例如在橫行F上,其和如下:

左面:$(3\times1)+(2\times2)+(2\times3)=13$

右面:$(10\times1)+(12\times2)+(11\times3)+(1\times5)+(1\times6)=78$

13與78二數,載在F-12與F-13方格上。

若用計算機運算,較爲迅速。其法如下:

先論橫行 f 之左面。置指於機之下右端 1- 鍵上。此鍵代表C-行上之差數 1。按鍵三下,代表次數三。移指至鍵 2,按二下,代表次數二。再移指上一鍵,按二下,代表次數二。總數13,記在機上,可以無須心算。計算時,祇須注目於相關表上,不必看計算機。

次論橫行F之右面。置指於機之下右端 1- 鍵上。此鍵可按十下,或此鍵左面之鍵,可按一下,代表次數十。移指於鍵之上一鍵,按十二下,或按二下,再按其上鍵一下,代表次數十二。再移指上一鍵,按十一下,或按一下,再按上一鍵一下。移指上二鍵,按一下,再移指上一鍵,按一下。總數78卽載在機上,可轉載於 F-13 方格內。移指按鍵,事極簡易,卽初學者,亦不難爲。

19. F-12方格內之13與F-13方格內之78,其差爲65,載在 F-15 方格內。凡縱行12與13之各數,均照此算法。

其差數載在總行14與15上。倘12－行之數較13－行之數爲大,則其差載在15－行上。二者不可紊亂。

20. 關於y－變量之假設原始點(橫行g)之上部,記載縱行9與14之積於縱行16上。縱行9與15之積,計載於縱行17上。此爲假設原始點上部之說明。原始點下部之說明適相反。縱行9與14之積,載在縱行17上;縱行9與15之積,載在縱行16上。以下算法,較爲清楚:

	縱行
在假設原始點之上:	(9) × (14) = (16)
	(9) × (15) = (17)
在假設原始點之下:	(9) × (15) = (16)
	(9) × (14) = (17)

尙有記憶之法,卽注意:縱行14與15,在原始點之上,具相同排列。但在原始點之下,則排列相反。

再有一記憶之法:卽縱行9之y－差數,在假設原始點之上爲正,在下爲負。此蓋由於作圖時,縱行1之y－變量之數之大者在上,小者在下。現若以縱行9之正號y－差數乘縱行15之正號Σx,則得正的Σxy,而記載於縱行17上。若以縱行9之負號y－差數乘縱行15之正號Σx,則得負的Σxy,而記載於縱行16上。若以縱行9之正號y－

差數乘縱行14之負號Σx,則得負的Σxy,而記載於縱行16上。若以縱行9之負號y－差數乘縱行14之負號Σx,則得正的Σxy,而記載於縱行17上。

21. 以上手續完後,即可求各數之和,而計算相關係數。將假設原始點左面N行上之數目相加。記載其和於表之上右角$\Sigma x_{neg}=$ 處。在本例此數爲744。

22. 將假設原始點右面N行上之數目相加。記載其和於表之$\Sigma x_{pos}=$ 處。在本例爲669。

23. 求Σx_{pos}與Σx_{neg}之差,而記載於$\Sigma x=$ 處。在本例爲－75。其關係爲$\Sigma x=\Sigma x_{pos}-\Sigma x_{neg}$切勿寫錯$\Sigma x$之符號。

24. 將假設原始點兩旁之P橫行上之數相加。記載(4201)於表之$\Sigma x^2=$ 處。

25. 將假設原始點上面縱行10之數目相加;記載其和(880)於表之$\Sigma y_{pos}=$ 處。

26. 將假設原始點下面縱行10之數目相加;記載其和(933)於表之$\Sigma y_{neg}=$ 處。

27. 求Σy_{pos}與Σy_{neg}之差,其數爲(－53)。而記載於表之$\Sigma y=$ 處。其關係爲$\Sigma y=\Sigma y_{pos}-\Sigma y_{neg}$。切勿寫錯$\Sigma y$之符號。

