生物統計學

上田常吉著

1935

岩波書店

QUETELET (1796–1874)

Quetelet 小傳　生物統計學ノ父, Lambert Adolphe Jacque Quetelet ハ 1796年2月22日べるぎー Gent (Gand) 市ニ生レ, 市ノ Lyzeum デ教育ヲウケ; 1814年市ノ Collège royal ニ, 次デ Brüssel 市ノ Athenäum ニ數學教師トナリ; 1828年ニハ彼ガ慫慂シテ建テサセタ天文臺ノ臺長トナツタ. 其他 1834年ヨリハ Brüssel ニ於ケル Akademie der Wissenschaften ノ常任祕書官, 1841年べるぎーノ統計中央委員長等ヲモ勤メタ.

天文學, 氣象學, 物理學, 數學等ニ關スル多數ノ研究報告ノ外ニ, 生物統計學ニ關スル次ノ著述ノ如キハ不滅ノ光ヲ放ツテ居ル.

Sur l'homme et le développement de ses facultés, ou essai de physique sociale, Paris, 1835, Bruxelles, 1869.

Sur la théorie des probabilités, Bruxelles, 1845.

Statistique internationale, Population, Bruxelles, 1865.

Anthropométrie, ou mesure des différentes facultés de l'homme, Bruxelles, 1871.

尙筆蹟ハ Baron de Ecienne Constantin Gerlache (1785—1871, 政治家デかとりつく黨ノ首領) ニ贈呈シタ彼ノ著 La physique de globe ニ署名シタモノヲ取ツタ.

QUETELET (1796–1874)

Quetelet 小傳　生物統計學ノ父, Lambert. Adolphe Jacque Quetelet ハ 1796年2月22日べるぎー Gent (Gand) 市ニ生レ, 市ノ Lyzeum デ教育ヲウケ; 1814年市ノ Collège royal ニ, 次デ Brüssel 市ノ Athenäum ニ數學教師トナリ; 1828年ニハ彼ガ慫慂シテ建テサセタ天文臺ノ臺長トナツタ. 其他1834年ヨリハ Brüssel ニ於ケル Akademie der Wissenschaften ノ常任祕書官, 1841年べるぎーノ統計中央委員長等ヲモ勤メタ.

天文學, 氣象學, 物理學, 數學等ニ關スル多數ノ研究報告ノ外ニ, 生物統計學ニ關スル次ノ著述ノ如キハ不滅ノ光ヲ放ツテ居ル.

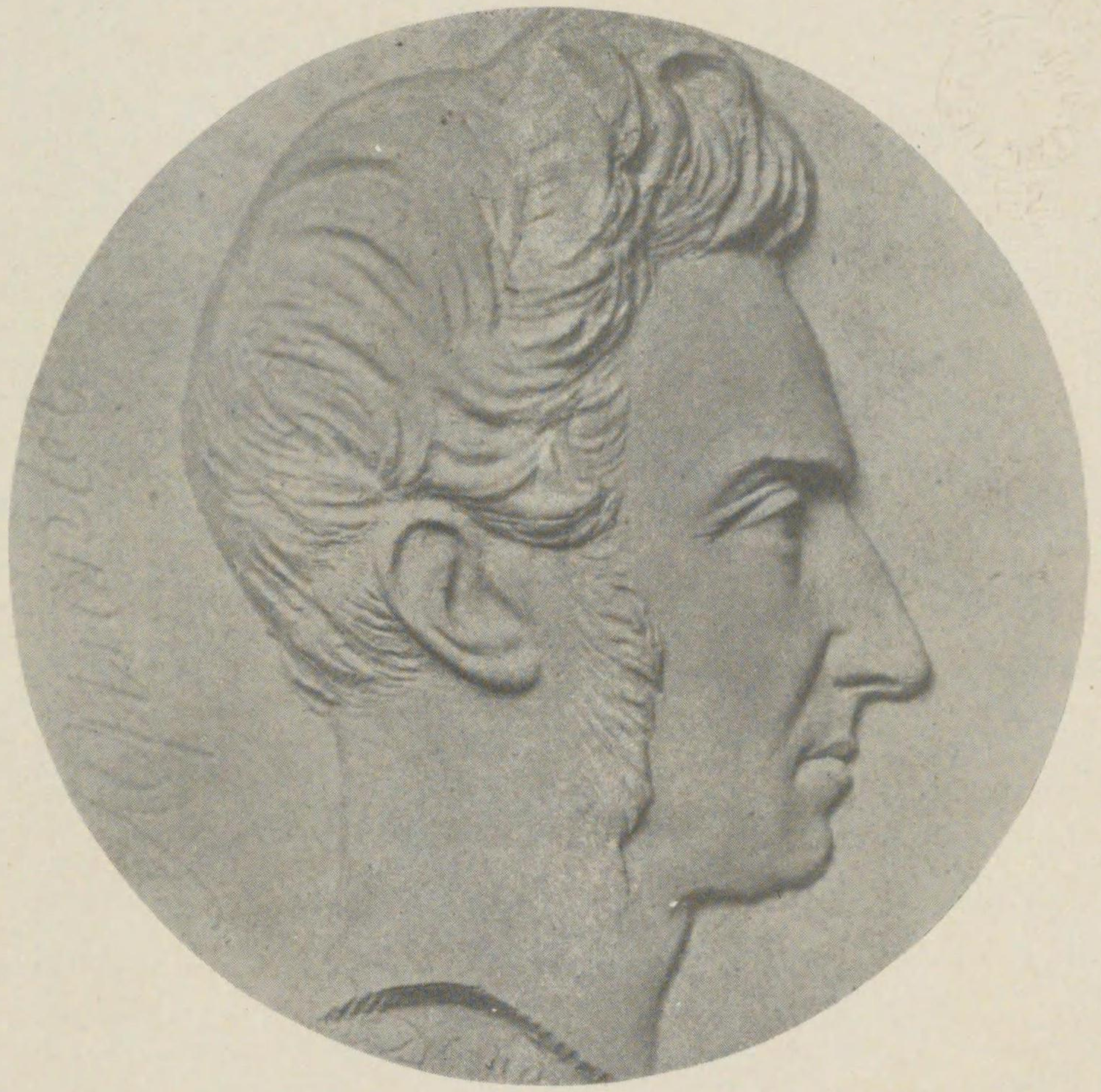
Sur l'homme et le développement de ses facultés, ou essai de physique sociale, Paris, 1835, Bruxelles, 1869.

Sur la théorie des probabilités, Bruxelles, 1845.

Statistique internationale, Population, Bruxelles, 1865.

Anthropométrie, ou mesure des différentes facultés de l'homme, Bruxelles, 1871.

尙筆蹟ハ Baron de Ecienne Constantin Gerlache (1785—1871, 政治家デかとりつく黨ノ首領) ニ贈呈シタ彼ノ著 La physique de globe ニ署名シタモノヲ取ツタ.



Quetelet

自　序

大正十一年私が朝鮮人の體質を調べた頃,生物統計學の理論を研究する人はまだ極めて少なかつた.その頃,一番困つた事は統計的理論を書いたものの無い事であつた. Yule などは良い書物だが,公式の誘導が書いて無かつたり,又書いて有つても,貧弱な確率論の知識では理解せられないことが多く,惱んだもので有つた.其後僅々十二年,學界の其の方面への認識は大分變つたと云つて良い.遺傳學,衛生學,人類學,畜產學,其他純正並に應用生物學のすべての範圍で,多少とも統計學的知識を必要としないものは無くなつた.又,邦語の數理統計學の著書も二三現はれ,英獨の教科書も數多くなつた.それでも,眞面目な學徒には,失望を與へる著書が多く,中には說明不親切で讀者を五里霧中に彷徨せしむる者すら有る.

生物統計學を現在の一般統計學から獨立させて,確率論と密接な關係を保たせることは,著者年來の希望で有つた.生物界の數量的現象は,他の人文科學に於ける統計的事項に比し遙に規則正しく,また自然哲學のそれに比し更に複雜特種なるものが有るからである.

本書は著者の此の希望へ踏み出した第一步であるが,變異統計學の中でも,記述漏れの部分がまだ相當ある.正常ならざる度數曲線及び相關の數學的研究や部分相關など,重要なる事項を大分省略したが,これ等は計算法極めて複雜なので,これを單

純化して後記述したいと思つたが爲,今回は省略したのである。改版ごとに順次增補したいと思ふ. 本書屢ば續篇といふ語を用ひた. 最初續篇をものする考で居たが,版を更めるごとに增補した方がいい樣である.

卷頭ケトレーの寫眞とその署名とは,城大教授中村拓博士の所藏にかかるもので有る. 請うて本書を飾ることを得たのは喜びに堪へない. 厚く御禮を申上げる. 又本書成るに當り激勵と援助とを賜つた先輩並に同僚諸賢に謹んで御禮を申上げ,併せて本書の不備不完全なる點,又犯してゐるだらうと危まるる魯魚の誤り等については,此上とも叱正の勞を惜しまれざらんことを御願ひする次第である.

昭和九年八月

著　者　識

目　次

附 錄

附　表

序　論

1. 生物統計學 (Biostatistics; Die Biostatistik)

生物統計學は生物學の一分科であつて生物を數量的に攻究する方法とその數學的根據を論じ,且つ數量的現象の法則を研究する部門を云ふのである.

自然科學發達の經路は,現象の記述及び考察に始り,次いで定量的研究に由つて自然現象を正確に認識し,終りに法則の樹立となるのが自然である. 生物學に於ては,定量的研究が他の自然科學,例へば物理學,化學,天文學等に比して,遙に遲れたのは事實である. 過去に於て生物統計學の不振であつた最大の理由は恐らく,生物は自然界に於ける最も複雜な且つ最も神祕的な存在であること,從てその定量的關係も他の自然界のそれに比し遙に複雜であつて,一見法則なきが如くに見ゆること等であらう. Quetelet や Galton 以來,この分科の進歩は遲々であつたが,最近遺傳學,民族衞生學,體質人類學等の勃興と共に,生物の定量的研究も進歩の目ざましいものがある.

然るに,生物を定量的に研究することは,今日尙ほ**生物測定學** (Biometry; Die Biometrie) と稱し,生物學研究の一の方法を教ふるものとして,即ち單なる**方法論** (Methodology; Die Methodologie) としての存在しか認められて居ない狀態にある. これも,或る程度迄は甘受しなければならなかつたかも知れぬ. そのわけは

生物統計學は遺傳學,衛生學,體質人類學等の補助學科としてこれまで發達して來たからで; 斯界に貢獻の有つた學者の中で眞に生物の量的法則樹立を目差して居たのは, Quetelet 一人位かも知れない.

しかし,この方面の著しく進歩せる今日,生物測定學なる方法論的の名稱は內容を忠實に表はすものでない.

本書冠するに**生物統計學**なる名を以てしたのは,生物測定學なる名稱が不適當である上に,人文科學に於ける**統計學**と截然と區別する必要を感じたからである. 即ち,生物界に於ける數量的現象は,人文科學界に於けるそれとは趣をことにするものがあつて,同時に論ずることは無益有害であると信ずるが爲である.

生物に關する數及び量は**計數**(肉眼又は顯徵鏡等にて觀察し存在する或るものの數を數へること; Die Zählung)と,**計量**(尺度,秤,寒暖計その他の器械を以て大さ,重量又は程度等を計ること; Die Messung)とによりて得られる. 計量と計數を總稱して**計測**と名づける. 計測の外に**觀察**も,亦,數量を以て表はしうる事が有る. 或る性情の有るか,無きかを觀察し,その性情を有するもの α%,有せざるもの β% 等と記載する如き其の一例で有る. 觀察と計測とを總稱して**觀測**と名づける. 觀測によりて得たる數の取扱方は,目的によつて異ふのであるから,廣義の生物統計學に次の三部門を大別することが出來る.

I. 函數生物學,

II. 變異統計學,

III. 誤差論.

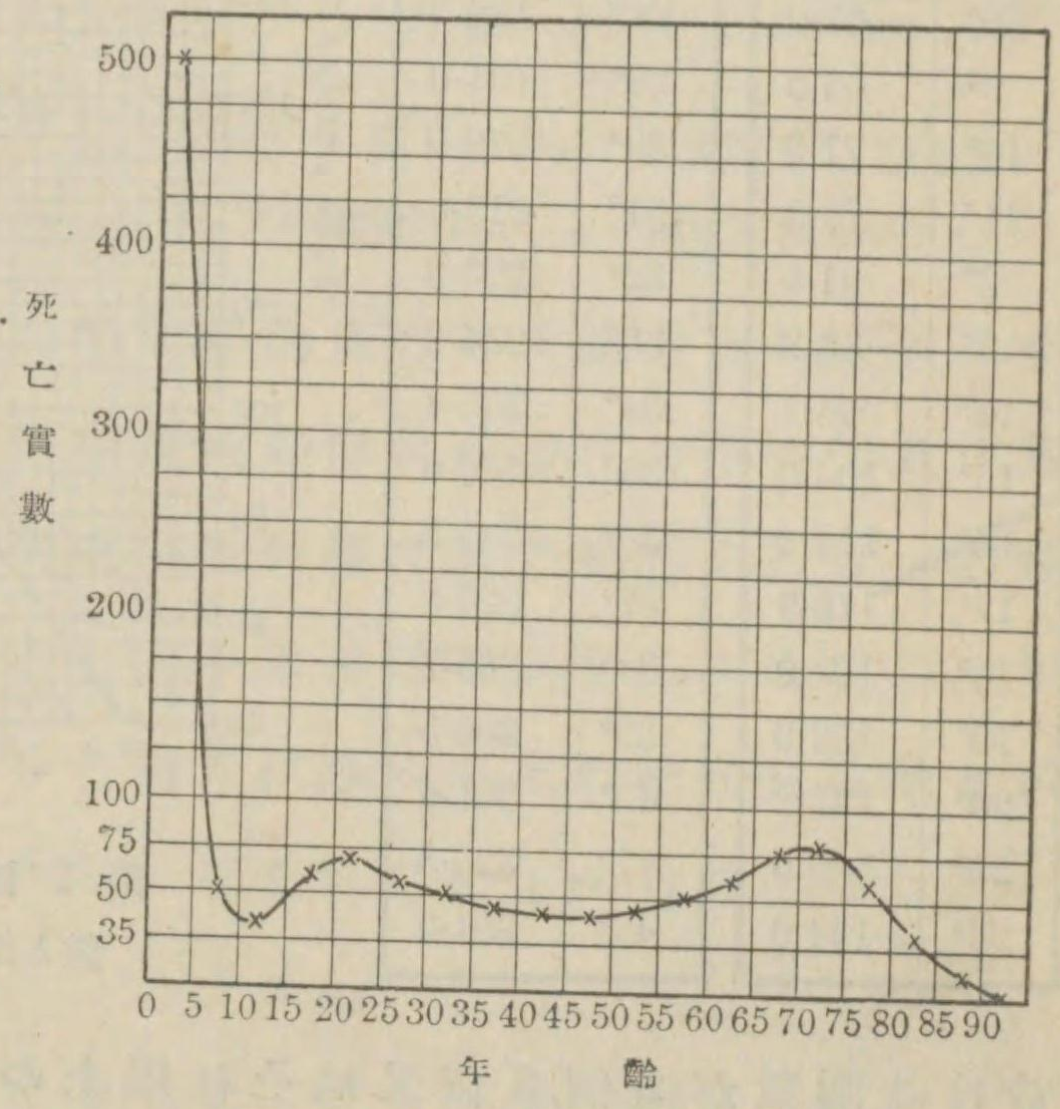
2. 函數生物學

出現する數又は量が時間,溫度,電氣や光の强さ,藥品の量等に從つて變化する; 換言すれば,數又は量が時,物理的又は化學的條件等の函數である場合がある. この函數關係を研究の對象とするものを**函數生物學**と云ふ. 二三の例をあげて見る.

(1) 或る特定の人の身長は年齡の增すに從ひ增大し,一定の年齡以後では固定するが,老年に至る時は却て減少するのが普通である. 即ち人の身長は年齡の函數である.

又他の例を引けば,1 歲の人の死亡率と,10 歲の人或は 25 歲の人の死亡率とは皆異つて居る. 年齡が變ればその死亡率(又は死亡實數)がこれに從つて變り,死亡率(又は死亡實數)は年齡の函數であると云ふことが出來る.

第 1 圖は大正 9 年日本に於ける死亡實數を年齡の函數として圖示したものである(4).



第 1 圖 大正 9 年日本に於ける死亡者數の年齡別分布圖.
茲に死亡實數は 1000 人を單位とした.

(2) 脈搏數と溫度との關係を調べる. 他の條件を悉く一定にし,溫度のみを變化する時は,脈搏數は溫度の函數で

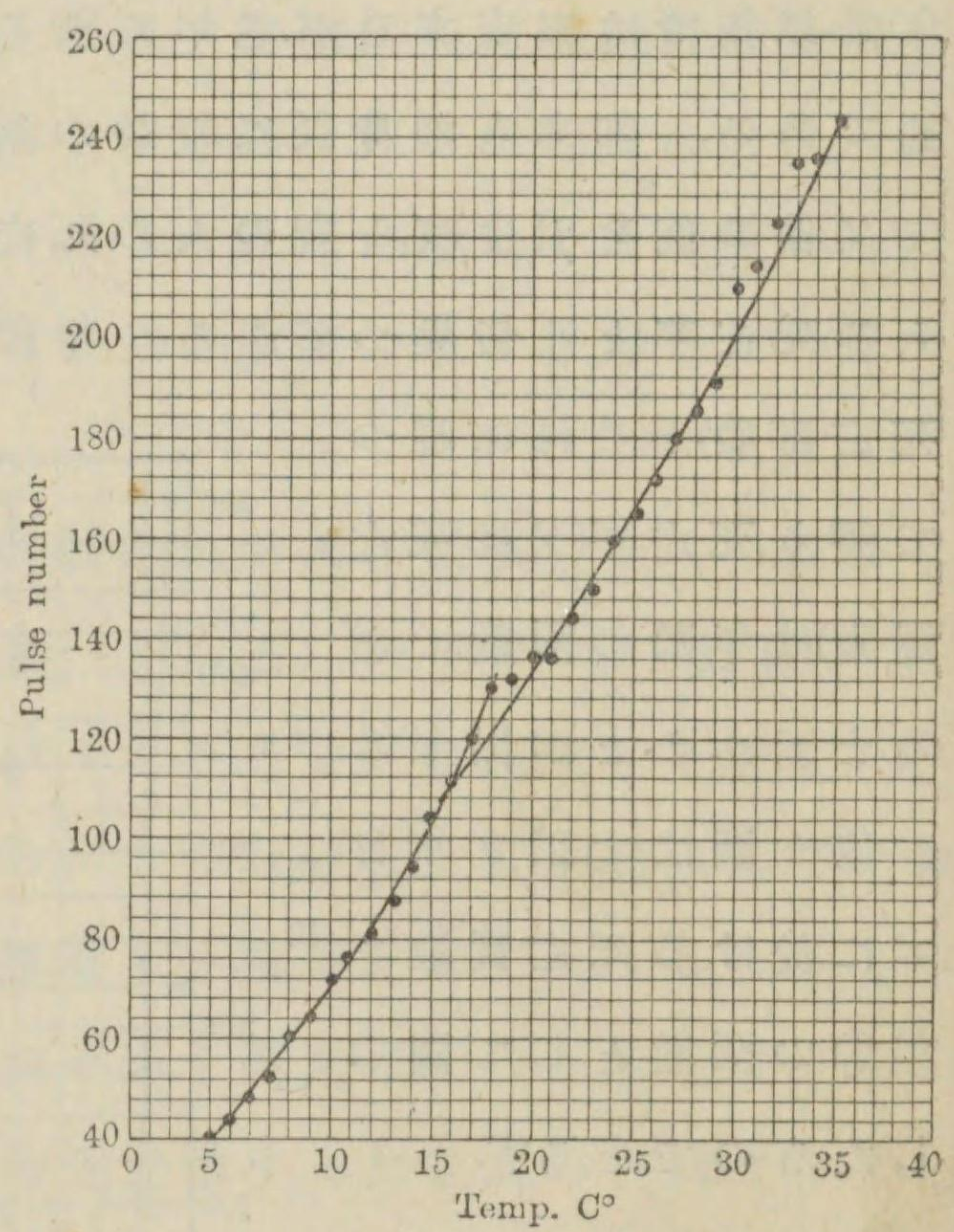
ある. 今, Chironomus(ユスリ蚊)の幼蟲の1分間に於ける心臟搏動數を溫度の函數として記錄したのが,第1表及び第2圖である[(2)]. 攝氏4°より42°まで溫度が昇るに從ひ,1分間の平均搏動數が漸次多くなつて居る. 卽ち,脈搏數は溫度の函數である.