28. 將縱行11之各數相加,記載其和(7377)於表之$\Sigma y^2=$ 處。

29. 將縱行16之各數相加,記載其和(零)於表之$\Sigma xy_{neg}=$ 處。

30.　將縱行17之各數相加,記載其和(2783)於表之$\Sigma xy_{pos}=$　處。

31.　求Σxy_{pos}與Σxy_{neg}之差,其數爲(+2783)記載於表之$\Sigma xy=$　處。其關係爲$\Sigma xy=\Sigma xy_{pos}-\Sigma xy_{neg}$。切勿寫錯$\Sigma xy$之符號。

以上各和數,可以直接代入公式,依式解算,可得係數爲+.50。上述記載及計算相關係數之方法,若慣於計算,約需時自四十分鐘至六十分鐘不等。若練習加多,時間尚可減少,若專靠說明法,而不參看相關表,反不易明瞭。苟慣於看表,而廢除文字之說明,卽一素無統計智識之通常書記,亦可訓練之而使之精於相關係數之探求。

第二十五章

等級相關

二變量爲等級(Ranks)而非爲原値(Original Values or Scores)時,則相關係數,祇可近似。均方相關係數或闢而生相關係數,祇可適用於對偶量數上,而不能適用於對偶等級上。故事實爲等級時,祇可用一與均方相關不同之方法。有時,即對偶量數爲原値而非爲等級,亦以用等級相關方法爲是,以其簡捷而便利,不若均方相關之繁瑣耳。

等級相關之係數,可用下式表出之:

$$\rho=1-\frac{6\Sigma(k_x-k_y)^2}{n(n^2-1)}=1-\frac{6\Sigma d^2}{n(n^2-1)},$$

在此式內:

ρ = 等級相關係數

r = 均方相關係數,或闢而生相關係數

n = 次數之總數

k_x = x-値之等級

k_y = y-値之等級

d = 對偶等級之差

用等級相關法,求二變量之相關係數時,必有一假設

卽二變量之分配,均各爲常態。實際上,此假設尙屬妥善。

第十七表說明此法之應用。第一縱行爲二十四個人。故常數 n 爲24。第二及第三縱行爲此二十四個人每人之原來對偶量數。此數代表測驗分數。第四行所列者,爲測驗 X (Test X) 分數之改爲等級者。例如個人 20 在測驗 X 之分數爲 17。因此數爲各數之最低者,故其等級爲 1, 記載於第四縱行。同樣,個人 17 之分數爲 198。因此數爲各數之最高者,故其等級亦爲最高,卽 24,記於第四縱行上。實則第四行之數,悉爲第二行各數之改爲絕對等級者。若有數個數同者,則各數須互相均勻,前已述之矣,茲不贅。第五行所載者爲第三行各數之改爲絕對等級者。

第二步卽在 d 縱行上,載第四與第五行等級之差。例如第一個人之二等級 12 與 14 之差爲 2, 此數卽載在 d 縱行,而不計其符號。最後一行,記載 d 行各數之方。因此爲二等級差數之方,故正負號可不計及。

將最後一行各數相加。其餘手續,卽將各數代入公式,如**第十七表**。等級相關係數,用 ρ 表之;在本例爲 +58。

若分配爲常態者,則 ρ 值可用式改爲 r 值:

$$r = 2\sin\frac{\pi}{6}\rho$$

但爲P值改爲 r 值之手續便利起見,特作**第十八表**。根據此表,等級相關之係數,可改爲均方相關之係數。

問題1. 用第十七表之事實而求均方相關係數,將此數與ρ相比較,並與ρ之r相當值相比較

個人	分數 X	分數 Y	等級 X	等級 Y	等級之差	差方
#	X	Y	k_x	k_y	d	d^2
1	104	18	12	14	2	4.00
2	83	14	9	8	1	1.00
3	155	17	20	12.5	7.5	56.25
4	165	23	21	20.5	.5	.25
5	178	19	22	15.5	6.5	42.25
6	147	25	18	23	5	25.00
7	22	12	2	5	3	9.00
8	84	10	10	1.5	8.5	72.25
9	27	19	3	15.5	12.5	156.25
10	63	21	7	18	11	121.00
11	117	17	14	12.5	1.5	2.25
12	94	23	11	20.5	9.5	90.25
13	74	16	8	11	3	9.00
14	59	13	5.5	7	1.5	2.25
15	118	15	15	9.5	5.5	30.25
16	131	21	16	18	2	4.00
17	198	24	24	22	2	4.00
18	108	12	13	5	8	64.00
19	59	11	5.5	3	2.5	6.25
20	17	15	1	9.5	8.5	72.25
21	29	10	4	1.5	2.5	6.25
22	189	21	23	18	5	25.00
23	152	26	19	24	5	25.00
24	146	12	17	5	12	144.00