第 1 表

ユスリ蚊の幼蟲の1分間に於ける脈搏數.

溫度(攝氏)	1分間の脈搏平均數	溫度(攝氏)	1分間の脈搏平均數
3°	停　止	23°	150·3
4°	41·0	24°	159·0
5°	44·5	25°	165·2
6°	48·7	26°	173·2
7°	54·1	27°	179·5
8°	61·0	28°	186·3
9°	65·5	29°	191·0
10°	71·3	30°	210·0
11°	75·8	31°	213·6
12°	81·9	32°	223·3
13°	88·3	33°	234·2
14°	96·7	34°	235·4
15°	104·0	35°	246·6
16°	112·2	36°	287·3
17°	119·9	37°	326·8
18°	128·9	38°	365·7
19°	132·9	39°	366·5
20°	137·2	40°	361·6
21°	139·0	41°	323·8
22°	144·6	42°	251·5

脈搏數は又年齡の函數とも云へる. 又同時に筋肉勞働の强さ等の函數とも云へるであらう. 數學に於ける函數が2個,3個又はそれ以上の變數の函數であり得ると同じく,1分間に於ける脈搏數は溫度,年齡等種々なる要約の函數である.



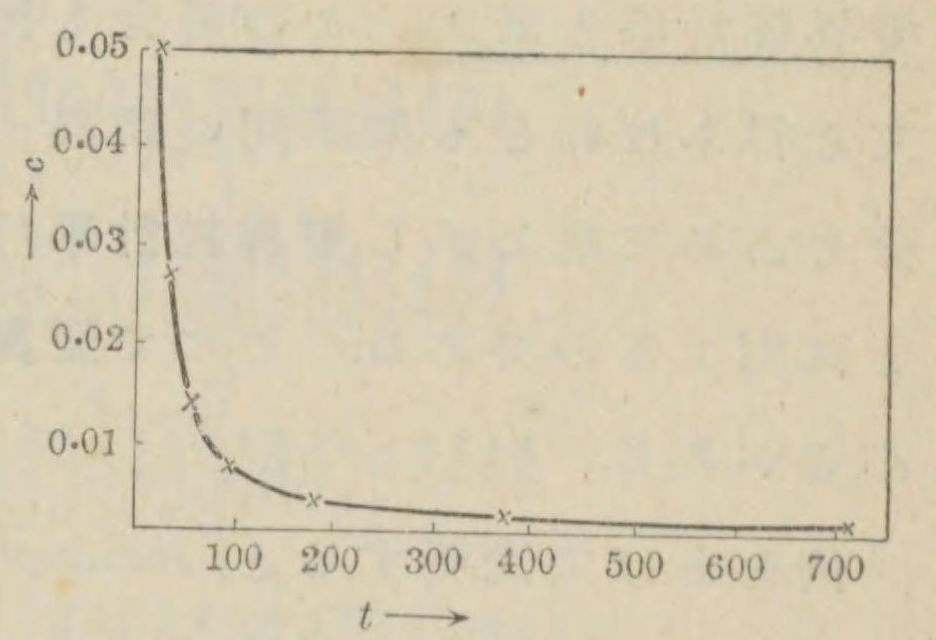
第 2 圖　ユスリ蚊の心臟搏動數と溫度との函數關係を示す.

(3) Koffein の溶液で Spirogyra 細胞の原形質の沈澱を起さしめた. Koffein の溶液の濃度 c と,沈澱の時間 t との關係は, K を常數とすれば,

$$t=\frac{K}{c}$$

で表示することが出來る. t を橫軸に c を縱軸にとれば,第3圖の如く雙曲線となるのである. t と c とは明に一定の函數關係を示す[(1)].

一體,多くの數を取扱ふ以上,常に何等かの函數關係を明にするのが目的の一であるけれども,函數生物學的取扱をなしうるのは,以上三個の例でも明な通り,數又は量が時間または化學的,物理學的,或は形態學的條件等の函數として表示せられ,その函數關係——一の變數の變化につれて他の變數が如何に變化するか——を研究の主目標に置く場合である. **函數生物學**なる名稱は,我國で八木,小泉兩氏[(1)]が初めて使用したものであつて,未だ一般的の名稱では無いが,次の變異統計學と區別するに適切な言葉であるから,これを使用することとした.



第 3 圖　Koffein の濃度 c と原形質の沈澱時間 t との函數關係を示す.

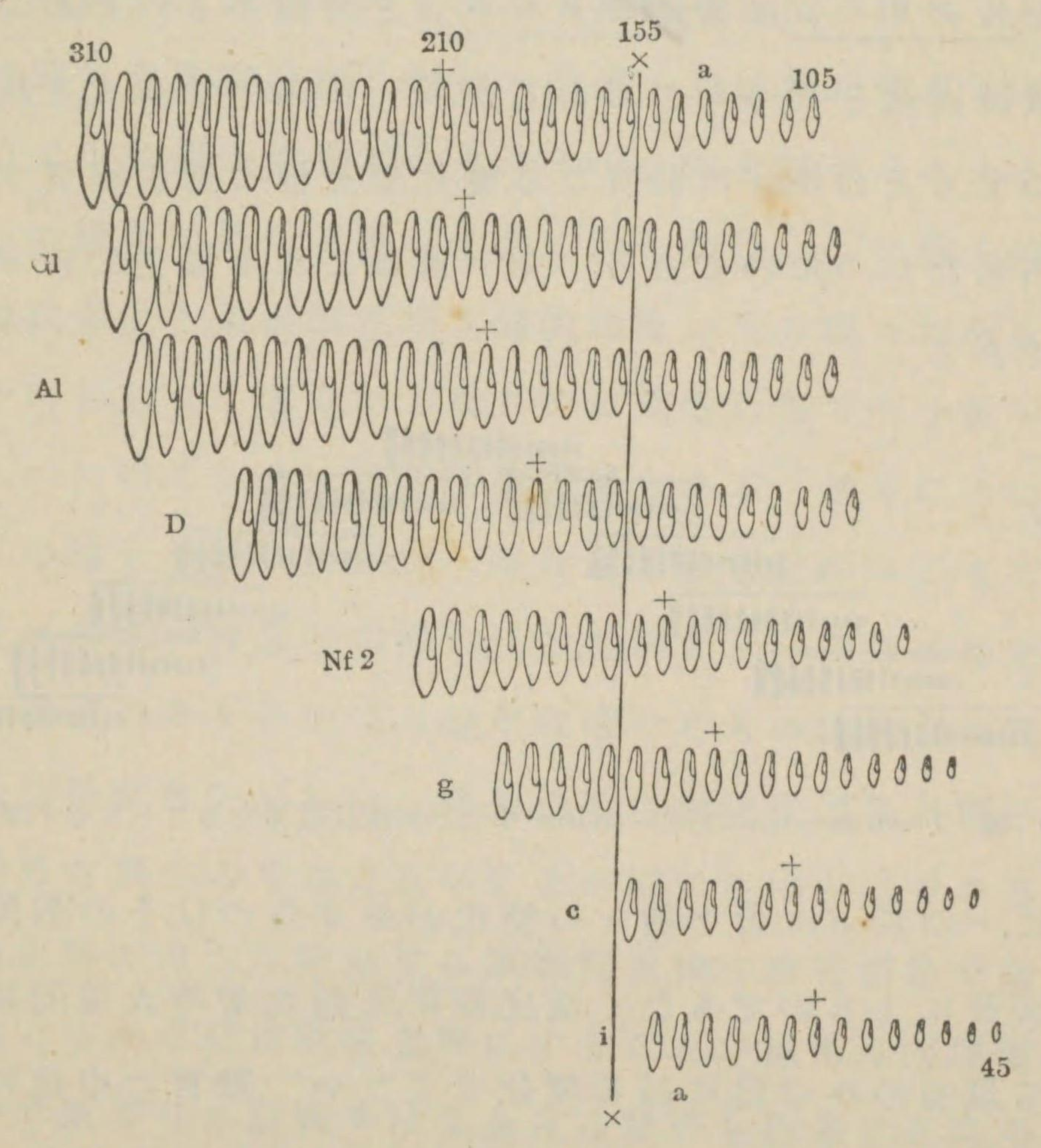
3. 變異統計學

函數生物學では條件を漸次變化させてその場合に,生物に所屬の數又は量が如何に變化するかと云ふことが研究の主題であつた. 此度は條件を一定にするのである. あらゆる條件を

同一にしても,——人力の及ぶ範圍であらゆる條件を同一にしても,——その生物に屬する數又は量が個體毎に相違が有るものである. この個體間の差異は全く不規則で何等共通の法則が無い樣であるが,多數の計測を繰返して行ふ時は,一定の法則のあることが明になる. この個體間の變化を研究の對象とし,更に進んで異なる個體群の比較研究に應用せんとするものを變異統計學と云ふ. この場合も後章で說く樣に函數關係の研究を行ふけれども,個體間の差異—變異—の研究が主要なる部分を占めて居るから,**變異統計學** (Die Variationsstatistik) として他と區別するのである. ここで**變異**と云ふ語の意義を明にする必要がある.

例をあげて說明すると, Jennings[3] が水溜りから1個のザウリムシ (*Paramaecium caudatum*) を取つてこれを一の水槽に飼養した所,この1個の *Paramaecium* から分裂によつて多數の個體が出來たが,飼養條件(溫度,食餌,水の量等)が何れも同一であると思考せらるるにかかはらず,大小種々の變化が出來たのである. かく1個の生物より無性生殖によつて出來た一の個體群を Klon と名づける. 次に同一の池より再び他の *Paramaecium* を取つて來て別の水槽に飼養し,各水槽の飼養條件を同一にしたのであるが,その各水槽からは何れも大小種々の大さのザウリムシが出來た. 又その變化は水槽ごとに必ずしも同一でなかつた. これは池から取つて來たザウリムシに素質の異つたものがあるためであらうと, Jennings はこれより體長を異にする8個の Klon を選別した. これが第4圖に示した八つの群であつて,數

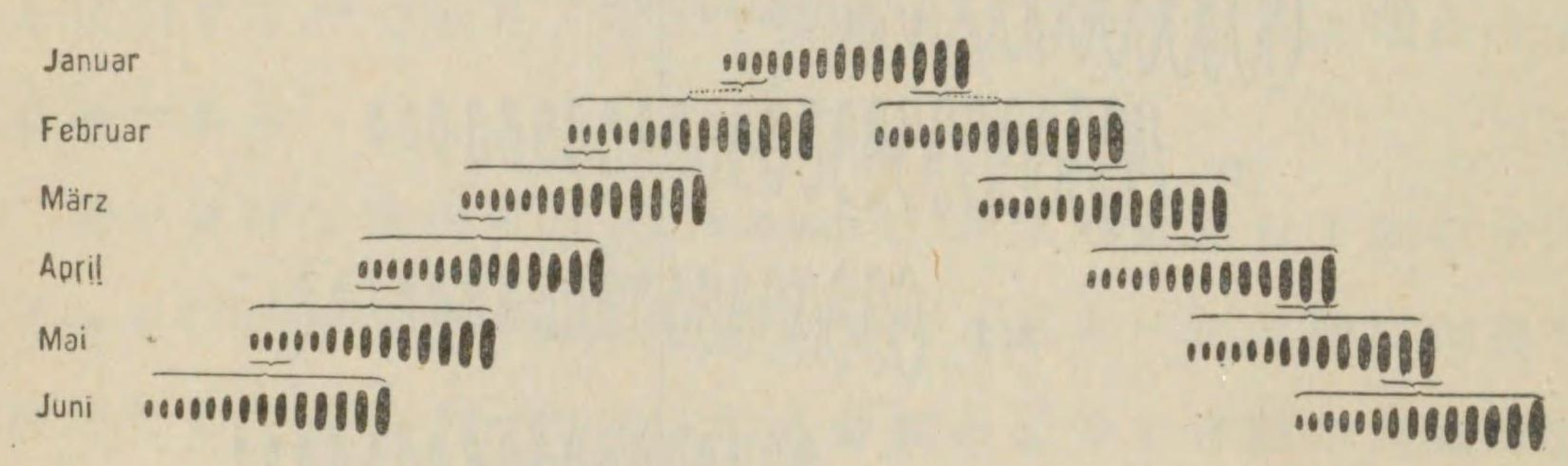
字はミクロン(μと記し,$\frac{1}{1,000}$ mm.)單位で計測した體長である. ×——× の線は8群合計の平均の位置のある所,又+は各 Klon の平均のある所である. 8個の群は何れも最大及び最小を異にし,その長さの平均もちがつて居るのである.



第4圖 *Paramaecium caudatum* の八つの Klon の概觀圖.

次に,或る一つの Klon の中最小なるものを取つて,これを別個の水槽に飼育して分裂を起さしめた所,必ずしも小なるもののみを生ぜず,再びもとの Klon と同一の變化を見た. 又原 Klon より最大なる個體を取つて同一條件のもとに飼養して見た所,

これも大なる個體のみを生ぜず,最初のKlonと同一の變化を見たのである. 一Klonの中でたまたま大なる又は小なる個體を取るも,これより發生する次代のKlonの個體は大又は小なりと限らない. 即ち母體その物の大さは娘體の大さを左右しない. 出來た新クローンは常に母クローンと同樣のものであつた. かく數回淘汰を行つたのであるが,同一Klon中で淘汰を繰返して行つても,そのKlonに特有なる變化型を他の型に變更することが出來なかつたのである. 第5圖はこれを標型的に示したものである.



第5圖 *Paramaecium* の同一 Klon 中での淘汰は效果なきことを示す標型圖.

故に,一のKlonに大小種々の變化が生ずるのは,その個體の遺傳的素質によるのでもなく,溫度,飼料,光線,水量等人爲的に變化しうる環境のみの爲に起る變化でもない. 發育の中途に,それぞれの個體に或は抑制的に或は促進的に作用する無數の原因が組合せられ,且つこの原因は或る個體には偶然に多く抑制的に,他の個體には偶然に多く促進的に作用するが爲に起る變化であると考へらる. かくの如き變化を**彷徨變異** (Die Fluktuation) と云ふ.

Paramaecium では大さのみならず,毒素に對する抵抗力や,高溫度に對する抵抗力等,生理的機能にも彷徨變異のあることが明にせられて居る.

彷徨變異は無性生殖を行ふ單細胞生物に限らず,自花授粉を行ふ生物,例へば豌豆の如き物にも觀察せられて居る. Johannsen[5]は豌豆を數代栽培し,初め一品種と思はれた豌豆より,淘汰によつて十數個の純系 (Die reine Rasse) を分離することが出來た. 一の純系に屬する豌豆を,全く同一條件にして蒔く時, (肥料,土地,日光,灌水,氣溫,濕度等人爲的になしうる限り,如何に條件を同一にしても), 實りたる豌豆には重量の點で大小種々の變化があつたのである. これ**彷徨變異**である. 然るに,この純系の何れの種子を蒔いても,例へば大なる種子を蒔いても,收穫する種子は最初の純系と同樣の彷徨變異を示し,反對に母系の中小なる種子のみを蒔いても,翌年收穫せるものは最初の純系と同樣の彷徨變異を示した. 純系內での淘汰は翌年の種子の大さに變化を與へるものでないことが明になつたのである.

無性生殖に由つて增加する單細胞生物や,自花授粉で繁殖する植物では,かく彷徨變異を明にすることが出來るが,雌雄異本の植物や,高等なる動物では完全なる純系を作ることは,不可能ではないまでも甚だ困難である. 從て,彷徨變異を分離することは甚しく困難であると云はねばならぬ. 多くの場合に變異は極めて複雜であつて,遺傳的に異なる,又環境的にも異なる多數の彷徨變異の集合して一の**變異**となつて出現して居るのである.

さて變異統計學は彷徨變異に於ける統計を行ひ,その變異の統計的方法,他の變異と比較する方法を論じ,又變異の法則を研究するものである. しかし,彷徨變異の統計が甚だ困難である場合が普通であつて,その時は,多數の彷徨變異の集合をあたかも一の彷徨變異であるが如く取扱ひ,これを研究するのである.

人類の場合では先天的條件を分離し得る場合極めて稀であり,たとひ爲し得ても極めて不完全であるから,常に彷徨變異の集合を一の彷徨變異なるが如く取扱つて居る. それのみならず後天的條件,例へば職業,地位,年齢等の如き條件すら分離して研究することが困難な程である.

變異統計法に於ても亦函數關係を研究する. 例へば,日本人6,000人の頭最大長徑を計測し,これを 2 mm. ごとに區分しならべる時(度數分布表)は,第5表(23頁)となるが,この表を讀めば明なる如く,身長の小なるものの個體數少く,身長の大なるものも亦少數で,身長中等のもの最も多數を占めて居る. 身長がきまればこの身長を有するものの人數が明になる. この人數を**度數**なる語であらはして居るが,度數は身長の函數であると云ふことが出來る. この函數の研究は變異統計學でも主要なる題目であつて,第三編で明にならう. しかし,同じく函數であつても,變異統計學及び誤差論に於けるこの函數の特徴は,函數が常に度數又は確率であることで,この點函數生物學と趣を異にする. かく說明すれば函數生物學的の事項と,變異統計學的事項とに,明確な區別が有る樣であるが,時には何れにも編入しうるものがある. 例へば第一章第6表に於けるディフテリア患者

の死亡實數の度數分布は變異統計學的に取扱つて居るが,死亡者數の如く年齢の函數と考へ,函數生物學的にも取扱ふことが出來る. すべて年齢を變數とする度數分布はどちらの方法でも研究し得る.

4. 誤差論

今,富士山の海水面上よりの高さを測量したとする. 10回測量すれば結果は10回とも多少の相違があり,又眞の高さに比べても亦必ず差がある. この差を**誤差**(詳しくは**觀測誤差**)と名づける. 觀測誤差は物理學的,天文學的,化學的,工學的計測等に常に出現することであるが,生物學にも亦屢あらはれる. 例へば頭蓋腔の容積を計測するに,常に著しい誤差を伴ふものである. これを矯正する爲に,監査頭骨(Der Kontrollschädel; 頭骨の大後頭孔を除き他の大孔に黃蠟をつめ,頭蓋腔に多量のシェルラックを流し込み充分頭蓋腔の內壁に流して,小孔を埋めて後,乾燥する. かくの如き操作を數回繰返して小孔を塞ぎ,乾燥せる後その容積は水を容れて正確に計られて居る)を用ひる.