$$\Sigma(k_x-k_y)^2=972$$
$$6\Sigma(k_x-k_y)^2=5832$$
$$n=24$$
$$n(n^2-1)=13{,}800$$
$$\rho=1-\frac{6\Sigma(k_x-k_y)^2}{n(n^2-1)}=+.58$$

根據第十八表改: 則r之相等值 $=+.60$

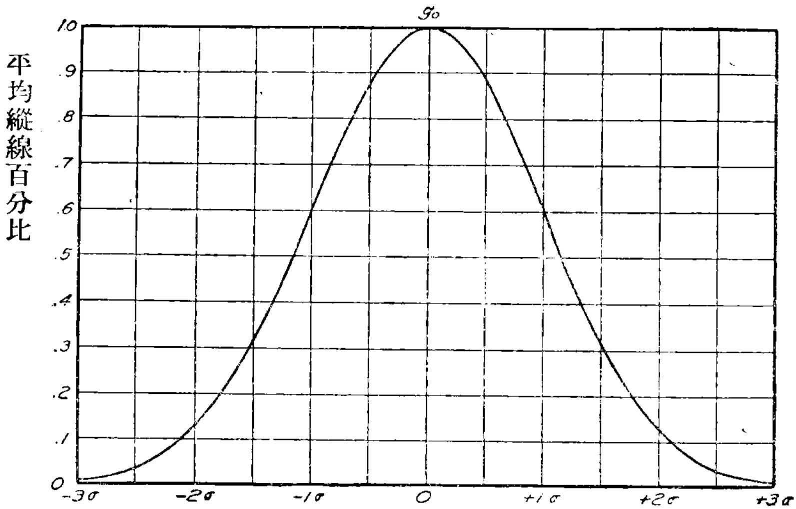
第十七表 等級相關係數之求法

ρ	r
.00	.000
.05	.052
.10	.105
.15	.157
.20	.210
.25	.261
.30	.313
.35	.364
.40	.416
.45	.467
.50	.518
.55	.568
.60	.618
.65	.668
.70	.717
.75	.765
.80	.813
.85	.861
.90	.908
.95	.954
1.00	1.000

第十八表　關而生相關係數及其相當之等級相關係數值

附 錄

機率曲線之縱線



第三十九圖　機率曲線之縱線

此爲機率曲線.平均數在式克碼量表零數上。最高縱線,y_0爲平均數.此縱線可用下式算出:

$$y_0=\frac{n}{\sigma\sqrt{2\pi}}=\frac{n}{2.5066\sigma}$$

其餘縱線爲平均縱線之分數,其分數之大小視其與平均縱線相距之遠近.

例　與平均數相距爲$+1\sigma$之處之縱線(在平均數之上或下),爲平均縱線之.606.與平均數相距爲1.34σ處之縱線,爲平均縱線之.407.此層可由**第十九表**及**第三十九圖**證明。