監査頭骨の容積1250 c.c. なるものに,粟粒を充分に塡めて後,この粟粒の容積を計るに,結果は一回ごとにちがつて居り,第2表の如くであつた.

第 2 表

監査頭骨を充たしたる粟粒の容積

1260 c.c.	なりしこと	1回	誤差 = +10
1270 c.c.	なりしこと	2回	+20
1280 c.c.	なりしこと	7回	+30
1290 c.c.	なりしこと	6回	+40
1300 c.c.	なりしこと	4回	+50
計		20回	

卽ち,その結果は全部は同一でなく,誤差10なること1回,誤差20なること2回,30なること7回,‥‥であつた.

かく誤差の分布が變異の分布によく似て居る.しかしこれは變異ではない.變異は個體間の差違であつて,多數の個體を計測して初めて,その間の差違を認めることが出來るものである.しかるに誤差のみが起る時の被計測體は1個である.今1個の監杳頭骨を20回計測したのであるから,變化ある値が出て來ても決して變異ではない.

自然科學界には誤差の全然ともなはない計測は有り得ない.唯誤差が上記の如く大なる際に問題になるが,小なる場合は無視していいのである.かくの如き誤差を研究の對象としたのが誤差論である.

誤差論の歴史は遠く Gauss, Legendre, Laplace 等に遡ることが出來る.變異統計法はこれに比べると新しく,しかも前者より分派し,發展を遂げたものである.

然らば,觀測誤差と變異との關係は如何であらうか? 上の頭蓋腔の例の樣に,計測方法の不完全なるもの,又は困難なものでは,誤差が可なり大きいのであるが,生物學では斯くの如きことは寧ろ例外であつて,普通は誤差は變異に比すると,極めて小である.故に生物學では誤差は普通考へない習慣になつて居る.しかし單に身長や頭の長徑を計る場合では誤差が小であるが,生物體の化學的分析や,心理學的計測等に至りては,變異の外に誤差も亦必ずしも小ではあるまいと思はれるから,將來は誤差も亦問題となるであらう.

そこで,函數生物學,變異統計學,誤差論,三つの研究對象を比較して見るに,函數生物學は條件にともなふ數又は量の函數的變化が研究の主目標であつて,條件の數量を順次かへれば,他の數量も亦函數的に變化する.云はば數又は量の**動態**が研究對象になるのである.第二の變異統計學ではすべての條件を出來るだけ一定にし,その條件内に於ける個體間の變異を研究するので,云はば研究對象の數又は量は**靜態**なのである.然るに,誤差論に至れば更に一層限定的であつて,すべての條件を一定にするのみならず,計測の個體も1個と限り,唯これを度々計測して,その誤差を研究するにあるので,研究の對象は全く**固定的**となる.

生物統計學的研究にはこの三階段を考慮せねばならぬ.精密なる計測には必ず,第一に誤差を考慮しなければならぬ.第二に又,多數の個體を計測すれば,誤差と個體差(即ち彷徨變異)とが混合して出現するものである.第三に,條件を漸次變化してゆけば,その條件の函數として,數量が變化するものである.物理學,天文學等に於けるが如く,變異なき自然現象では誤差を考慮して,數又は量を函數として取扱へばいいが,生物學ではこれと異つて誤差は普通考慮する必要ないが,變異に關する基礎知識が無ければ,數又は量は殆ど論ずることが出來ない.唯特殊なる題目,即ち誤差の大なるものでは,同時に誤差をも考へねばならぬ.

廣義の生物統計學は以上三つの領域を有することを說いた.狹義に**生物統計學** (Die Biostatistik) とは變異統計學 (Die Variationsstatistik) を意味するが,また誤差論 (Die Fehlerlehre) をも含ませることは必要であらう.函數生物學はこれに反して取扱法

に大分相違があり,對象も相違があるから,別個に取扱はれて居るのが普通であり,又それが便利である. 本書では狹義の生物統計學,卽ち函數生物學を除いた意味で數量的現象を論ずることとする.

5. 生物統計學と生物學

今日生物統計學と最も密接な關係のある學科は,遺傳學,衞生學,解剖學,體質人類學,心理學等であるが,その他動物並に植物分類學,畜産學,昆蟲學,寄生蟲學,細菌學等にも漸次應用せられつつあり,將來は生化學,病理學等,その應用生物學たると純正生物學たるとを問はず,廣く,更に一般的に利用せらるるものたるを著者は信じて居る.

かく廣汎に涉り利用せらるるから,生物統計學は統計の方法を講ずる學,卽ち方法論 (Methodology; Die Methodologie) に陷り易い. しかし生物統計學の主目的は生物界の數量的現象及びその法則を研究するのにある. しかも,その現象は確率と深い關係があり,確率論の定理は驚く程正確に生物學的現象に適合するのであつて,後者は生物學の爲に存在するかの如き感をさへ懷かせるものが有る.

本書では三編に分ち,第一編及び第二編では主として統計の方法と其の理論を論じ,第三編は確率論より生物統計學に入り,その深い關係を明にすることとしたのである.

参考文獻

此處に引用したものは,

1. 八木誠政, 小泉清明(昭和4年):- 函數生物學.

2. 小泉清明(昭和3年):- Chironumus 幼蟲の心臟搏動數とその溫度係數,特に Temperature characteristic に就いて. 動物學雜誌,第40卷.

3. Jennings, H. S. (1908/9):- *Paramaecium* についての實驗は, Handbuch der Vererbungswissenschaft. Bd. I, Lief. 11:- Dauermodifikationen. v. J. Hämmerling (1929) より.

4. 小倉金之助(附錄を見よ).

5. Johannsen (附錄を見よ).

第一編
一變數の統計法

第一章
度數分布

1. 屬性統計及び變數統計

本書で取扱ふのは**變異**と**誤差**とに關する事項であるが,誤差は生物統計學では變異の如く顯著でないから,續編に於て述べることとし,本書では誤差が無いものとして,變異のみにつき論ずる. **變異統計法**は生物の性情の變異を取扱ふのであるが,これに二つの場合がある. 例へば生れる子供が女であるか或は男であるかと云ふ如く,或は一群の人の眼を調べてその盲者であるか明視者であるかと云ふ如く,或は咲いた花が白か有色かと云ふ如く,或る一の性情を有するか有しないかを檢査する場合がある. かくの如き場合に適當なる統計法を**屬性統計法**と云ふ. この場合は,多數の個體を檢査して初めて數を以て表示しうるのである.

此度はそれと異つて或る植物の花瓣の數とか,豌豆の一莢中の種子の數とか,動物の一腹に生れる子供の數とか,或は人の身長とか,豌豆の種子の重さとか云ふ風に,何かの數を計へたり,又は大さ或は程度等を計測したりする場合がある. 斯く計數又

は計量によつて得たる數の整理を行ふ方法を**變數統計法**と云ふ. 計數又は計量による場合は,唯一個の觀測でも既に數として取扱はねばならぬのみならず,屬性統計法に比し複雜であるから,主として變數統計法につき記述し,屬性統計法は確率論の知識を要するから第十章で說くこととする. 又第一編では變數統計法のうち一變數を取扱ふこととし,第二編で二個の變數の場合を論ずる. 三個又は以上の變數の場合は一層複雜であり,且つ抽象的であるのみならず,實際に利用せられることも未だ稀であるから,續編に讓る. 今,この區分を表とすれば次の通りとなる.

變異統計法
- (1) 屬性統計法
- (2) 變數統計法
  - (a) 一變數を取扱ふ場合
  - (b) 二變數を取扱ふ場合
  - (c) 三變數以上を取扱ふ場合

2. 原表又は記錄の整理

計測項目が少なければ問題でないが,項目が多い時は,原表又は記錄は一個體につき一枚づつ觀測表を作つて置くのが便利である. 記錄は精細に全體に目を通して,人類ならば本籍(又は生地),職業,年齡,疾患等その目的に應じて分類し,或は動植物ならば遺傳の條件,飼養法や培養の條件等によつて分類し,それぞれ別個に取扱ふ. 成るべく條件を單純に,卽ち材料を純粹にするのである.

生物統計學ではその材料が純粋なる程,その權威に永續性があり,價値が大である. 例へば,單に日本人の身長とせずして,大阪府下土着人の年齡24歳より45歳までの男子の身長と云ふ風にする. 日本人でも地方的差異が甚しく,若し全日本より集つた男子の身長平均を求めたとすれば,各地方より集合する割合によつて平均身長が或は大に或は小になる虞がある. 故に,若し東京府下と京都府下との農民を檢査したとすれば,初めより全然別個に取扱ふべきである. 初めより別個に取扱つて來た材料を,最後に至り合一することは,極めて容易であるが(第二章及び第三章參照)初めから混同して取扱つて來たものを後に至り分離することは不可能である. 單一純粋なることは研究材料の具備すべき要件である. しかし,人體の計測の場合の如き,條件を餘りに嚴にして多種多様に類別したが爲に,各類の觀測數が少數となる虞が有る時は,影響の少い條件は大目に見ることが多い. 例へば,身長を計測するに滿24歳以上50歳以下では變化が少いから,24歳以上を一に類別し23歳以下を一歳ごとに取扱ふ. かく,研究の目的に應じて寛嚴宜しきを得なければならぬこと勿論である.

尙ほ,記錄の整理に當り,偶然混つた異分子を見別けること,計測の誤りまたは記載の誤謬の發見に力むることを忘れてはならぬ. それがすめば,指數の計算或はその他必要なる換算を行ふ.

3. 度數分布表

變異統計では常に多數の觀測を行ふのであるから(大量觀察),

數百又は數千の觀測値を無秩序にならべた記錄よりは,何等の意義をも捕捉することが出來ない. 必ず,或る方法で分類し,一目瞭然たらしめねばならぬ. 度數分布表の作製はその第一歩である.

例 1. 油菊(*Chrysanthemum lavandulaefolium*)の花は頭狀花であるが,その花の周邊に舌狀花と稱するのがある. 今,油菊の頭狀花499につきてその舌狀花數を計へて,第3表を得た.

第 3 表

油菊の舌狀花數の分布, $M=21{\cdot}88$. (池野)

舌狀花數	16	17	18	19	20	21	22	23	24	25	26	27
度　數	1	3	11	33	66	100	104	77	62	33	7	2

この表の第一行は舌狀花の數であつて,最小16より始つて最大27に至る. 即ち,調査せられた499個の花では舌狀花數は16—27であることがわかる. 又,第二行は花の數である. 總計499個のうち舌狀花16なるもの1個,舌狀花17なるもの3個,舌狀花18なるもの11個,‥‥‥なることを示すのである.

第一行の數を**變數**(又は基礎變數; Variable; Das Argument)と稱し,統計の基礎となる數列である. 第二行に示すものを**度數**(Frequency; Die Frequenz)又は**級度數**(Class frequency; Die Klassenhäufigkeit)と云ひ,各變數の頻度を示すのである. この表を**度數分布表**,又は單に**度數表**或は**分布表**(Frequency-table; Die Frequenztabelle od. Variationstabelle)と云ふ.

例 2. 第4表甲はQueteletが約26,000の兵士につきて身長を

計測し,吋以下を四捨五入し度數分布表としたものである.

第4表 甲

26,000人の兵士身長の千分比分布表(I). (Quetelet)

身表	X-60吋	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75-X吋
度數	2	2	20	48	75	117	134	157	140	121	80	57	26	13	5	3

第一行の變數はここでは身長である. 身長は1吋を單位とし,それ以下は四捨五入したのであつた. 故に嚴格に云へば,表の61吋と云ふのは60·5吋以上61·5吋以下の身長を有する者を全部,これに入れたのである. 之を61吋の**階級**と云ふ. 第4表甲は16個の階級に分つたのである. 又,最初の階級は60·5吋以下の身長を有するもの全部,最後の階級は74·5吋以上を有する者全部をこれに入れたのである. かく,變數を階級に區分した場合は,第4表甲の如く**階級中央の値**を記入することもあるが,又乙の如く**階級境界の値**を記入することもある. 何れでも宜しく,第4表乙と甲とは全く同一の事を表はすものである.

第4表 乙

26,000人の兵士身長の千分比分布表 (II).

身長	度數
X—60·5	2
60·5—61·5	2
61·5—62·5	20
62·5—63·5	48
63·5—64·5	75
64·5—65·5	117
65·5—66·5	134
66·5—67·5	157
67·5—68·5	140
68·5—69·5	121
69·5—70·5	80
70·5—71·5	57
71·5—72·5	26
72·5—73·5	13
73·5—74·5	5
74·5—X	3
計	1000

尙ほ第4表甲及び乙は**千分比分布表**で,**度數の絕對數**を示して居ない. この點も例1とちがつて居る.

4. 變數の連續と不連續

變數統計の基礎となる變數には二種ある. 例1の變數は整數値のみを取り,その中間の値を取ることは有り得ない. 即ち,實數値を飛び飛びに取るのである. かくの如き變數を**不連續變數**と稱し,第3表,第9表等,これに類似の例は少くない. その多くは生物の器官の數又はその部分を數へる場合であつて,花瓣の數とか,複葉の小葉の數とか,或は莢中の種子,鰭の棘,舌輪廓,乳頭,一回に生れる子供,細胞の染色體等の數は,何れも不連續變數の例である[1].

例2の場合は全くこれと異つて,身長は最低約60吋から最高約75吋まで,整數に限らず如何なる値でも取り得るのである. 少數の觀測ではこの間の實數値を飛び飛びに取るであらうが,若し,無限に多くの計測を行ふ時は,この區間の總ての實數値を取るものと考へられる. かく一定區間の如何なる實數値をも取り得る變數を**連續變數**と稱す. 生物又はその一部分の長さ,面積,容積,重量,角度の如く一般に大さに關する量; 刺戟に反應する時間,現象の終るまでに要する時間,速度,溫度,握力,その他一般に强さや程度をあらはす量; 又はこれらの比,例へば頭長幅指數等は何れも連續變數である.

連續變數の統計では一の變數値の度數は無限小である. 例

[1] 1分間の脈搏數とか,膵の1平方cm.中に出現するLangerhans氏島の數とか云ふものは連續變數とも考へ得る. 中間の小數値をも取りうるからである. 又分數又は小數の不連續變數も生物統計學には存在するが,この時も一定の數のみを飛び飛びに取るものである.

へば,例2に於て身長65吋なるもの117人有りと云ふがこれが皆丁度65吋であつたとは考へ難い．若し,精密な器械を用ひて,嚴格に少數點以下三位又は四位まで計りたりとすれば,丁度65吋であつた者は26,000人中恐らく一人も無かつたであらう．若し無限に精密に計り得たとすれば,正確に身長65吋有りし者の數は無限小の筈である．故に,連續變數の場合は變數は一定の所で區分して,その區分內の度數を取り,度數分布表とするのである．この區分は**階級** (Class; Die Klasse) である[2].

階級の區分の大さを簡單に**階級の幅** (Magnitude of classinterval; Die Grösse des Klassenspielraums) と云ふ．例2では階級の幅は1吋であるが,例3では2mm.である．**階級の境界の値**には**下限** (Die Minusgrenze) と**上限** (Die Plusgrenze) とがある．下限とは一階級の小なる境界値,上限は大なる境界の値である．一階級の上限と下限との和を二分せるものは**階級の中央値**であり,差は**階級の幅**である．

例 3.　日本人男學生6,000人を計測し,その頭最大長徑の度數分布を作つたのが第5表である．

第5表

日本人男學生の頭最大長徑の度數分布表．(松村)

階數	度數	階數	度數	
162·5—164·5	1	188·5—190·5	746	$n = 6,000$,
164·5—166·5	2	190·5—192·5	713	$M = 188{\cdot}55$
166·5—168·5	4	192·5—194·5	576	
168·5—170·5	9	194·5—196·5	389	
170·5—172·5	21	196·5—198·5	269	
172·5—174·5	34	198·5—200·5	171	
174·5—176·5	87	200·5—202·5	76	
176·5—178·5	128	202·5—204·5	34	
178·5—180·5	278	204·5—206·5	24	
180·5—182·5	402	206·5—208·5	12	
182·5—184·5	531	208·5—210·5	2	
184·5—186·5	721	210·5—212·5	1	
186·5—188·5	768	212·5—214·5	1	

(2) 次に不連續變數の度數分布には階級がない．けれども,便宜上屢變數の各値を階級と呼ぶことがある．

例2では最初と最終との階級幅が不定であつたが,それでは平均や,散布度の計算に差支へるから階級は,第5表の如く選ぶべきである．また,第6表のディフテリアの死亡數の表の如く,各階級の幅が同一でないのも計算に不便であるから,成るべくこの例(第5表)の如く,各階級の幅をことごとく等しくすべきである．

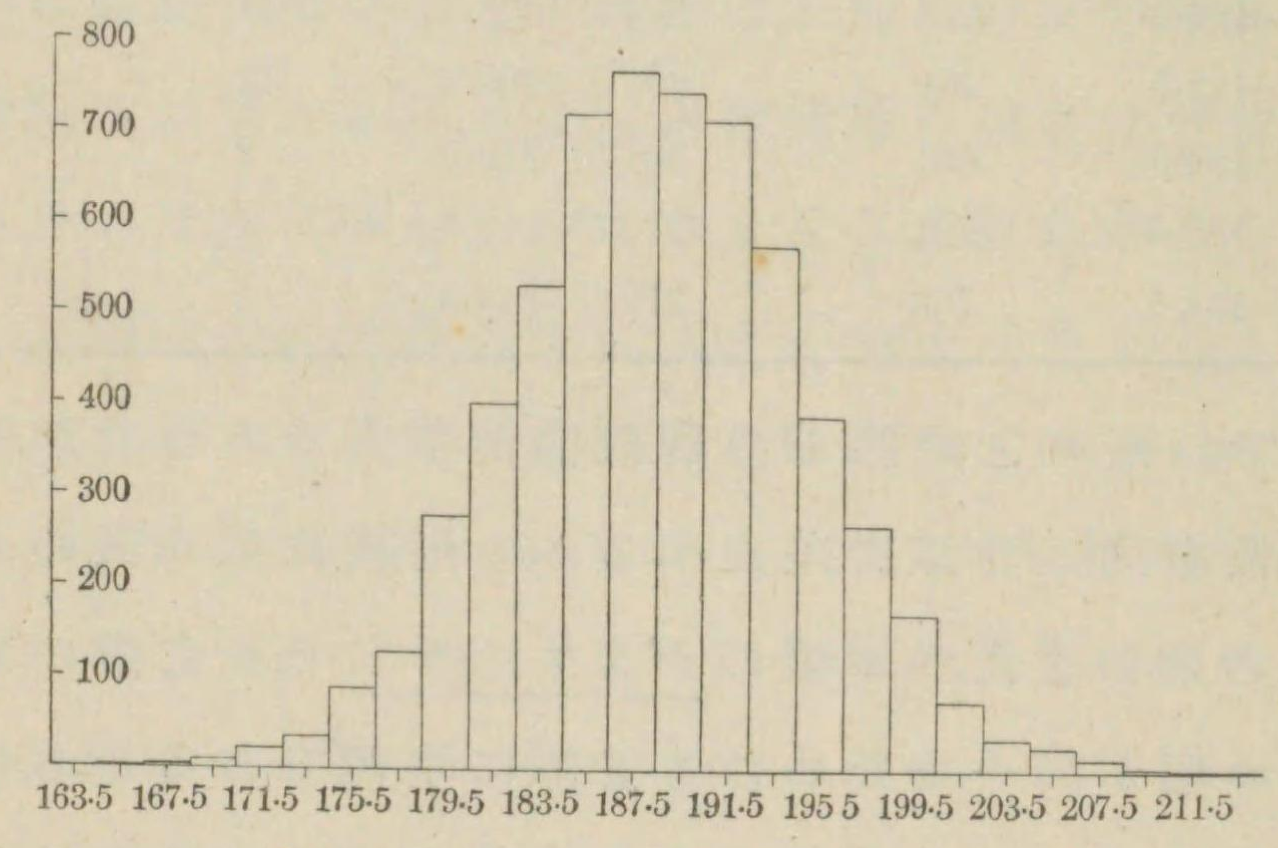
5.　度數分布の圖示

分布表のみでも充分度數分布の概略を理解することが出來るが,これをグラフに現はせば,分布の性情を一目瞭然ならしむることが出來る．度數分布の圖示法に三種あるが,最も普通にはヒストグラムと度數多角形とが用ひられる．