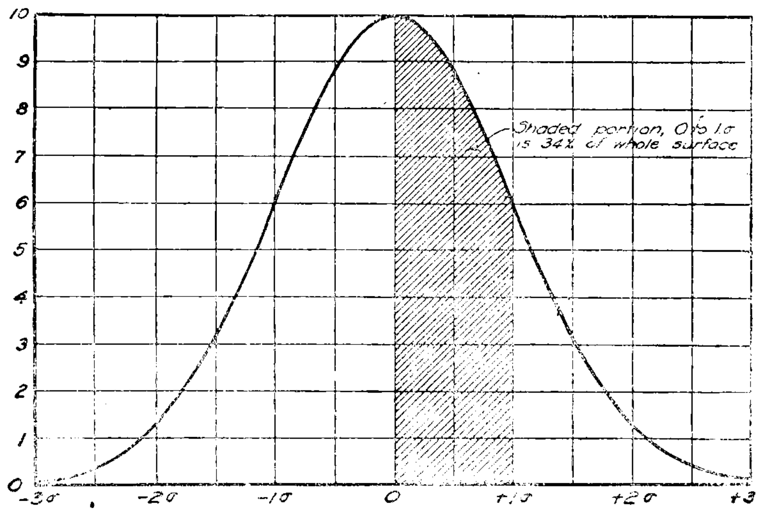
欲作機率曲線,必將x－量表譯成σ值.若分數$=65,\sigma=10$, $m=50$,則分數65在式克碼量表上爲$+1.5\sigma$。其縱線在機率曲線圖上爲y_0之.325。

Sigma	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	1.00000	.99995	.99980	.99955	.99920	.99875	.99820	.99755	.99680	.99596
0.1	.99501	.99397	.99283	.99159	.99025	.98881	.98728	.98565	.98393	.98211
0.2	.98020	.97819	.97609	.97390	.97161	.96923	.96676	.96421	.96156	.95882
0.3	.95600	.95309	.95009	.94701	.94384	.94059	.93725	.93384	.93034	.92677
0.4	.92312	.91939	.91558	.91169	.90774	.90371	.89960	.89543	.89119	.88688
0.5	.88250	.87805	.87354	.86897	.86433	.85963	.85487	.85006	.84518	.84025
0.6	.83527	.83023	.82514	.82000	.81481	.80957	.80429	.79896	.79358	.78816
0.7	.78270	.77721	.77167	.76609	.76048	.75484	.74916	.74345	.73771	.73194
0.8	.72615	.72033	.71448	.70861	.70272	.69680	.69087	.68492	.67896	.67297
0.9	.66698	.66097	.65495	.64892	.64288	.63683	.63078	.62472	.61866	.61259
1.0	.60653	.60047	.59440	.58834	.58228	.57623	.57018	.56414	.55811	.55209
1.1	.54607	.54007	.53409	.52811	.52215	.51620	.51028	.50437	.49848	.49260
1.2	.48675	.48092	.47511	.46933	.46357	.45783	.45212	.44644	.44078	.43516
1.3	.42956	.42399	.41845	.41294	.40747	.40202	.39661	.39123	.38589	.38058
1.4	.37531	.37007	.36488	.35971	.35459	.34950	.34445	.33944	.33447	.32954
1.5	.32465	.31980	.31500	.31023	.30550	.30082	.29618	.29158	.28702	.28251
1.6	.27804	.27361	.26923	.26489	.26059	.25634	.25213	.24797	.24385	.23978
1.7	.23575	.23176	.22782	.22392	.22007	.21627	.21251	.20879	.20511	.20148
1.8	.19790	.19436	.19086	.18741	.18400	.18064	.17732	.17404	.17081	.16762
1.9	.16447	.16137	.15831	.15529	.15232	.14938	.14649	.14364	.14083	.13806
2.0	.13534	.13265	.13000	.12740	.12483	.12230	.11982	.11737	.11496	.11259
2.1	.11025	.10795	.10570	.10347	.10129	.09914	.09702	.09495	.09290	.09090
2.2	.08892	.08698	.08508	.08320	.08137	.07956	.07779	.07604	.07433	.07265
2.3	.07101	.06939	.06780	.06624	.06471	.06321	.06174	.06030	.05888	.05750
2.4	.05613	.05480	.05349	.05221	.05096	.04972	.04852	.04734	.04618	.04505
2.5	.04394	.04285	.04179	.04074	.03972	.03873	.03775	.03679	.03586	.03494
2.6	.03405	.03317	.03232	.03148	.03066	.02986	.02908	.02831	.02757	.02683
2.7	.02612	.02542	.02474	.02408	.02343	.02279	.02217	.02157	.02098	.02040
2.8	.01984	.01929	.01876	.01823	.01772	.01723	.01674	.01627	.01581	.01536
2.9	.01492	.01449	.01408	.01367	.01328	.01289	.01252	.01215	.01179	.01145
3.0	.01111									
3.1	.00819									
3.2	.00598									
3.3	.00432									
3.4	.00309									
3.5	.00219									
3.6	.00153									
3.7	.00106									
3.8	.00073									
3.9	.00050									
4.0	.00034									
5.0	.000004									

第十九表　機率曲線之縱線[1]

[1] 採取于 W. F. Sheppard, 所算之表。此表見 "Tables for Statisticians and Biometricians," edited by Karl Pearson, Cambridge University Press.