(i)　**ヒストグラム** (Histogram; Die Treppenkurve)

先づ橫線を引き,階級の幅に正比例して橫線を區分し,その上に各階級の度數に正比例する樣に高さを取り,長方形を畫け．各階級につきて長方形を畫けば第6圖の如き圖形となる．これをヒストグラムと稱す．第6圖は第5表のヒストグラムで

ある．一階級の幅を紙上に何 mm. にとるか,度數1を高さ何 mm. に取るとかは全く任意であるが,圖形が大に過ぎず,小に過ぎず,高きに過ぎず,低きに過ぎないやうに,適宜に尺度を選ぶべきである．



第6圖 日本人男學生の頭最大長徑の分布を示すヒストグラム．

二つの度數分布を比較する場合は,勿論尺度を一定にしなければならぬのみならず,度數が觀測實數では誤解をおこすから,千分比分布,または百分比分布等に更めてのち,圖示すべきである．

例 4. 第6表甲及び乙はディフテリアに罹りて死亡せる者の年齢の度數分布表である．甲は初め5歳を階級の幅とし,25歳以上は10歳を階級の幅としたのである．又乙では1歳より5歳迄を更に細かく刻んで分布表とした[1]．

(1) 年齢を變數とする度數分布は函數生物學的にも攻究することが出來る．

第6表

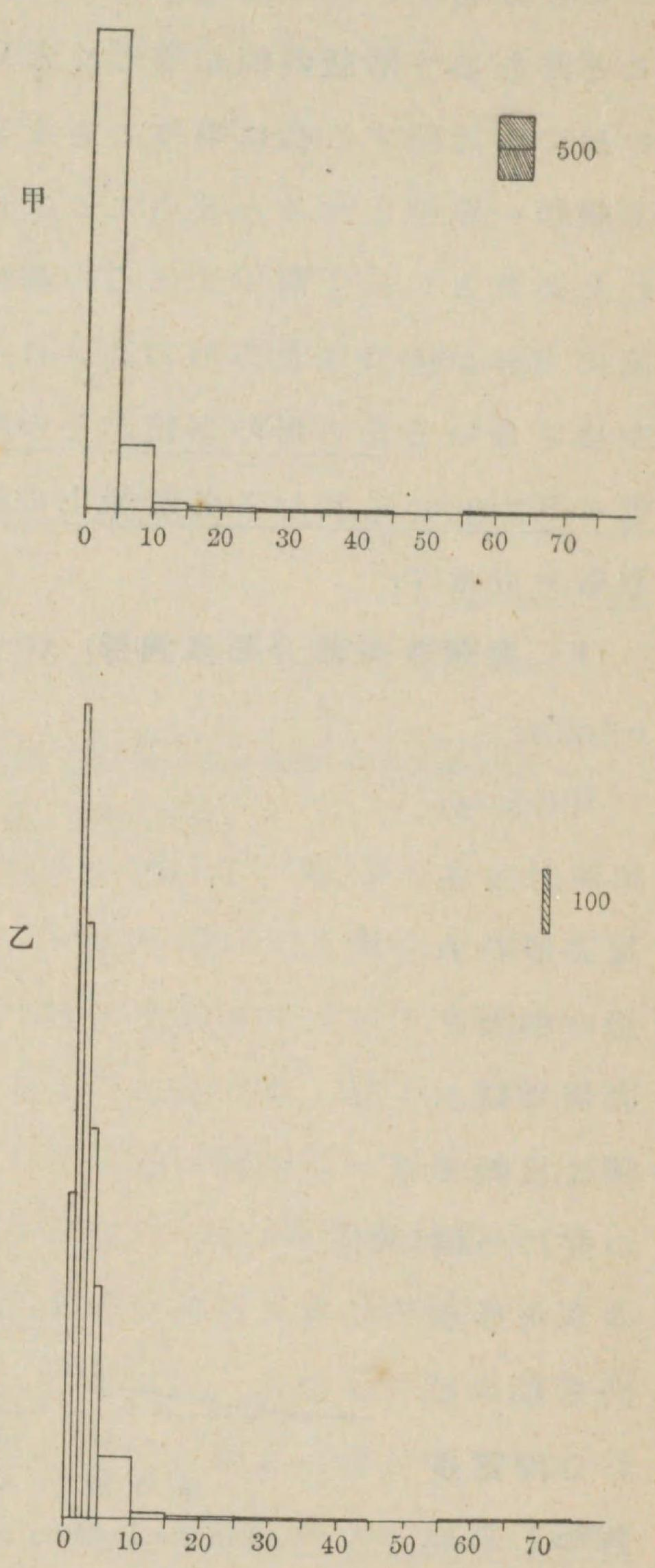
大正5年ディフテリアに罹りて死亡せる者の年齢による度數分布表．(森)

甲

死亡年齢	死亡者數	1年平均の死亡者數
0—5	3588	718
5—10	479	96
10—15	51	10
15—20	12	2
20—25	11	2
25—35	8	1
35—45	7	1
45—55	2	—
55—65	4	1
65—75	5	1
75—X	2	—
計	4169	

乙

死亡年齢	死亡者數	各歳の平均死亡者數
0—1	497	497
1—2	1236	1236
2—3	906	906
3—4	595	595
4—5	354	354
5—10	479	96
10—15	51	10
15—20	12	2
20—25	11	2
25—35	8	1
35—45	7	1
45—55	2	—
55—65	4	1
65—75	5	1
75—X	2	—
計	4169	

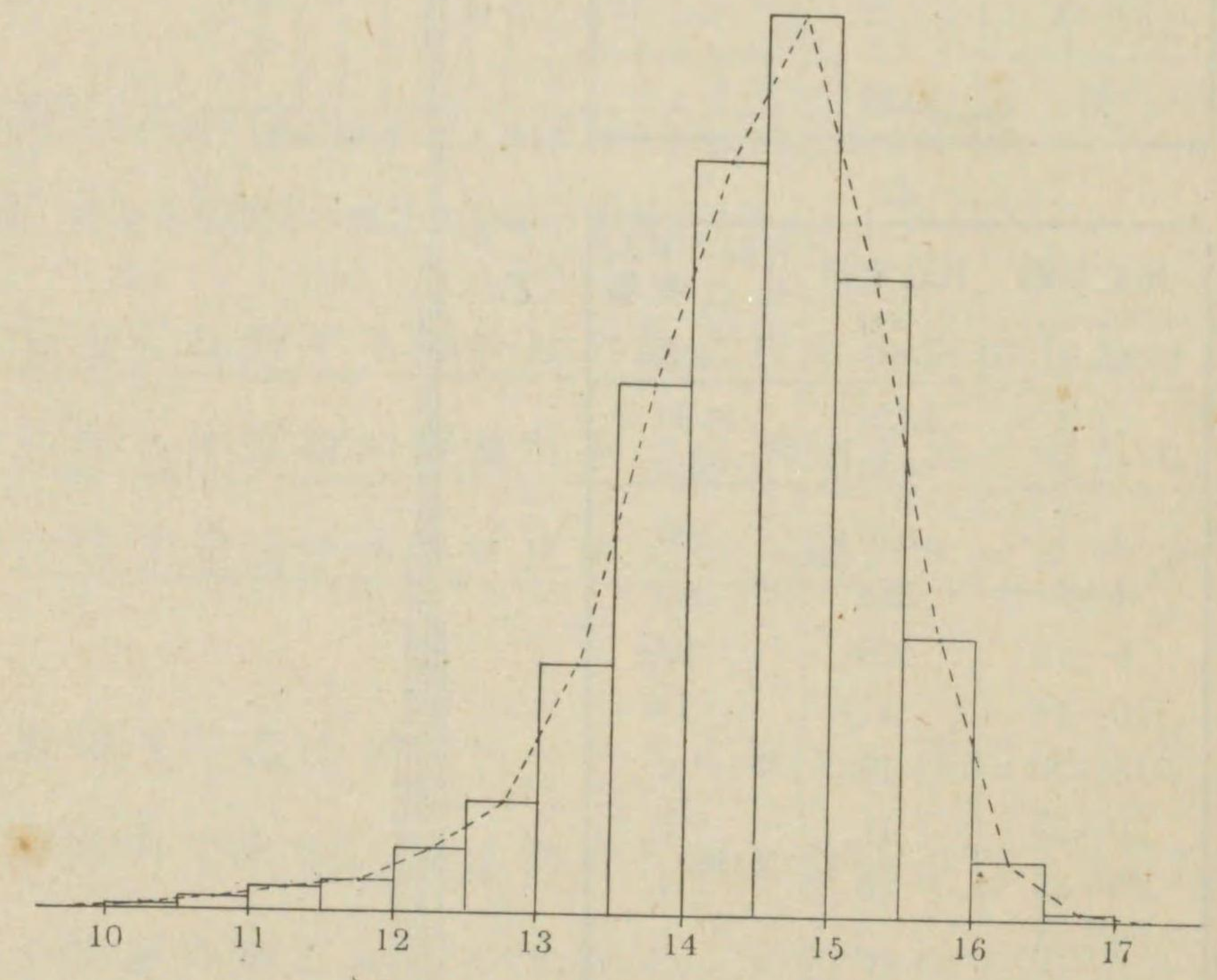


第7圖 ディフテリアに罹りて死亡せるものの度數分布を示すヒストグラム．

この分布表を Histogram であらはすに,注意しなければならぬことがある. 階級の幅が等しくない時には横線の區分は階級の幅に正比例する様に取ることを忘れてはならぬ. 又度數は階級幅一單位とせるときの(ここでは1年の)平均度數を求めこれを高さとし第7圖甲又は乙の様に描くのが合理的である. 故に,Histogram では長方形の高さは必ずしもその階級の度數を表はさないで,長方形の面積がその階級度數になるのである. 從て Histogram に於ける各階級上の長方形の面積の總和は觀測總數を表はす.

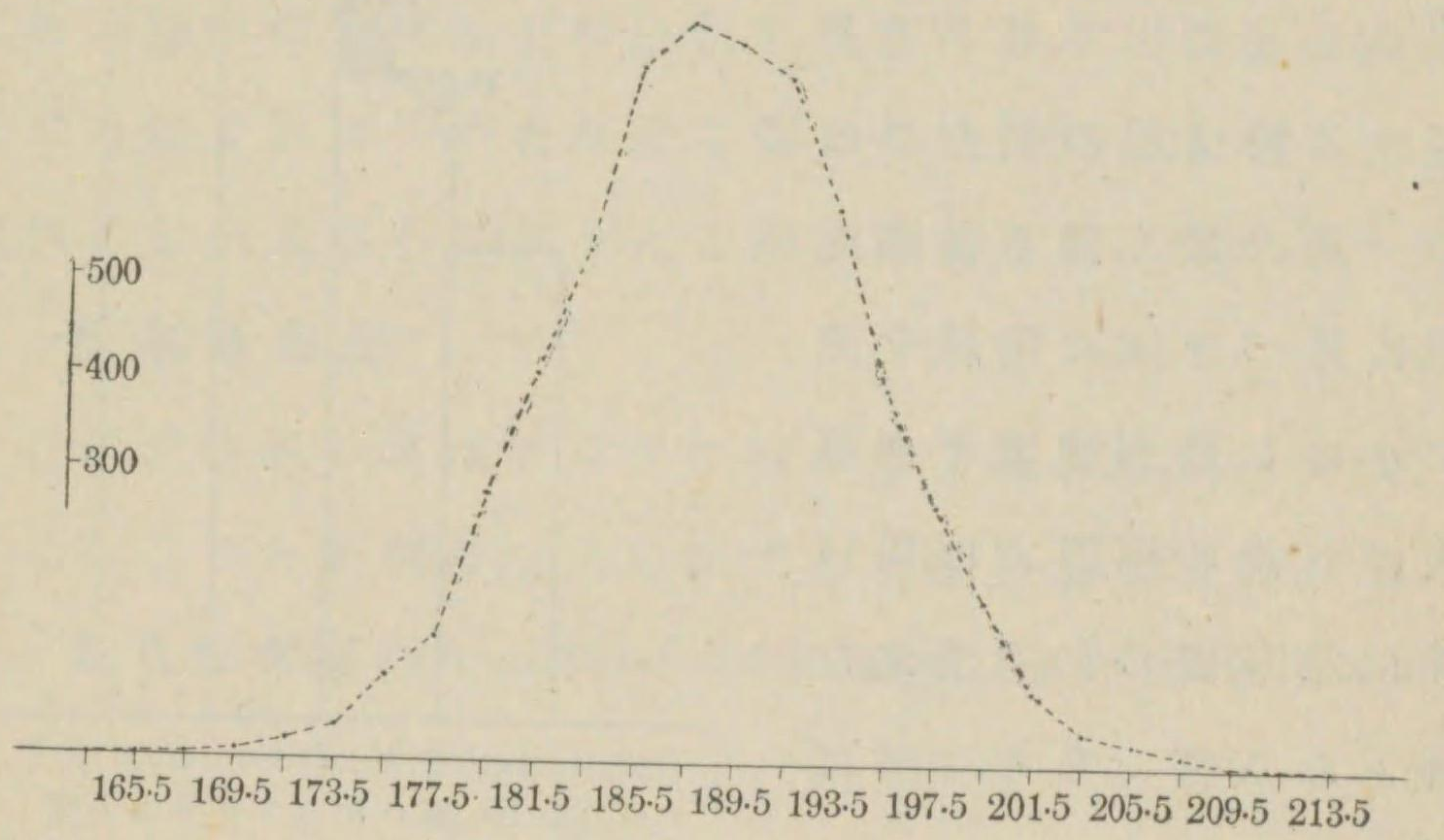
(ii) **度數多角形(分布多角形)** (Frequencypolygon; Die Frequenzpolygon)

Histogram に於ける各長方形の上邊の中央を直線で結ぶ時は,横軸との間に一種の多角形が出來る. これを**度數多角形**と云ふ.



第8圖 ヒストグラムと度數多角形.
Braune Bohnen の種子の長さの分布 (第7表).

第8圖は例5に於ける Braune Bohnen の分布を Histogram と,度數多角形とで示したものである. しかし,かく Histogram をも添へて描く必要はなく,第9圖の如く度數多角形のみで充分である. 後者は日本人男子の頭最大長徑の分布を示すもので,第6圖の Histogram と同一の分布をあらはす.

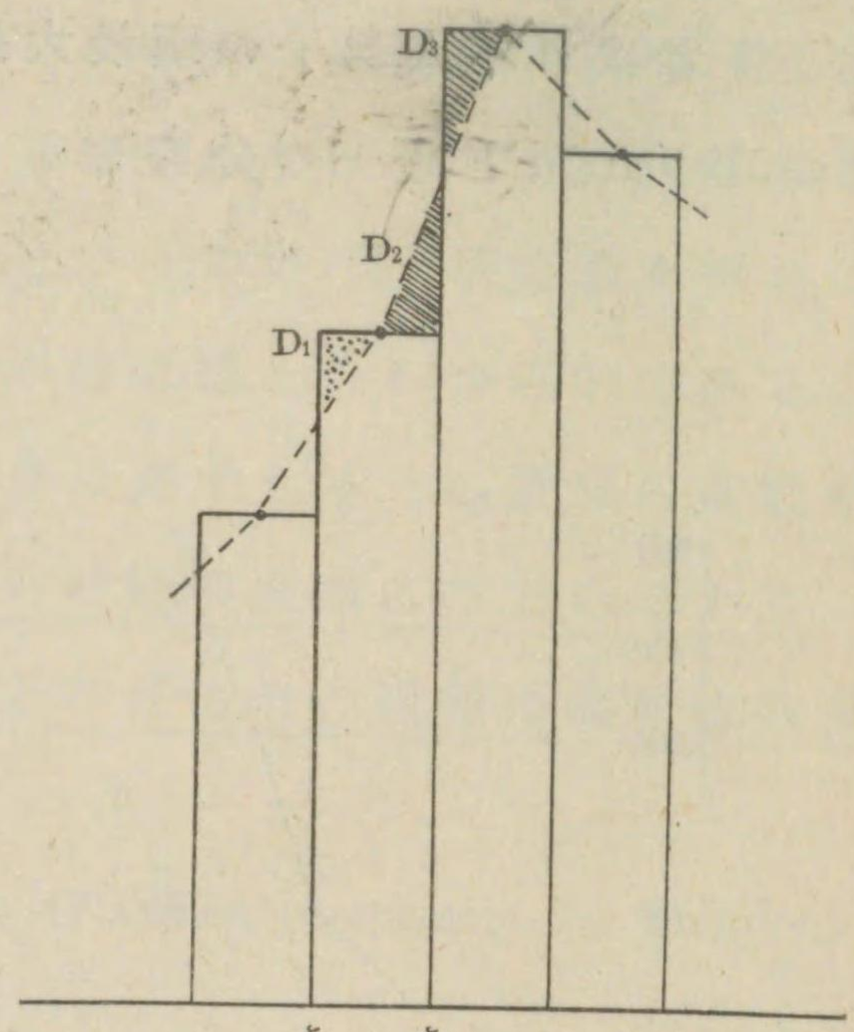


第9圖 度數多角形.
日本人男學生の頭最大長徑の分布を示す (第5表).

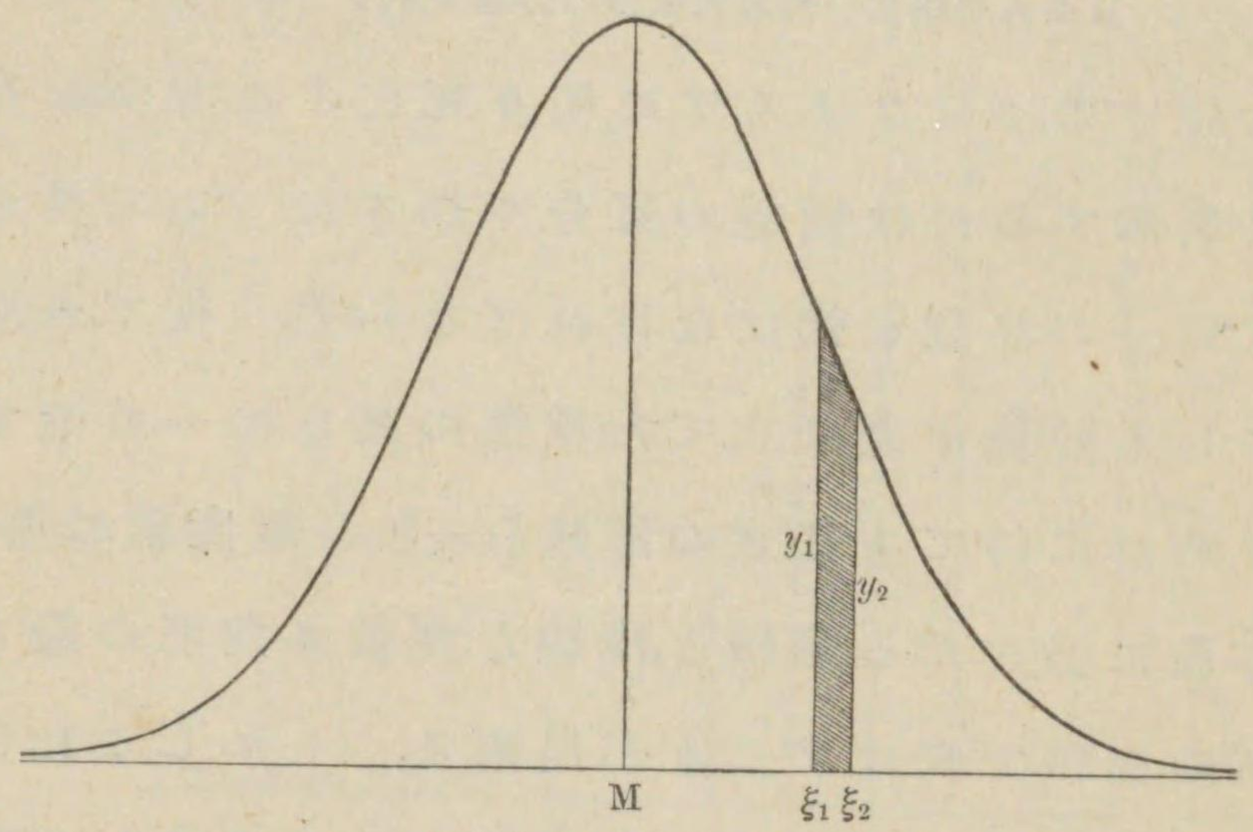
例4に於ける,ディフテリアに罹り死亡する者の分布を度數多角形に表示するには,横線の區分や高さについて Histogram に於けると同一の注意を拂はなければならぬ. 從て,一階級の度數は必ずしも階級中央に立てた垂線の高さに一致せず. 一般に度數多角形に於ても任意の階級 ξ_1—ξ_2 の級度數は略々(第10圖) ξ_1 及び ξ_2 に引きたる垂線と,横線と,度數多角形の邊とで圍まれたる面の面積で表はすことが出來る. しかし,これは嚴格な意味では正しくない. 圖の三角形 D_1 と D_2 とは等しくないからである. 然し D_2 と D_3 とは等しいから,度數多角形の總面積は觀測總數に等しい.

(iii) **度數曲線(又分布曲線)** (Variation curve; Die Variationskurve, Die Frequenzkurve, Die ideale Kurve)

ヒストグラムに於て階級を極めて小にとり,觀測總數を甚しく大にして,グラフを作製する時は,その長方形の頂の一邊の畫く圖形は漸次曲線に近づき,遂に階級を無限に小にし,觀測總數を無限に大にしたる極限では,第11圖又は第12圖の如く滑かな曲線となる筈である. Mは算術的平均の位置である.



第10圖　ヒストグラムと度數多角形との面積の比較.



第11圖　分布曲線の一例 (正常分布曲線).

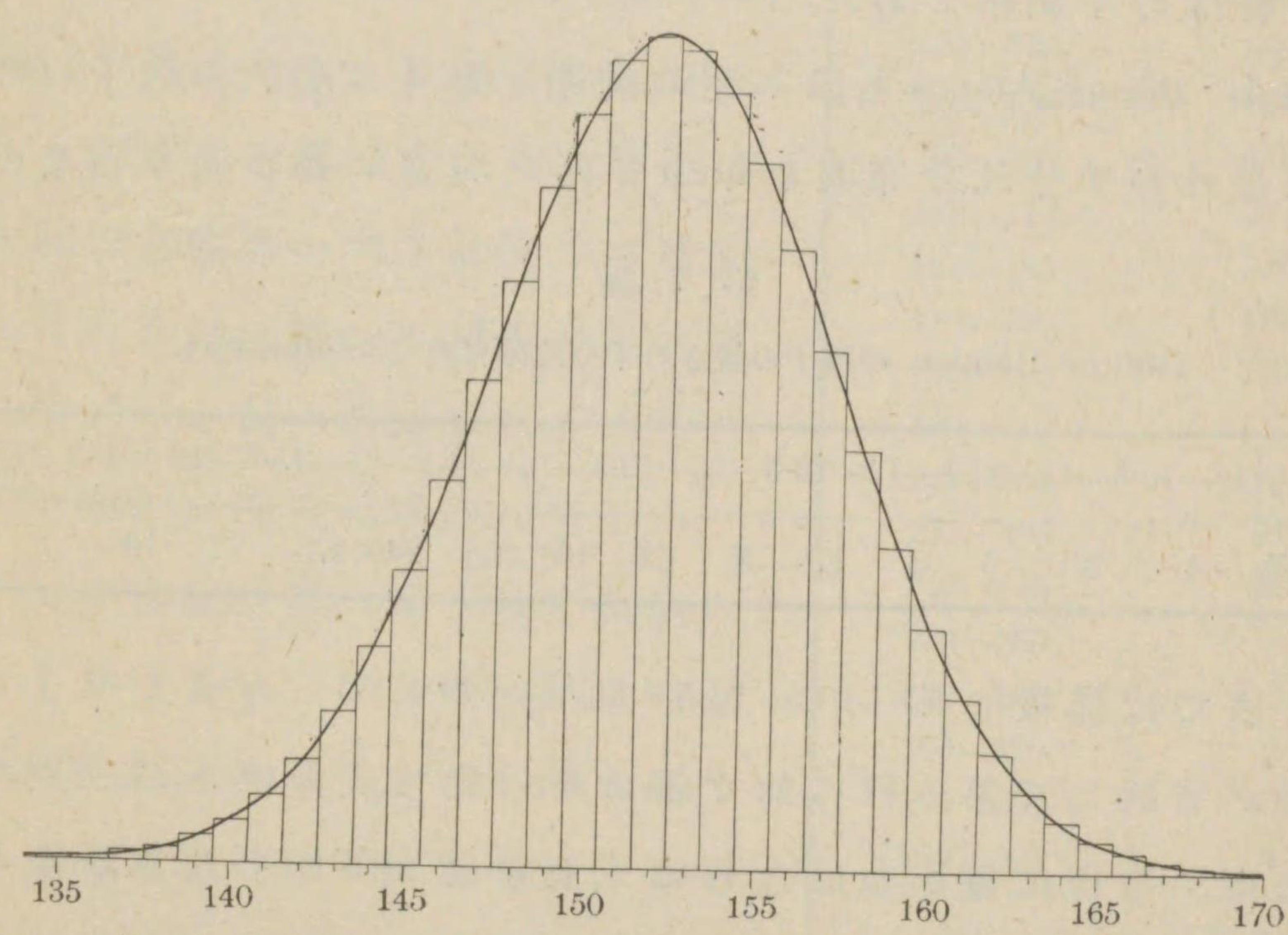
しかし實際に於ては階級は適當に小にすることが出來るけれども,觀測總數を甚しく大とすることは不可能であるが爲に,級度數は偶然の影響を受け,決して滑かな曲線とはならない. 故

に,分布曲線は實測の値からは描くことが不可能であつて,數學的に計算せる結果を圖示するか,又は,目分量で圖示する外はないのである. 曲線と横軸の間の面積は觀測總數をあらはすこと,又,第11圖の如く横軸上に ξ_1, ξ_2 を任意に取り,この上に縱線 y_1, y_2 を引く時,これにて夾まれたる圖形の面積(陰影を施したる部分)は,境界を ξ_1 及び ξ_2 に取りたる階級の級度數に等しい.

6. 基本分布型

分布表を作製し,又はグラフに表示して見れば,種々異つた分布型のあることが判る. その中でも三種の基本型がある.

I. 正常分布型 (Symmetrical distribution ; Die normale Variationsreihe, Die symmetrische Verteilung)



第12圖　正常分布型.
分布曲線とヒストグラムとを示す.

單に相稱型,又は對稱型とも云ふ. 第4表又は第5表を見れば中央の階級に屬する度數が最も大で,それより兩側にゆくに從つて度數は漸次減少し,兩端に至る時は遂に0となる. その狀況を圖示すると更に明になり,第6圖の Histogram,第9圖の度數多角形等何れもよくこれを表示する. 曲線であらはせば,第11圖の如く分布の中央より兩側に向ひ,脚が對稱的に展開し,遂に殆ど橫線に合するのである. 第12圖は正常分布型を度數曲線と Histogram とで表はして比較したものである.

II. 輕度の不相稱型 (Moderately asymmetrical distribution; Die leicht asymmetrische Verteilung)

正常相稱型は生物學上最も屢出現するが,これに次いで多い例は輕度の不相稱である.

例 5. Braune Bohnen と云ふ豆の純系を蒔きて,得たる種子5,000個の長さの千分比分布表を作つて見たが,第7表の通りとなつ

第 7 表

Braune Bohnen の種子の長さの千分比分布表 (Johannsen).

長さ	10—10·5	—11	—11·5	—12	—12·5	—13	—13·5	—14	—14·5	—15	—15·5	—16	—16·5	—17
級度數	1	3	6	8	17	30	68	145	206	246	175	77	16	2

た. 表では階級は 10—10·5, 10·5—11, 11—11.5, ････と書くべきであるが,省略して第一行の如く書いた. 從て,度數は上の二數の間に落ちる樣に書かねばならぬ. 上表は 10—10·5 なる階級の度數が 1, 10·5—11 なる階級の度數が 3, ････なることを示すのである.

最大の級度數は 14·5—15 なる階級にあつて,その千分比度數は 246 である. これより兩側に級度數は減少して居るが,右方は急速に左方は緩慢に減少して居る. これをヒストグラムで表示するならば一層明で,第8圖となる. 第3表に示す油菊の舌狀花の數の分布は一見不相稱型に見えるが顯著でなく,先づ相稱と考へていい.

例 6. 英國成人男子 7,749 人の體重度數分布.

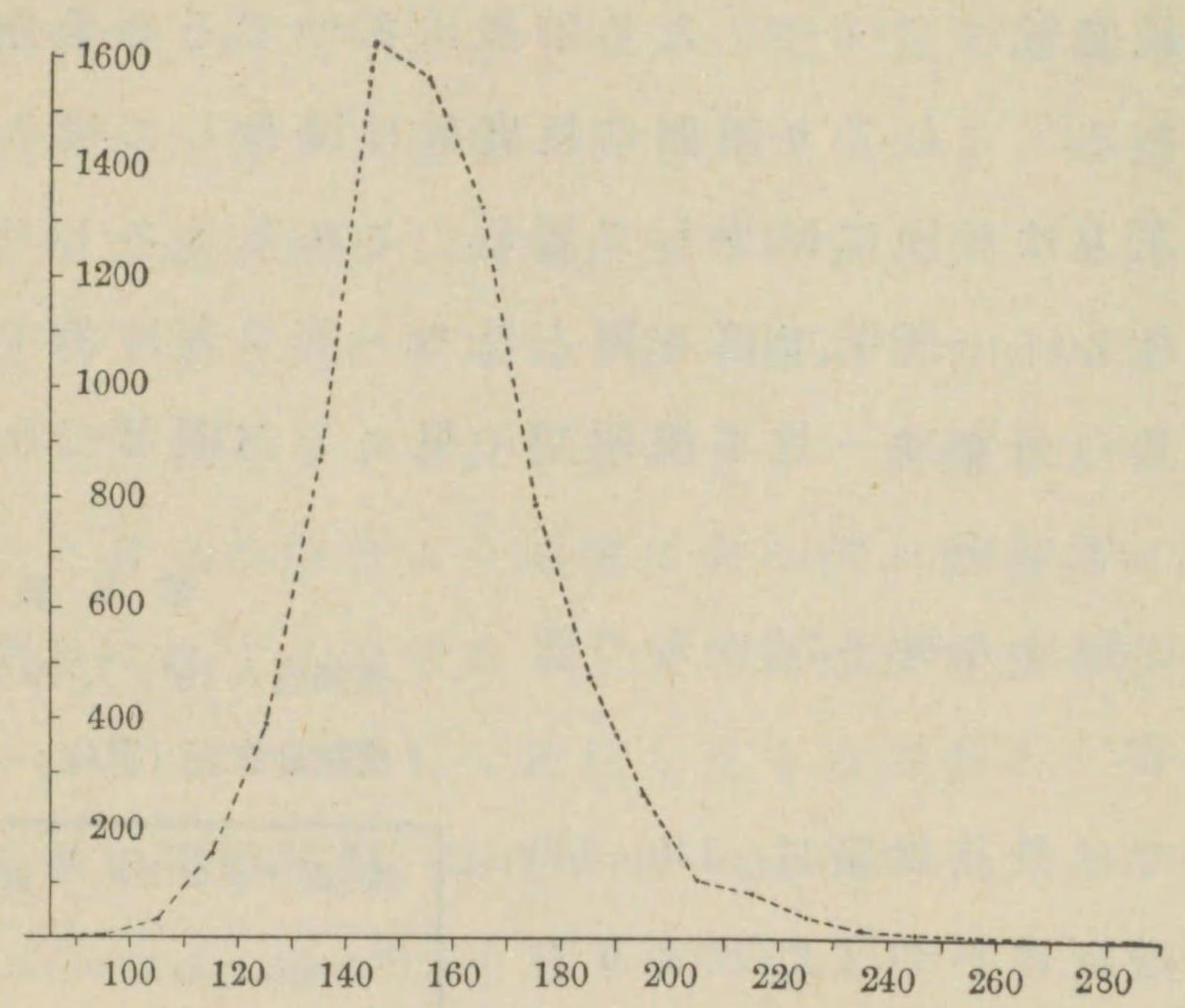
この例では最高度數は 140—150 ポンドの階級にあつて,これより右側には緩慢に,左側には急に傾斜した曲線を得るのである(第13圖). 傾斜の方向は例5と反對であるが,これも輕度の不相稱を示して居る.

輕度の不相稱の例は植物界に屢あらはれるが,人體にも體重などに屢來る樣である. 花瓣の數とか,舌狀花の數とか,指紋の隆線の數とか,一般に不連續變數の場合は不相稱なることが多い. 第13圖は例6の分布を度數多角形で表示したもの,又第14圖は輕度に不相稱なる分布の度數曲線を示したのである.

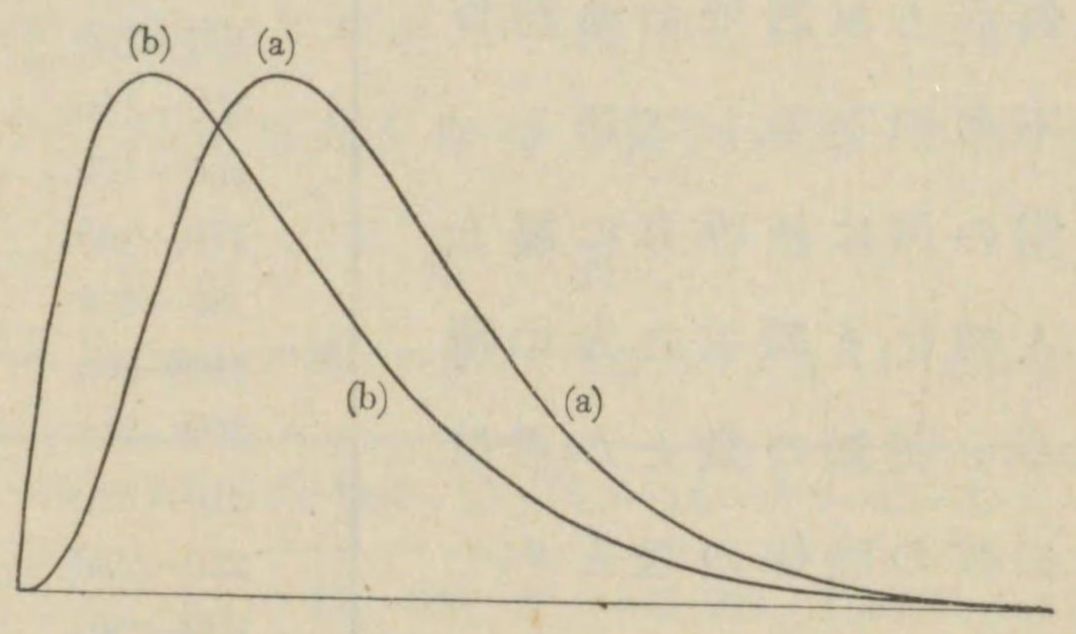
第 8 表

英國成人 (♂) 7,749 人の體重度數分布表 (單位: ポンド).

體重 (ポンド)	度數
90—100	2
100—110	34
110—120	152
120—130	390
130—140	867
140—150	1623
150—160	1559
160—170	1326
170—180	787
180—190	476
190—200	263
200—210	107
210—220	85
220—230	41
230—240	16
240—250	11
250—260	8
260—270	1
270—280	0
280—290	1
計	7749



第 13 圖　輕度に不相稱なる度數分布.
第 8 表を度數多角形にて表はせるもの.