機率曲線面之面積



式克碼量表　　　中數

第四十圖　機率曲線面之面積

平均數在式克碼量表零數上。在第四十圖量表上各點,均用式克碼值表示。此圖所示者爲在平均數與任何式克碼點內之面積分數。

例: 在平均數與 $+1\sigma$ 之間之面積爲 .3413。在平均數與 -1.65σ 之間之面積爲 .4505 或 45%。在 $+1\sigma$ 與 -1.65σ 之間之面積爲二者之和或 79%。

設一機率面之平均數爲 50,均方差爲 10,則在平均數 50 與 65 之間之面積爲多少。65 點處爲 $+1.5\sigma$; 按第二十表在平均數與 $+1.5\sigma$ 之間之面積爲 .4332 或約 43%。

Sigma	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.00000	.00399	.00798	.01197	.01595	.01994	.02392	.02790	.03188	.03586
0.1	.03983	.04380	.04776	.05172	.05567	.05962	.06356	.06749	.07142	.07535
0.2	.07926	.08317	.08706	.09095	.09483	.09871	.10257	.10642	.11026	.11409
0.3	.11791	.12172	.12552	.12930	.13307	.13683	.14058	.14431	.14803	.15173
0.4	.15542	.15910	.16276	.16640	.17003	.17364	.17724	.18082	.18439	.18793
0.5	.19146	.19497	.19847	.20194	.20540	.20884	.21226	.21566	.21904	.22240
0.6	.22575	.22907	.23237	.23565	.23891	.24215	.24537	.24857	.25175	.25490
0.7	.25804	.26115	.26424	.26730	.27035	.27337	.27637	.27935	.28230	.28524
0.8	.28814	.29103	.29389	.29673	.29955	.30234	.30511	.30785	.31057	.31327
0.9	.31594	.31859	.32121	.32381	.32639	.32894	.33147	.33398	.33646	.33891
1.0	.34134	.34375	.34614	.34850	.35083	.35314	.35543	.35769	.35993	.36214
1.1	.36433	.36650	.36864	.37076	.37286	.37493	.37698	.37900	.38100	.38298
1.2	.38493	.38686	.38877	.39065	.39251	.39435	.39617	.39796	.39973	.40147
1.3	.40320	.40490	.40658	.40824	.40988	.41149	.41309	.41466	.41621	.41774
1.4	.41924	.42073	.42220	.42364	.42507	.42647	.42786	.42922	.43056	.43198
1.5	.43319	.43448	.43574	.43699	.43822	.43943	.44062	.44179	.44295	.44408
1.6	.44520	.44630	.44738	.44845	.44950	.45053	.45154	.45254	.45352	.45449
1.7	.45543	.45637	.45728	.45818	.45907	.45994	.46080	.46164	.46246	.46327
1.8	.46407	.46485	.46562	.46638	.46712	.46784	.46856	.46926	.46995	.47062
1.9	.47128	.47193	.47257	.47320	.47381	.47441	.47500	.47558	.47615	.47670
2.0	.47725	.47778	.47831	.47882	.47932	.47982	.48030	.48077	.48124	.48169
2.1	.48214	.48257	.48300	.48341	.48382	.48422	.48461	.48500	.48537	.48574
2.2	.48610	.48645	.48679	.48713	.48745	.48778	.48809	.48840	.48870	.48899
2.3	.48928	.48956	.48983	.49010	.49036	.49061	.49086	.49111	.49134	.49158
2.4	.49180	.49202	.49224	.49245	.49266	.49286	.49305	.49324	.49343	.49361
2.5	.49379	.49396	.49413	.49430	.49446	.49461	.49477	.49492	.49506	.49520
2.6	.49534	.49547	.49560	.49573	.49585	.49598	.49609	.49621	.49632	.49643
2.7	.49653	.49664	.49674	.49683	.49693	.49702	.49711	.49720	.49728	.49736
2.8	.49744	.49752	.49760	.49767	.49774	.49781	.49788	.49795	.49801	.49807
2.9	.49813	.49819	.49825	.49831	.49836	.49841	.49846	.49851	.49856	.49860
3.0	.49865									
3.1	.49903									
3.2	.49931									
3.3	.49952									
3.4	.49966									
3.5	.49977									
4.0	.49997									
5.0	.4999997									

第二十表　機率曲線面之面積 ❶

❶ 採取于 W. F. Sheppard 所算之表。此表見 "Tables for Statisticians and Biometricians," edited by Karl Pearson, Cambridge University Press.

本書統計名詞中西對照表

Abscissa: 橫坐標

Additive constant: 常加數

Arithmetic mean: 算術平均數

Asymmetrical: 不對稱

Asymptote: 漸近線

Average: 平均數

Bi-modality: 雙峯形

Binomial expansion: 二項展開式

Central tendency: 集中趨勢

Chance: 機遇

Class frequency: 組距次數

Class interval: 組距

Class limits: 組限

Column diagram: 直方圖

Combinations: 組合

Correction: 校正數

Correlation by ranks: 等級相關

Correlation coefficient: 相關係數

Correlation table: 相關表

Curve: 曲線

Smoothed curve: 已修勻曲線

Frequency curve: 次數曲線

Asymmetrical curve: 不對稱曲線

Skewed curve: 偏態曲線

Normal curve: 常態曲線

Percentile curve: 百分曲線

Probability curve: 機率曲線

Curve plotting: 曲線之繪法

Data sheet: 事實張

Deviation: 差數

Quartile deviation: 二十五分差(又譯四分差)

Mean deviation: 平均差

Standard deviation: 均方差(又譯標準差)

Equation for straight line: 直線之方程式

Equation for straight line through origin: 經過原始點之直線方程式

Equivalent scale: 等值量表

Eta coefficient: 曲線相關係數

Frequency: 次數

Balanced frequency: 均衡次數

Frequency polygon: 次數多邊圖

Frequency surface: 次數面

Frequency table: 次數表

Geometric mean: 幾何平均數

Graphical division: 圖示除法

Graphical representation: 圖示

Graphical tabulation: 圖示法

Harmonic mean: 倒數平均數 (又譯調和平均數)

Histogram: 直方圖

Linear relations: 直線相關

Line-plotting: 繪線法

Locus of equation: 方程之軌跡

Median: 中數

Mode: 衆數(又譯範數)

Moment: 重距

Multiplying constant: 常倍數

Non-linear relations: 非直線相關

Ordinates: 縱線

Axis of ordinate: 縱坐標

Origin: 原始點

Arbitrary or assumed origin: 假設原始點

Pearson coefficient of correlation: 關而生相關係數

Percentile curve: 百分曲線

Percentile rank: 百分等級

Permutations: 錯列

Probability: 機率

Probability of single event: 一事之機率

Probability of compound event: 複事之機率

Probability curve: 機率曲線

Probability surface: 機率面

Probable error: 機誤(又譯概誤)

Quadrants: 四分方

Quartiles: 二十五分值

Upper quartiles: 上二十五分值

Lower quartiles: 下二十五

分値

Range: 全距離

Rank: 等級

Regression: 迴歸法

Lines of regression: 迴歸線

Equations of regression: 迴歸方程

Relationship between two variables: 二變量之相關

Sample: 樣本

Sampling: 取樣

Scatter diagram: 分布圖

Slope of line: 線之傾斜度

Standing: 地位

Symmetrical: 對稱

Transmutation of measures: 量數之蛻變

Units of measurements: 測量之單位

Variability: 離中趨勢

Continuous variables: 繼續變量

Discontinuous variable: 分立變量

Dependent variables: 倚變量

Independent variables: 自變量

X-axis: X-軸

X-intercept: X-交切

Y-axis: Y-軸

Y-intercept: Y-交切