第 14 圖　輕度に不相稱なる度數分布.
度數曲線を以て示す. (Elderton 原圖)

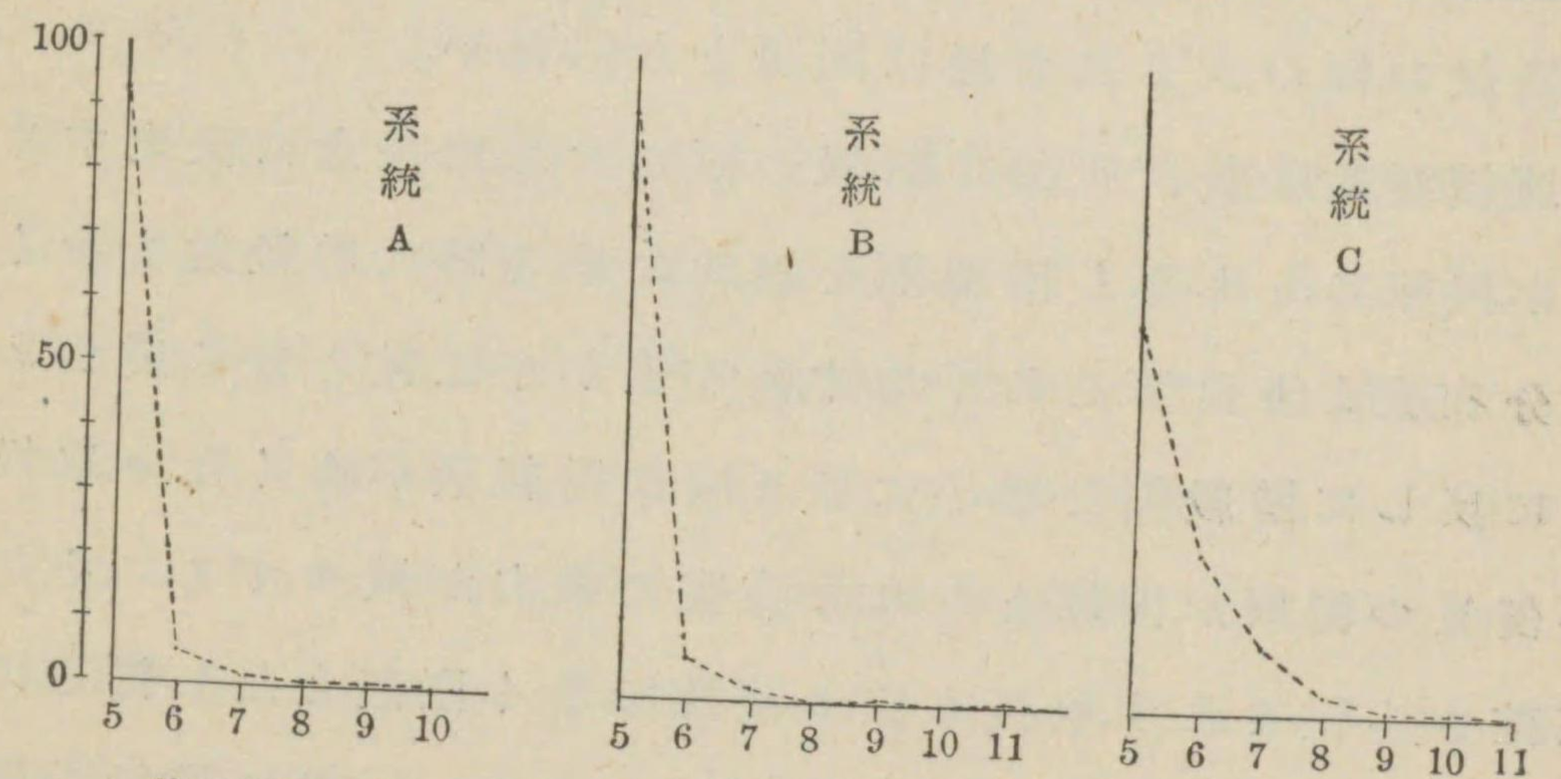
III. 甚しき不相稱型 (J-shaped Distribution; Die stark asymmetrische Verteilung, Die einseitige Verteilung)

分布が甚しく不相稱であつて,最高度數の變數(又は最高度數の階級)が全く一端に偏せるもので,度數はこれより一方に漸次低下して,遂に反對側では 0 となるものである. 故に,度數多角形又は分布曲線は J-字形を示す(或は裏返したる J-字形)が故に, **J-字形分布**とも云ふ.

例 7. ウマノアシガタ屬の一種 *Ranunculus bulbosus* の三系統につきて花瓣數を計へて,第 9 表を得た. 括弧中の數字は百分比の度數である. 又第 15 圖は百分比分布を度數多角形に表示

第 9 表
Ranunculus bulbosus の三系統の花瓣數.

花瓣の數	度數		
	系統 A	系統 B	系統 C
5	312 (92·6)	345 (90·8)	133 (59·9)
6	17 (5·0)	24 (6·3)	55 (24·8)
7	4 (1·2)	7 (1·9)	23 (10·4)
8	2 (0·6)	— —	7 (3·1)
9	2 (0·6)	2 (0·5)	2 (0·9)
10	— —	— —	2 (0·9)
11	— —	2 (0·5)	— —
計	337(100·0)	380(100·0)	222(100·0)



第 15 圖 *Ranunculus bulbosus* の花瓣數の分布を示す多角形.

したものであるが,明に J-型分布である.

J-型分布は不連續變數の分布に屢出現する. 連續變數の場合は極めて稀であるが,例 4 に於けるディフテリアに罹りて死亡せる者の,年齡に關する度數分布は,その一例であつて,第 6 表甲(第 7 圖甲)では全く J-型の分布をして居る. しかし,これは階級を粗大に取つたが爲に起る現象であつて,若し階級の幅を一年とすれば, J-型分布でなくなること乙の如くなる. しかし不相稱の程度は輕度でないから强度の不相稱型に屬する.

かく見て來れば,輕度の不相稱型と强度の不相稱型との間には,中間型があつて完全に區別出來ないことは,前者を正常型と區別すること困難なのと同じである.

7. 度數分布表を作製するに當り注意すべき事項

(i) 度數分布表を作製するに當り,最初に考慮すべきことは階級の幅の大さと,階級の數とである. 不連續變數の場合は變數値がそのまま階級となる. 唯,階級數が餘り多すぎる場合に,二三の變數値を合同する場合もあるが,これは寧ろ例外で有つて,階級の幅の大さは普通は問題とならない.

連續變數の統計に於て,階級の幅の大さは任意に決定することが出來る. しかし,階級幅を過大に取る時は,階級數も少く,度數分布表より計算せる平均や,散布度が不正確となる虞がある. 之に反して,階級幅を過小に取る時は階級數が多くなり,取扱に不便,且つ概觀が困難となり,分布表の效力が減少する. そこで,正確さも侵されず,取扱も餘り困難でなく,概觀もなし得るには何の位の階級幅にすべきかが問題となる. 一般に,階級の幅の

大さは計測の目的物によつて異ふのであつて,身長に於ては幅の大さ 1 cm. としても未だ小に過ぎる感があるが,鼻幅徑などは 1 mm. としても尙大に過ぎ,毛髮の直徑などは $\frac{1}{100}$ mm. 以下としなければならぬ. しかし階級數を標準として階級幅の大さを決定すれば,概略正しい結果を得る樣である. 階級數が 40 以上になるのは煩雜であつて概觀も困難であるし, 12 以下になる樣では正確さが侵される. 一般に階級數 20—30 位になるやう階級幅を選ぶが一番便利である. この點より見て例 2 及び例 5 は階級數が稍少きに過ぎるが,これは概觀に便利な例を選んだ結果である.

例 8. 第 10 表甲は牛乳搾取量の度數分布表であるが,階級幅は大で 50 ガロンとした. (A) では階級數 28 となり, (B) では 21 となつた. 又,乙では階級幅を 3 週間とし,從て階級數 27 となつた. この例では階級幅が大であるから,不正確に過ぎはしないかとの懸念も起るであらうが,變異の甚しき場合では僅少の誤差は問題とならないので有つて,これなども適當に分類せられて居る一の例としてよからう.

常に正確にして概觀を良くする爲には,先づ階級數 20—50 位の分布表を作りて,平均や散布度等を計算し,後に必要に應じその二階級乃至三階級を合同せしめて,簡單な度數分布表を作り,分布の狀況を精査するのも過の無い一の方法である. 幅の大さはなるべく一單位,二單位,五單位,十單位等にとるがいい. 頭最大長徑ならば 1 mm., 胸圍ならば 5 mm., 身長ならば 1 cm. 等とする樣なものである.

第10表　甲

6,648頭の牝牛の牛乳搾取量の度數分布表（單位：ガロン）．(Tocher)

搾取量（階級中央の値）	全部の乳牛につきての度數(A)	年齢2歳及び3歳なる牝牛につきての度數(B)
25	1	—
75	1	—
125	2	1
175	5	2
225	25	16
275	57	39
325	100	60
375	177	113
425	302	212
475	503	291
525	637	308
575	806	396
625	829	301
675	762	231
725	688	156
775	527	114
825	448	80
875	248	45
925	189	20
975	108	14
1025	91	3
1075	43	7
1125	40	1
1175	19	—
1225	21	—
1275	12	—
1325	6	—
1375	1	—
計	6648	2410

第10表　乙

牝牛の乳汁分泌期間の長さの度數分布表（期間の長さは週を單位とす）．(Tocher)

期間の長さ（階級中央の値）	度數
8	1
11	—
14	4
17	28
20	40
23	76
26	197
29	444
32	831
35	1443
38	1399
41	1016
44	534
47	294
50	165
53	75
56	41
59	21
62	13
65	8
68	9
71	2
74	1
77	2
80	2
83	1
86	1
計	6648

又,階級幅は計測器械の目盛の歩みの爲に左右せられる．目盛が精密でない爲に止むを得ず幅を大に止めて置き,從てその時は階級數が少な過ぎるも止むを得ない．目盛の歩みが1mm.の器械で計測して$\frac{1}{10}$mm.を階級幅とするなどは滑稽である．

尙ほ,注意すべきことは,階級幅は等分に取るべきであつて,第6表のディフテリアの死亡數分布の如く階級幅が同一で無かつたり,例2の如く兩端の階級の幅が不定であつたりするのは避けるがいい．例1(第3表),例3(第5表),例6—例9(第7表—第11表)等は模範と出來る例であらう．

(ii) 次に階級の境界を決定すべきである．

連續變數の場合ははつきりと境界値を決定して置かないと,屢誤解を招く．例6(第8表)の如く整數値を階級の境界の値とすることもあり,例2の如く整數値を階級の中央の値として階級を構成することもある．何れにしても連續せる二階級の間には,隙間のない樣に心掛くべきである．

例 9. 第11表は朝鮮人の頭骨指數の度數分布表を二樣の階級表示法で示したのである．嚴密には11表乙の如く,各階級間に間隙の無い樣に境界を選び,その通り表記せねばならぬが,往往甲の如く表記した報告に接することが有る．これでは連續せる二階級間に間隙が有るが,變數はこの間隙の値を取らないわけでは無い．この値を取つた時は四捨五入して何れかの階級に編入したるものと解釋すべきである．かくの如き記法も決して誤りではないが,理論的度數分布の計算の際(第13章參照)過失を起し易いから,なるべくは乙の如く表記するのが安全で

第11表 甲

分類	階級	度數 ♂	度數 ♀
超長頭	67—67·9	1	—
	68—68·9	—	1
	69—69·9	1	1
長頭	70—70·9	3	—
	71—71·9	3	1
	72—72·9	5	2
	73—73·9	4	4
	74—74·9	6	3
中頭	75—75·9	9	3
	76—76·9	6	6
	77—77·9	10	10
	78—78·9	11	9
	79—79·9	7	7
短頭	80—80·9	8	14
	81—81·9	11	5
	82—82·9	11	6
	83—83·9	10	13
	84—84·9	11	10
超短頭	85—85·9	10	9
	86—86·9	9	8
	87—87·9	8	9
	88—88·9	3	2
	89—89·9	3	1
過超短頭	90—90·9	1	1
	91—91·9	—	3
	92—92·9	—	1
	93—93·9	—	1
	94—94·9	1	1
	95—95·9	—	1
	96—96·9	—	—
	97—97·9	—	—
	98—98·9	—	1

第11表 乙

分類	階級	度數 ♂	度數 ♀
超長頭	66·95—67·95	1	—
	67·95—68·95	—	1
	68·95—69·95	1	1
長頭	69·95—70·95	3	—
	70·95—71·95	3	1
	71·95—72·95	5	2
	72·95—73·95	4	4
	73·95—74·95	6	3
中頭	74·95—75·95	9	3
	75·95—76·95	6	6
	76·95—77·95	10	10
	77·95—78·95	11	9
	78·95—79·95	7	7
短頭	79·95—80·95	8	14
	80·95—81·95	11	5
	81·95—82·95	11	6
	82·95—83·95	10	13
	83·95—84·95	11	10
超短頭	84·95—85·95	10	9
	85·95—86·95	9	8
	86·95—87·95	8	9
	87·95—88·95	3	2
	88·95—89·95	3	1
過超短頭	89·95—90·95	1	1
	90·95—91·95	—	3
	91·95—92·95	—	1
	92·95—93·98	—	1
	93·95—94·95	1	1
	94·95—95·95	—	1
	95·95—96·95	—	—
	96·95—97·95	—	—
	97·95—98·95	—	1

ある. 甲表の方が簡單ではあるが,乙表の方が正確である.

(iii) 階級の幅と,境界の値とが定まれば,階級がすべて書き下すことが出來るから,次に計測せる値を分類して何れかの階級に編入するのである. 一階級に一觀測値を增すことに,一丅下𤴓正又は | || ||| |||| 卌 等の記號を記入する. 分類の際丁度境界の値があつた時は,その境界の上下の二階級に$\frac{1}{2}$づつ度數を增すこととする. その時は度數に0·5が出來る(例10). それでも差支ないが,今後の計算が不便であるから,成るべく度數に小數が出來ない樣に階級の値を選擇するのが上分別である. その爲に階級の中央の値が整數にならずとも差支ない. 例へば指數(例9)の度數分布表を作る際,(70—71),(71—72),······等を階級と決定しても,(70·5—71·5),(71·5—72·5),······等を階級と決定しても,境界の値を取るものが多く,度數に小數 0·5 が出て來る. 若し(70·0—70·9),(71·0—71·9),······等を階級と選ぶ時は,變數が境界の値を取ることが普通無いから,階級を第11表甲の如く選ぶ. 但しこれでは階級間に間隙が有る樣に見えるから,乙表の如く記して置くがいい. 極めて多數の記錄を分類する時は分類に器械を用ふることがあるが,生物學方面ではそれ程の必要もない.

例 10. 英國人1,078人の身長の分布.

第 12 表

英國人の身長の度數分布表 (吋).

階級	度數	階級	度數
58·5—59·5	3	61·5—62·5	17
59·5—60·5	3·5	62·5—63·5	33·5
60·5—61·5	8	63·5—64·5	61·5

64·5—65·5	95·5
65·5—66·5	142
66·5—67·5	137·5
67·5—68·5	154
68·5—69·5	141·5
69·5—70·5	116
70·5—71·5	78
71·5—72·5	49
72·5—73·5	28·5
73·5—74·5	4
74·5—75·5	5·5
計	1078

第12表は度數に小數が出て來た例である.

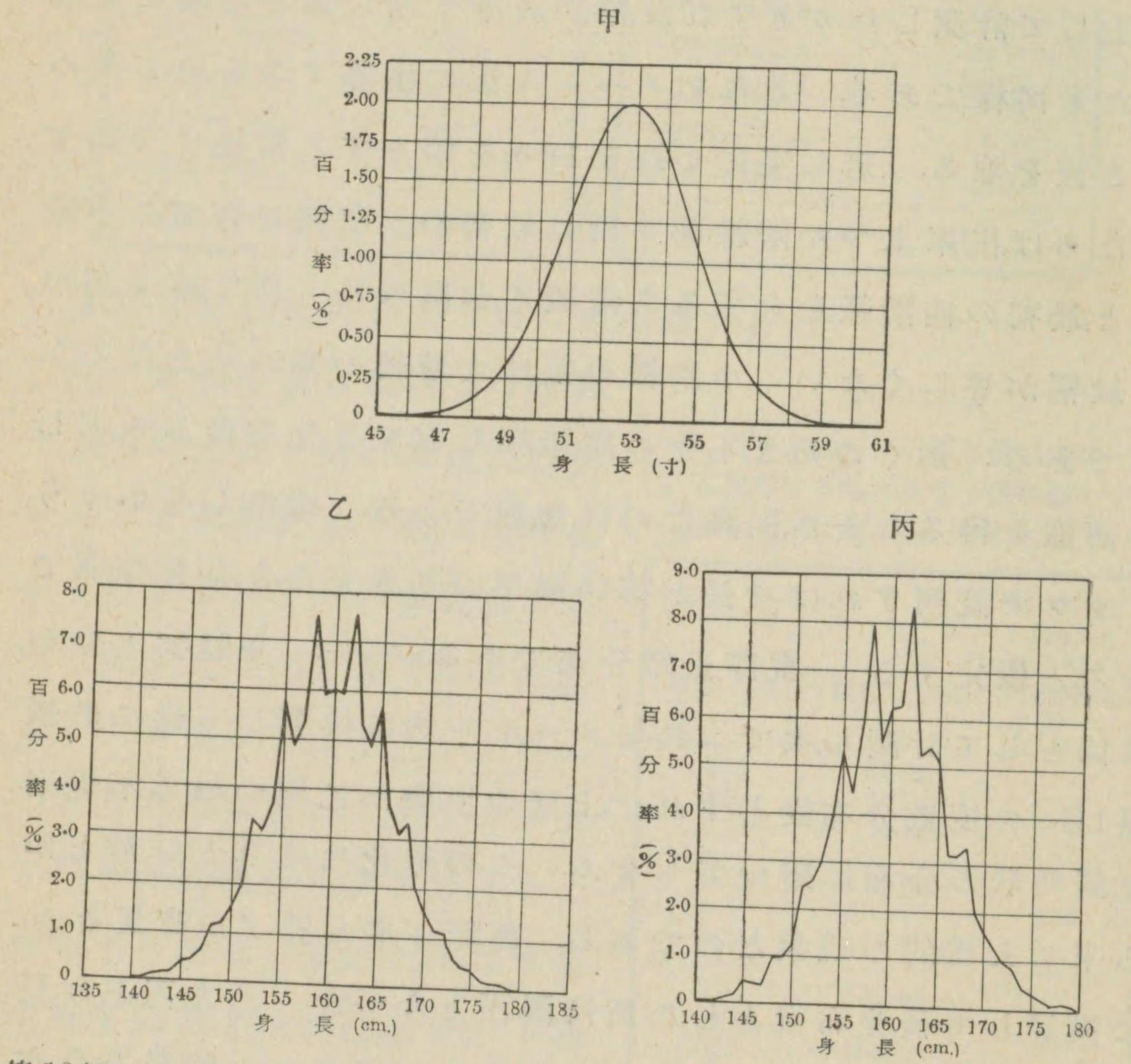
(iv) 計測が或る單位で行はれてある時,これを他の單位に換算したいことがある. 例へば身長が1分を單位とし,日本固有の尺度で表示せられてある時,これをメートル法に換算したいと云ふ場合,日本の尺度の2分又は3分を幅とし度數分布表を作り,その分布型を精査し,圖示し,又は平均や標準偏差を計算し,計算が一通り終了して後,これらの平均や標準偏差をメートル法に換算するがいい. 決して計測値を米法に換算して後,度數分布表を作製してはならぬ. でないと,計算に多くの時間が費えるのみならず,分布に著しい不規則が出來て,却て誤解を招くことがある. 今身長を徴兵檢査の時の如く1分を單

第 13 表

日本尺度にて計測し後に米法に換算せる時に生ずる階級幅の不同を示す表. (* 大階級)

尺寸法 (單位：寸)	メートル法 (單位：cm)	階級幅1 cm. とせる時の階級中央の値
514 515 516	155·76 156·06 156·36	156
517 518 519	156·67 156·97 157·27	157
520 521 522 523	157·58 157·88 158·18 158·48	*158
524 525 526	158·79 158·09 159·39	159
527 528 529	159·70 160·00 160·30	160
530 531 532	160·61 160·91 161·21	161
533 534 535 536	161·52 161·82 162·12 162·42	*162
537 538 539	162·73 163·03 163·33	163

位として計測したりとすれば,第13表第1列の如き階級に分類したも同樣である. これをメートル法に換算すると,第2列の如き値を取る. 若し更にこれを1 cm. を幅とする階級に分類するならば,出來上つた階級(第3列)は最初の三階級を合する小階級と,最初の四階級を合する大階級とが出來る. 從て,第3列の階級幅が等しくない. 小階級の幅は大階級の幅の$\frac{3}{4}$となるからである. かくの如き,不定の階級幅を有する分布表から,正しき認識を得ることが困難なのは勿論である. 理解し易いやうグラフで說明すれば,身長が第15圖甲の如き完全な正常分布であると假定する——實際正常分布であるが——. 今最初1分を單位として計測し,次でこれをメートル法に換算して後に,階級幅1 cm. の度數分布表とすれば,上記の甲圖が乙圖の如き不規則な鋸齒狀の分布に變つてしまふ. この變化は換算と分類とから來る人爲的の現象なのである. 實際日本人壯丁の身長をcm. に換算して後,分布表とした所同圖丙となつた. 乙と丙とは見別けるに困難する程よく似通つて居る. 丙のこの特異な分布も,換算と分類とから來る人爲的現象であつて,元來は甲に近き正常分布であつたことが,推しはかられるのである. 乙にしても丙にしても,著しいことは,身長158と162とに特に高い尖つた峯のあることで,これ大階級に屬するからである. 日本人壯丁身長の度數分布に丙の如く,二つの高き峯があることから,日本人の起源につき二元を力說して居る者もあるが,大きな誤解で,日本人の二元であるか否かはこの分布形からは證明できない. 二つの峯は人爲的のものであるからである.



第16圖 尺寸法にて計測せる身長をメートル法に換算せるが爲に分布曲線に起る異狀.
甲. 原分布—日本人壯丁の身長分布は甲の度數曲線で示しうると假定す.
乙. メートル法に換算せるが爲に原分布甲に起る異狀—不規則なる多くの峯を見よ.
丙. 日本人壯丁を1分を單位として計測しメートル法に換算して後に作れる度數多角形—よく乙に一致せるを見よ. (水島原圖)

以上は一例を擧げたのであるが,斯の如き不愉快な人爲的現象が現れるから,すべて單位の換算は最後の數値にのみ行ふべきで,度數分布表作製以前にこの換算を行つてはならぬことを繰返して置く.

(v) 分布の中心を遠く距つた觀測値を棄ててはならぬ. たとひ如何に大なる觀測値でも,又如何に小なる觀測値でも,それ

が過失による數字で有れば格別だが,正しい數であるなら決して棄ててはならぬ. 平均より如何に顯しく偏れる値であつても,眞に存在する値で有るから,これを無視するは大量觀察の趣旨に反する.

分布の兩端の觀測値を棄てることは,平均に大なる影響が無いかも知れぬが,散布度(標準偏差等)を不當に小にする結果となり,分布の眞實を誤らしむることとなる. 或る場合には兩端に飛び離れて大なる値,又は小なる値が存在することは,その分布に特異なる一性情である事もある. 從て過失の爲に過大又は過小なる値を混入せしむることは絕對に避けねばならぬ. 分布の特異性を惹き起すからである.

演習問題

1. 第9表の三系統の分布をHistogramに描き表はせ.
2. 第15圖,三系統の分布を表はす度數多角形の面積が等しくないのは何故なりや.
3. 第10表,甲のA, B二分布を度數多角形に描き,相稱,不相稱の程度を比較せよ.
4. 第10表甲のA分布は,B分布に比し多少不相稱であるのは,何に由來するかを考察せよ. (第四章參照.)

參考文獻

本章の記述につきてはYule, Johannsen, Pearl, Czuber等を見よ.
メートル法に換算することの可否については,

水島治夫(1931);- 東京醫專新誌, No. 2726.
同 人(1931);- 同誌, No. 2734.

第二章
平　均　値

第一項　種々なる平均値

1. 平均値

度數分布表を作り,グラフを作製しただけでは,分布の中心が何處にあるか,分布の密度がどう配分して居るかを,正確に把握するに尙不充分である. その爲に,平均値と散布度が計算せられねばならぬ.

定義 I. 平均値 (Averages; Die Mittelwerte) とは分布の中心にしてその代表的數値を云ふ.

平均値は分布の中心を表はす數値であるから,その中心の決定の仕方が異なれば,平均値にも種々なるものを作りうる. その主なるものは次の六つである.

(1) 算術的平均 (Arithmetic mean; Das arithmetische Mittelwert),

(2) 中央値 (Median; Die Median),

(3) 形態(流行値) (Mode; Das Modalwert),

(4) 幾何平均 (Geometric mean; Das geometrische Mittelwert),

(5) 調和平均 (Harmonic mean; Das harmonische Mittelwert),

(6) 重みづけられたる平均 (Weighted mean; Das gewichtete Mittelwert)

2. 平均の具備すべき條件

(1) 平均値は確定的の數値であつて,觀る人によつて勝手に評價し,又は計算する人によつて勝手に變動し得る値であつてはならぬ.

(2) 平均はすべての觀測値を基礎とすべきで,或る觀測値が忘れられたり,又は或る觀測値が特に重く,或は特に輕く平均の構成に與つて居てはならぬ.

(3) 平均値は觀測値と同次でなければならぬ. 觀測値が長さならば平均値も長さ,前者が時間,力,重量,指數等ならば平均も亦時間,力,重量,指數等でなければならぬ.

(4) 平均は理解し易く,且つ平易に觀測値を代表するものでなければならぬ. あまりに數學的又は餘りに抽象的のものは不可である.

(5) 平均は迅速且つ容易に計算しうるものでなければならぬ.

(6) 代數的取扱が可能でなければならぬ. これも必要な條件であつて,さうでない平均値は理論的研究が出來ない.

(7) 偶然の影響の爲に平均値の動搖すること小なるものでなければならぬ(第14章參照).

平均の中多く用ひられるのは前記の六つであるが,以上の條件をことごとく備へて居るものは少く,算術的平均が有るのみである. 以下各項にわけてこれを說明する(1).

(1) 分布の中心(平均値)に種々なるものがあると云ふことは一寸不思議に思へるが,中心の定義の決め樣次第で一個とは限らない. 恰かも三角形の中心に重心,內心,外心,垂心等,種々の中心のあると同樣である. 正三角形ではこれら四つの中心が一致する如く,正常分布では算術的平均と中央値と形態とが一致するものである.

第二項　算術的平均

3. 算術的平均

定義 II. 觀測値の總和を觀測總數で除したるものを算術的平均と云ふ.

觀測總數を N とし, N 個の觀測値の各を $V_1, V_2, V_3, \cdots\cdots, V_N$ とすれば,算術的平均 M は

$$\left.\begin{aligned} & M=\frac{V_1+V_2+V_3+\cdots\cdots+V_N}{N}, \cdots\cdots \\ \text{或は}\quad & M=\frac{\Sigma V}{N}, \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \text{又は}\quad & M=\frac{1}{N}\sum_{i=1}^{N} V_i \cdots\cdots\cdots\cdots\cdots\cdots\cdots \end{aligned}\right\} \text{(II. 1)}$$

Σ は總和を表はす記號であつて, $\sum_{i=1}^{N} V_i$ は V の副附號(Subscript) i を 1 より N まで變化せしめ $V_1, V_2, \cdots\cdots, V_N$ を作り,これらの總和をとることを意味する. 又 i の變化が明なる時は單に ΣV とのみ記號することがある. 便利であるから,今後もこの記號を用ふる.

算術的平均はすべての觀測値が同じ程度にその計算に關與し,平易で理解し易く,計算に容易で,代數的取扱が可能である等,平均に必要なるすべての條件を具備して居り,最も屢使用せらるる故に,單に平均とのみ云へば,算術的平均を指す程である.

4. 度數分布表より算術的平均を計算する法

觀測總數の少い時,上記の定義通り,各觀測値より直接に算術的平均を計算することもあるが,普通は多數の計測値が度數分布表となつて居るのであるから,分布表より計算するのが便利である. 度數分布表は分布型を概觀する爲に,又平均値,散布度等を容易に計算する爲に作製するものである[(1)].

今, K を階級數とし, $X_1, X_2, X_3, \cdots\cdots, X_K$ を各階級の中央の値とし, $f_1, f_2, f_3, \cdots\cdots, f_K$ を各階級の度數とすれば,算術的平均 M は,

$$M=\frac{f_1X_1+f_2X_2+f_3X_3+\cdots\cdots+f_KX_K}{f_1+f_2+f_3+\cdots\cdots\cdots\cdots+f_K}, \cdots\cdots\text{(II. 2)}$$

但し　　$f_1+f_2+f_3+\cdots\cdots=N,$

であるから,算術的平均は次の通りに表はすことが出來る.

$$\left.\begin{aligned} & M=\frac{1}{N}\sum(fX)=\frac{\Sigma(fX)}{\Sigma f}, \\ \text{又は}\quad & M=\frac{1}{N}\sum_{i=1}^{K}(f_iX_i)=\frac{\sum_{i=1}^{K}(f_iX_i)}{\sum_{i=1}^{K}f_i}. \end{aligned}\right\}\cdots\cdots\cdots\cdots\text{(II. 2)}$$

$\sum_{i=1}^{K}(f_iX_i)$ も亦前條と同一記號を用ひたので, f_iX_i の副附號 i を 1 より K まで變更して得る積, $f_1X_1, f_2X_2, \cdots\cdots$ の K 個の和を意味する. この場合 i を記入せずとも亦明である.

例 1. 第14表は朝鮮人頭骨(男)の全顏面角の度數分布表であるが,之より算術的平均を計算するには,第三列にある通り fX を求め,これを記入し,これらの總和を求め,これを觀測數93で除す. 第四列以下の計算は第三章を見よ.

(1) この爲に分布表作製にあたり,最初は階級の幅をあまり大にし階級數を少くしてはならぬ. 必ず,最初の分布表は階級數20以上になる樣にしなければならぬ. 又,度數分布表より算術的平均を計算するには,階級幅がすべて齊一でなければならぬ.

第 14 表
朝鮮人頭骨(♂)の全額面角.
(平均及び標準偏差の計算.)

X (階級中央の値)	f (度數)	fX	x	x^2	fx^2
77°	2	154	−7·38	54·46	108·9
78°	3	234	−6·38	40·70	122·1
79°	4	316	−5·38	28·94	115·8
80°	6	480	−4·38	19·18	115·1
81°	8	648	−3·38	11·42	91·4
82°	7	574	−2·38	5·66	39·6
83°	9	747	−1·38	1·90	17·1
84°	6	504	− ·38	·14	0·8
85°	11	935	+ ·62	·39	4·3
86°	6	516	+1·62	2·63	15·8
87°	10	870	+2·62	6·87	68·7
88°	9	792	+3·62	13·11	118·0
89°	5	445	+4·62	21·35	106·7
90°	6	540	+5·62	31·59	189·5
91°	—	—	+6·62	—	—
92°	1	92	+7·62	58·07	58·1
計	$N=93.$	$\sum fX$ $=7847.$		—	1171·9

平均 $M=\dfrac{7847}{93}=84{\cdot}38°.$ $\sigma^2=\dfrac{1171{\cdot}9}{93}=12{\cdot}60,$

$\sum|a|=279{\cdot}86,$ $\sigma=3{\cdot}55.$

$\varepsilon=3{\cdot}01.$

Med.$=84{\cdot}64.$

5. 算術的平均の計算に原點 A を選ぶこと(算術的平均の計算法. II.)

適當に原點を選ぶことに由りて,大なる數を使用することを避け,計算を容易にすることが出來る. 今 A を任意に選びたる

原點とし,

$$V_1-A=x_1,\quad V_2-A=x_2,\quad V_3-A=x_3,\cdots\cdots \quad とす.$$

然る時は $V_1+V_2+V_3+\cdots\cdot+V_N-NA=x_1+x_2+x_3+\cdots\cdot,$

各邊を N にて除す時は

$$M-A=\frac{1}{N}\sum_{i=1}^{N}x_i,$$

即ち

$$M=A+\frac{1}{N}\sum_{i=1}^{N}x_i.\quad\cdots\cdots\cdots\cdots\cdots\cdots(\mathrm{II.}\ 3)$$

A を適當に選ぶ時, $\sum\limits_{i=1}^{N}x_i$ の計算は $\sum\limits_{i=1}^{N}V_i$ の計算に比し容易である. 又,觀測値が度數分布表になつて居る時は,原點 A に或る階級の中央の値を選ぶを常とす; この時は偏差 ξ は普通簡單な整數となるからである.

定理. 各階級中央の値の,任意に定められたる原點 A よりの偏差を ξ とすれば,算術的平均 M は

$$\left.\begin{aligned}M&=A+\frac{1}{N}\sum_{i=1}^{K}(f_i\xi_i)\\&=A+\frac{1}{N}\sum(f\xi)\end{aligned}\right\}\quad\cdots\cdots\cdots\cdots(\mathrm{II.}\ 4)$$

である. 但し $X_i-A=\xi_i$ とす.

算術的平均の計算には,A を分布の中央に近き何れかの階級中央の値に選ぶ時は,ξ は小なる數となり,計算は極めて容易となる.

例 2. 第15表は日本人男學生の頭長幅指數の分布であるが,若しこれを例 1 の如く計算すれば, fX 從て ΣfX が甚だ大なる數となり,面倒であるから,公式 (II. 4) を應用して計算するがいい. 今,其の順序をのべる.

第 15 表　日本人男學生の頭長幅指數.（平均及び標準偏差の計算）

階級	f_i	ξ_i	$f_i\xi_i$	Med. の計算	ξ^2	$f\xi^2$
68—69	1	−12	− 12	1	144	144
69—70	4	−11	− 44	5	121	484
70—71	8	−10	− 80	13	100	800
71—72	14	− 9	− 126	27	81	1134
72—73	36	− 8	− 288	63	64	2304
73—74	82	− 7	− 574	145	49	4018
74—75	147	− 6	− 882	292	36	5292
75—76	238	− 5	−1190	530	25	5950
76—77	326	− 4	−1304	856	16	5216
77—78	434	− 3	−1302	1290	9	3906
78—79	619	− 2	−1238	1909	4	2476
79—80	618	− 1	− 618	2527	1	618
◎ 80—81	709	0	小計−7658	3236	0	0
81—82	657	1	+ 657	3893	1	657
82—83	549	2	+1098	4442	4	2196
83—84	497	3	+1491	4939	9	4473
84—85	338	4	+1352	5277	16	5408
85—86	248	5	+1240	5525	25	6200
86—87	156	6	+ 936	5681	36	5616
87—88	103	7	+ 721	5784	49	5047
88—89	94	8	+ 752	5878	64	6016
89—90	57	9	+ 513	5935	81	4617
90—91	22	10	+ 220	5957	100	2200
91—92	20	11	+ 220	5977	121	2420
92—93	10	12	+ 120	5987	144	1440
93—94	5	13	+ 65	5992	169	845
94—95	3	14	+ 42	5995	196	588
95—96	5	15	+ 75	6000	225	1125
計	6000	—	小計+9502	—	—	81190

$$\sum f\xi = -7658 + 9502 = +1844, \qquad \therefore \ M = 80{\cdot}5 + \frac{1844}{6000} = 80{\cdot}81.$$

$$\sum f\xi^2 = 81190, \qquad s^2 = 13{\cdot}532,$$

$$\sigma = \sqrt{13{\cdot}532 - 0{\cdot}096} = \sqrt{13{\cdot}436} = 3{\cdot}666,$$

$$\bar{\sigma} = \sqrt{13{\cdot}436 - \frac{1}{12}} = \sqrt{13{\cdot}343} = 3{\cdot}653.$$

$$\text{Med.} = 80 + \frac{3000 - 2527}{709} = 80.67.$$

(i) 第一に原點を決めねばならぬ. 分布の中心に近き階級に原點をおくのが便利であるから, 80—81 なる階級の中央の値卽ち 80·5 を原點, A と定める.

(ii) 次に, 各階級中央の値の, A よりの偏差卽ち $X_i - A = \xi_i$ を求め, これを第三列に記入した. 負號は必要である.

(iii) 次に, $f_i\xi_i$ をそれぞれ計算して後, 負の偏差と正の偏差とで別々に $\Sigma f_i\xi_i$ を求め, 次で全體について $\Sigma f_i\xi_i$ を求める.

卽ち負の偏差につきて小計 −7658, 正の偏差につきて小計 +9502, 卽ち,

$$\sum f\xi = -7658 + 9502 = 1844,$$

$$M = A + \frac{1}{N}\sum f\xi = 80{\cdot}5 + 0{\cdot}307 = 80{\cdot}81.$$

6. 算術的平均の計算方法. III.

階級の幅が簡單である時は, 例 2 の通り計算することが出來るが, 階級の幅 0·5, 2, 3, 4, 5 等であることが屢ある. そんな場合は

$$\xi = \lambda\zeta$$

と置き, λ を階級の幅とすれば, ζ は整數となる.

$$M = A + \frac{1}{N}\sum(f\xi) = A + \frac{1}{N}\sum(f\lambda\zeta).$$

而して, λ は各項に共通なる常數であるから, Σ の記號の外に出すことが出來る.

$$\left.\begin{aligned} \therefore \quad & M = A + \frac{\lambda}{N}\sum(f\zeta), \\ \text{或は} \quad & M = A + \frac{\lambda}{N}(f_1\zeta_1 + f_2\zeta_2 + f_3\zeta_3 + \cdots + f_k\zeta_k) \end{aligned}\right\} \cdots \text{(II. 5)}$$

例 3.　前章例 8 に於ける牛乳搾取量の A 分布より算術的平均を計算せよ.

解.　$\lambda=50$ ガロンである(第 16 表). 今 A を 675 ガロンに選び, 第三列に ζ を記入すれば表の如くなる.

第四列は $f\zeta$ の計算の順序を示す. これより

$$\sum f\zeta = 7669 - 10414 = -2745,$$

$$M = A + \frac{\lambda}{N}\sum f\xi = 675 - \frac{50\times 2745}{6648} = 654{\cdot}36,$$

一頭當り平均 654·36 ガロンを搾取したことになる. $\Sigma f\zeta$ の符號を注意せよ.

第 16 表

牝牛の牛乳搾取量の度數分布表.

(平均及び標準偏差の計算.)

X (單位: ガロン)	f (度　數)	ζ	$f\zeta$	累積度數	ζ^2	$f\zeta^2$
25	1	−13	13	1	169	169
75	1	−12	12	2	144	144
125	2	−11	22	4	121	242
175	5	−10	50	9	100	500
225	25	− 9	225	34	81	2025
275	57	− 8	456	91	64	3648
325	100	− 7	700	191	49	4900
375	177	− 6	1062	268	36	6372
425	302	− 5	1510	670	25	7550
475	503	− 4	2012	1173	16	8048
525	637	− 3	1911	1810	9	5733
575	806	− 2	1612	2616	4	3224
625	829	− 1	829	3445	1	829
◎ 675	762	0	小計 −10414			
725	688	1	688		1	688
775	527	2	1054		4	2108
825	448	3	1344		9	4032
875	248	4	992		16	3968
925	189	5	945		25	4725
975	108	6	648		36	3888
1025	91	7	637		49	4457
1075	43	8	344		64	2752
1125	40	9	360		81	3240
1175	19	10	190		100	1900
1225	21	11	231		121	2541
1275	12	12	144		144	1728
1325	6	13	78		169	1014
1375	1	14	14		196	196
	6648		小計 7669			80623

$$\Sigma f\zeta = -10414 + 7669 = -2745,$$

$$c = \frac{\Sigma f\zeta}{N} = 0{\cdot}4129.$$

$$M = 675 - \frac{\lambda}{N}\sum f\zeta = 675 - 50\times 0{\cdot}4129 = 675 - 20{\cdot}645 = 654{\cdot}355.$$

$$\text{Med.} = 600 + \frac{50\times 708}{829} = 642{\cdot}70.$$

$$\Sigma f\zeta^2 = 80623,$$

$$s^2 = \frac{\lambda^2}{N}\Sigma f\zeta^2 = 50^2\times 12{\cdot}127,$$

$$\sigma^2 = s^2 - d^2 = 50^2\times 12{\cdot}127 - c^2\lambda^2 = 50^2(12{\cdot}127 - 0{\cdot}171),$$

$$\sigma = 50\times\sqrt{11{\cdot}916} = 50\times 3{\cdot}4525 = 172{\cdot}63,$$

$$\bar{\sigma} = 50\times\sqrt{11{\cdot}916 - 0{\cdot}088} = 50\times\sqrt{11{\cdot}833} = 172{\cdot}01.$$

7. 算術的平均を度數分布表より計算するが爲に起る誤差

算術的平均は元來觀測値 V につきて,公式(II. 1)によつて計算せらるべきであるが,多數を取扱ふ時は,度數分布表より算術的平均等を計算する. しかし眞の計測値必ずしも階級中央の値を取らないから,常に多少の誤差は免れない. この誤差は多數計測せる場合,完全に相稱なる度數分布では0である. 又完全に相稱でなくとも,吾々が普通遭遇する樣な輕度な不相稱型では,誤差は極めて僅微であるから,これは無視していいのである. けれども,J-型分布,又は高度の不相稱型分布では,負の誤差と正の誤差とが相等しくはないから,度數分布表を作製して計算せる算術的平均は可なり大なる誤差をともなふものである. 故に,これらの甚しく不相稱なる度數分布では,階級を成るだけ細かく刻み,その數30又はなるべく30以上とする必要があるのである.

前章例4に於けるDiphtherie死亡者の度數分布の場合は,平均計算の爲には分布表粗大に過ぎる.

例 4. 第17表は金澤地方に於ける日本人頭骨(男)の上顏面指數の度數分布表である. 初めに階級幅を0·1に取つて度數分布表を作つた. 度數Iの列はそれである. 次に階級幅を1とした分布が度數IIの列で,又階級幅を2としたのがIIIの列である. これらの三つの列より平均を計算して比較するに,階級幅を0·1とした時と1した時とは,算術的平均が殆ど同一であるが,階級幅を2とし從て階級の數を10とした時,稍大なる誤差を生じた. 階級數を一層少くすれば最早用をなさぬであらう.

第17表

金澤地方における日本人男子頭骨の上顏面指數.

(階級分類と算術的平均との關係)

階數	度數(I)	度數(II)	度數(III)
44·8	1	1	5
45·2	2	4	
45·9	2		
46·4	1	5	11
46·7	1		
46·8	1		
46·9	2		
47·0	1	6	
47·3	1		
47·4	1		
47·5	2		
47·8	1		
48·0	1	11	22
48·2	1		
48·3	1		
48·5	2		
48·6	4		
48·9	2		
49·2	3	11	
49·3	3		
49·6	3		
49·7	2		
50·0	10	24	40
50·4	4		
50·7	4		
50·8	6		
51·0	1	16	
51·1	5		
51·3	1		
51·4	2		
51·5	2		
51·6	1		
51·8	1		
51·9	3		

階數	度數(I)	度數(II)	度數(III)
52·1	1	12	26
52·2	2		
52·3	1		
52·4	2		
52·5	1		
52·6	2		
52·7	1		
52·8	1		
52·9	1		
53·0	1	14	
53·1	3		
53·2	1		
53·3	2		
53·4	2		
53·5	1		
53·6	1		
53·7	3		
54·1	1	11	18
54·3	2		
54·4	1		
54·5	1		
54·6	2		
54·7	2		
54·8	2		
55·0	2	7	
55·3	1		
55·6	2		
55·8	1		
55·9	1		
56·1	1	4	6
56·3	1		
56·5	1		
56·9	1		
57·3	1	2	
57·5	1		

階數	度數(I)	度數(II)	度數(III)
58·3	1	3	4
58·5	1		
58·9	1		
59·0	1	1	
61·8	1	1	1
67·4	1	1	1
N	134	134	134
M	51·670	51·666	51·696

8. 算術的平均の持つ特性

算術的平均は初めに述べた様な必要なる條件を具備して居る. 就中,必要な事は代數的取扱の可能なることである. 代數的取扱とはどんな事であるか,今その二三を次に定理として述べる.

定理 I. 各觀測値の,算術的平均Mよりの偏差の總和は0である. 今,N個の觀測に於て,各偏差を $x_1, x_2, x_3, x_4, \cdots, x_N$ とすれば, $x_1 = V_1 - M,\ x_2 = V_2 - M,\ \cdots\cdots,\ x_N = V_N - M,$

故に, $x_1 + x_2 + x_3 + \cdots + x_N = V_1 + V_2 + \cdots + V_N - NM.$

或は,

$$\sum(x) = \sum(V) - NM.$$

しかるに算術的平均の定義により $\sum V = NM.$ であるから,

或は,

$$\left.\begin{aligned} &\sum x = 0, \\ &\sum_{i=1}^{N} x_i = 0. \end{aligned}\right\} \cdots\cdots\cdots\cdots\cdots\cdots\cdots (\text{II. } 6)$$

定理 II. 或る値Aよりの,觀測値の總和が0である時は,この値Aは算術的平均である.

これは定理Iの逆で,證明は極めて容易であるから,讀者これを試みられよ.

定理 III. 觀測値の二つの群あり,第一群に於て觀測總數N_1,算術的平均M_1,又第二群に於てN_2,及びM_2,なりとす. 今,二群を合同して一群とする時,この合同觀測値の算術的平均,Mは

$$M = \frac{N_1M_1 + N_2M_2}{N_1 + N_2} \cdots\cdots\cdots\cdots\cdots\cdots (\text{II. } 7)$$

である.

證明は極めて容易であるから讀者自ら證明を試みられよ.

系 I. 三個以上の群を合同する時は

$$\left.\begin{aligned} M &= \frac{\Sigma(NM)}{\Sigma N}, \\ \text{卽ち}\quad &= \frac{N_1M_1 + N_2M_2 + \cdots + N_kM_k}{N_1 + N_2 + N_3 + \cdots + N_k}. \end{aligned}\right\} \cdots\cdots (\text{II. } 8)$$

系 II. 公式(II. 7)に於て $N_1 = N_2$ とすれば,

$$M = \frac{M_1 + M_2}{2}.$$

卽ち,二群の觀測總數互に等しい時は,合同群の算術的平均は各群の算術的平均の和の二分の一である.

例 5. 朝鮮咸鏡南,北道,平安南,北道にて計測せる男子身長平均は次表(第18表)の通りである. これより北鮮男子の平均身長を求む.

この四群の算術的平均は,

$$M = \frac{90 \times 166{\cdot}77 + 88 \times 166{\cdot}18 + 160 \times 165{\cdot}46 + 167 \times 163{\cdot}46}{90 + 88 + 160 + 167}$$

$$= 165{\cdot}171.$$

第 18 表

計測地方	n	M
咸　北	90	166·77
咸　南	88	166·18
平　北	160	165·46
平　南	167	163·50
計	505	165·171

しかし表を見るに,平安南道及び北道に於ける觀測總數は他道に比し多いので,上式の様な平均の出し方は,必ずしも,北鮮に於ける朝鮮人の眞の平均身長を表示するとは云はれない. 各道の人口が同一なりと見なして,

$$M = \frac{166{\cdot}77 + 166{\cdot}18 + 165{\cdot}46 + 163{\cdot}50}{4} = 165{\cdot}475$$

とするを以て寧ろ正しとす. 一層,正しくは重みづけられたる平均を見よ.

定理 IV. N 對の觀測値

$$X_1,\ X_2,\ X_3,\ \cdots\cdots,\ X_N \quad と$$

$$Y_1,\ Y_2,\ Y_3,\ \cdots\cdots,\ Y_N \quad と$$

あり. 各對の觀測値の和又は差,

$$X_1 \pm Y_1,\ X_2 \pm Y_2,\ X_3 \pm Y_3,\ \cdots\cdots,\ X_n \pm Y_n$$

を作れば,その算術的平均は X の算術的平均 $M_{(x)}$ と, Y の算術的平均 $M_{(y)}$ との和又は差に等しい. 即ち,

$$M_{(x \pm y)} = M_{(x)} \pm M_{(y)} \quad \cdots\cdots\cdots\cdots\cdots (\text{II. }9)$$

である.

證明.

$$\Sigma(X \pm Y) = \Sigma(X) \pm \Sigma(Y) = NM_{(x)} \pm NM_{(y)},$$

$$\therefore \qquad M_{(x \pm y)} = M_{(x)} \pm M_{(y)}$$

である. 即ち定理を證明し得たり.

觀測數列が二個以上一組となれる場合も,亦同様である.

$$M_{(x \pm y \pm z \pm \cdots)} = M_{(x)} \pm M_{(y)} \pm M_{(z)} \pm \cdots \quad \cdots\cdots (\text{II. }10)$$

これらの諸定理は常に用ひられ,亦重要なものである.

第三項 中央値

9. 中央値の定義

觀測値をその大さの順に並べる時,その中央の觀測値を中央値 (Median; Der Medianwert, Der Zentralwert) と云ふ. 例へば身長の中央値を求めるには,計測する個體をその身長の順にならべ,

その中央にあたる個體の身長を以て,中央値とする. 若し,觀測總數 N が奇數即ち $2m+1$ なる時は小なる方より,又は大なる方より計へて,第 $(m+1)$ 番目にあたる個體の身長が中央値となる. 又,偶數 $2m$ ならば,第 m 番目と第 $(m+1)$ 番目の個體の身長の中間にある値である. 故に,若し中央値のみを以て滿足する場合ならば,變數の大さの順にならべて,其の中唯一個又は二個體を計測すればいいので,全部の計測を行ふ必要はない.

觀測値が度數分布表となつて居る時は,中央値の計算はやや複雑であつて,次の通りの方法で計算する.

10. 中央値の計算法

例 6. 第19表に示した朝鮮人頭骨全顔面角の度數分布表より,中央値を計算せよ.

$N = 93$, 奇數であるから,初めより數へて $\dfrac{93}{2} = 46{\cdot}5$ を超えた所,即ち47番目にあたる個體の全顔面角が中央値である. 然るに度數分布表が與へられて有るのみである. 角度の小なる方より順次數へ上げて行く(これを**累積度數**と云ふ).

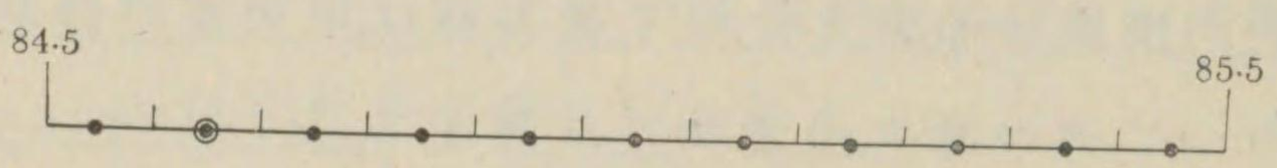
表に累積度數を順次記入すれば,第三列の通り, 77·5 以下の度數 2, 78·5 以下の度數 5, 79·5 以下の度數 9, ····, 85·5 以下の度數 56 となり,この時豫定の47を通り越すが故に,第

第 19 表

中央の計算法

階級	度數	累積度數
76·5—77·5	2	2
77·5—78·5	3	5
78·5—79·5	4	9
79·5—80·5	6	15
80·5—81·5	8	23
81·5—82·5	7	30
82·5—83·5	9	39
83·5—84·5	6	45
84·5—85·5	11	56
85·5—86·5	6	
86·5—87·5	10	
87·5—88·5	9	
88·5—89·5	5	
89·5—90·5	6	
90·5—91·5	—	
91·5—92·5	1	

47番目の觀測値は84·5と85·5との間になければならぬ．而して,階級84·5—85·5には11個の觀測値があり,第47番目の觀測値は其のうち第2番目の觀測値であることも直ちに知られる,この11個のうち第2番目の觀測値を知るには,11個が等間隔を保つて居ると假定する．即ちλを階級幅とすれば,$\frac{\lambda}{11}$なる小區劃に分ち,各觀測値は$\frac{\lambda}{11}$だけの差を保つて分布して居ると假定するのである(第17圖).



第17圖　一階級内における觀測値の分布を示す假想圖.

然らば,圖により第2番目の觀測値は$84·5+\frac{\lambda}{11}\times 1·5$であることが明に知られるであらう．但しここでは$\lambda=1$である.

次の通り計算する.

$$\frac{N}{2}=\frac{93}{2}=46·5,$$

$$\text{Med.}=84·5+\frac{1}{11}\left(\frac{93}{2}-45\right)=84·636.$$

一般には次の公式による．累積度數を順次$\frac{N}{2}$まで計算し,第$\frac{N}{2}$番の個體が存在する階級の下限をα_1,上限をα_2,その階級の度數をf,階級幅をλ,又α_1までの累積度數をzとすれば,Med.は

$$\text{Med.}=\alpha_1+\frac{\lambda}{f}\left(\frac{N}{2}-z\right).\quad\cdots\cdots\cdots\cdots(\text{II. }11)$$

この計算法はNが奇數でも,偶數の場合でも同樣に,實行することが出來る.

例 7.　第15表に於ける日本人男學生6,000人の頭長幅指數のMed.は,次の通りに計算することが出來る(第15表第5列)．順次累積度數を計へ,79—80の階級まで數へ上げたる度數は2,527である．次の階級の度數を加ふれば3,000を超えてしまふ故に,中央値は80—81なる階級になければならぬ．この階級には709人あり,故に階級幅を709等分して,$\left(\frac{6000}{2}-2527\right)\times\frac{1}{709}$を求める．即ち$\frac{473}{709}=0·667$.

$$\text{Med.}=80+0·667\fallingdotseq 80·67.$$

11.　中央値の圖計算法

中央値の計算を上記の如く比例部分的に行ふ代りに,ミリメーター紙を用ひて,圖計算を行つてもいい.

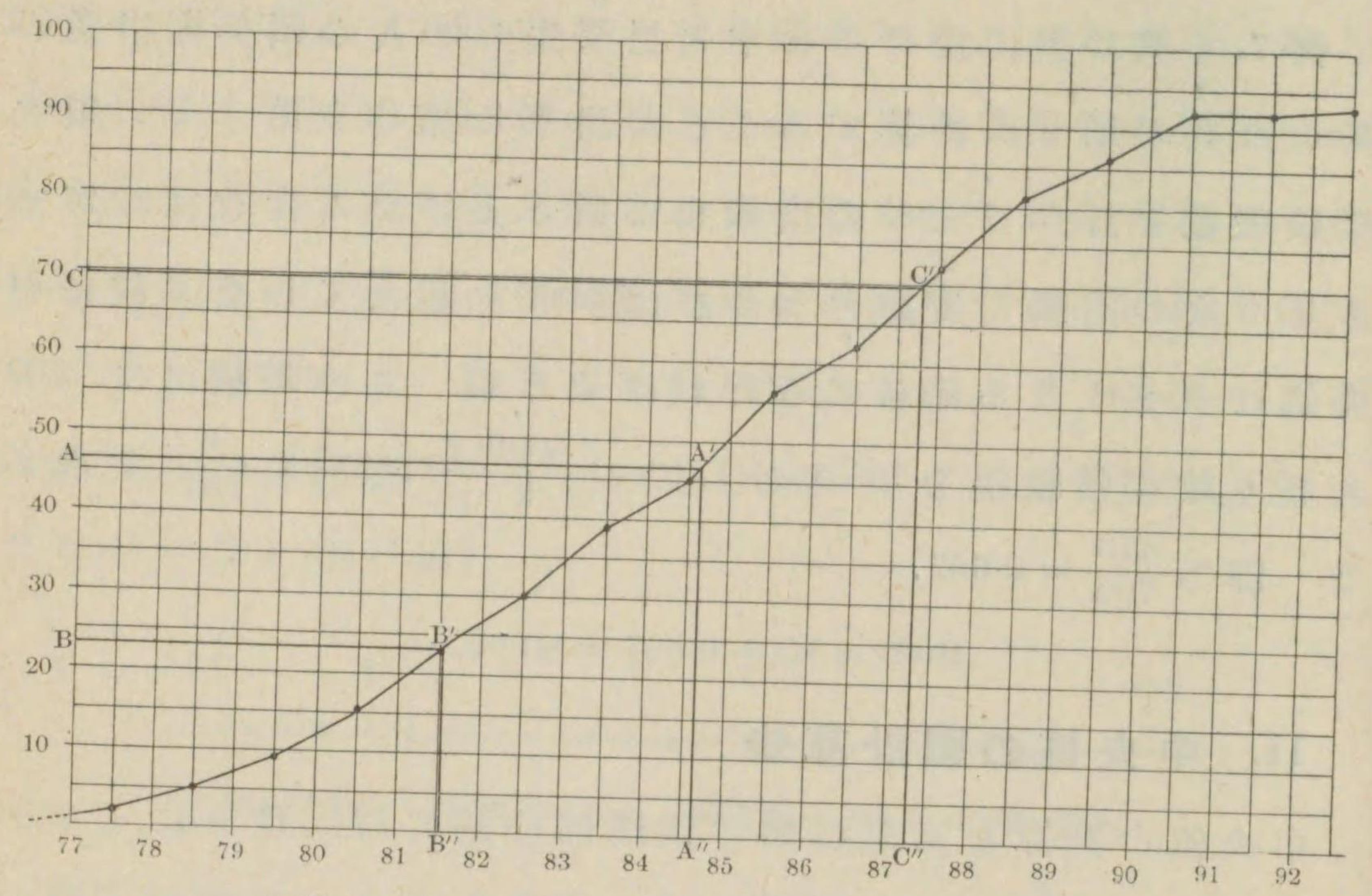
例 8.　例6の頭骨全顏面角の中央値を圖計算で求めよ(第18圖).

(i)　ミリメーター紙に横線Xを取り,その上に階級を適當に取る,第17圖では一階級を2cm.に取つた(數字を記入した所は階級の中央の値である)．次にY軸を定め,

(ii)　階級境界(階級の中央にあらず)に垂線を取り,その上にそれぞれ累積度數を取り,圖の如く**累積曲線**$B'A'C'$を畫く.

(iii)　次に$\frac{N}{2}$をY軸に取り,この點をAとす．AよりX軸に並行線AA'を引き,累積曲線とA'で交らしむ．更にA'よりY軸に並行なる線$A'A''$を引き,X軸とA''に交らしむ.

(iv)　A''の位置に於ける目盛を讀む,84·65,これ中央値で例6の計算結果と略一致す．並行線を引くことは,ミリメーター紙では容易であるが,必ずしも鉛筆で線を引く必要はない．Aより



第 18 圖 中央値及び四分の一位の計算圖 (例 8).

A′まで直線に沿ひて眼を動かすだけで,A′を決定することが出來,A″も亦同様である. この方法は中央値のみならず,四分の一位,百分法の計算にも用ひうる.

例 9. 例 2 日本人男學生の頭長幅指數の分布表より, Med. を精密に計算せよ.

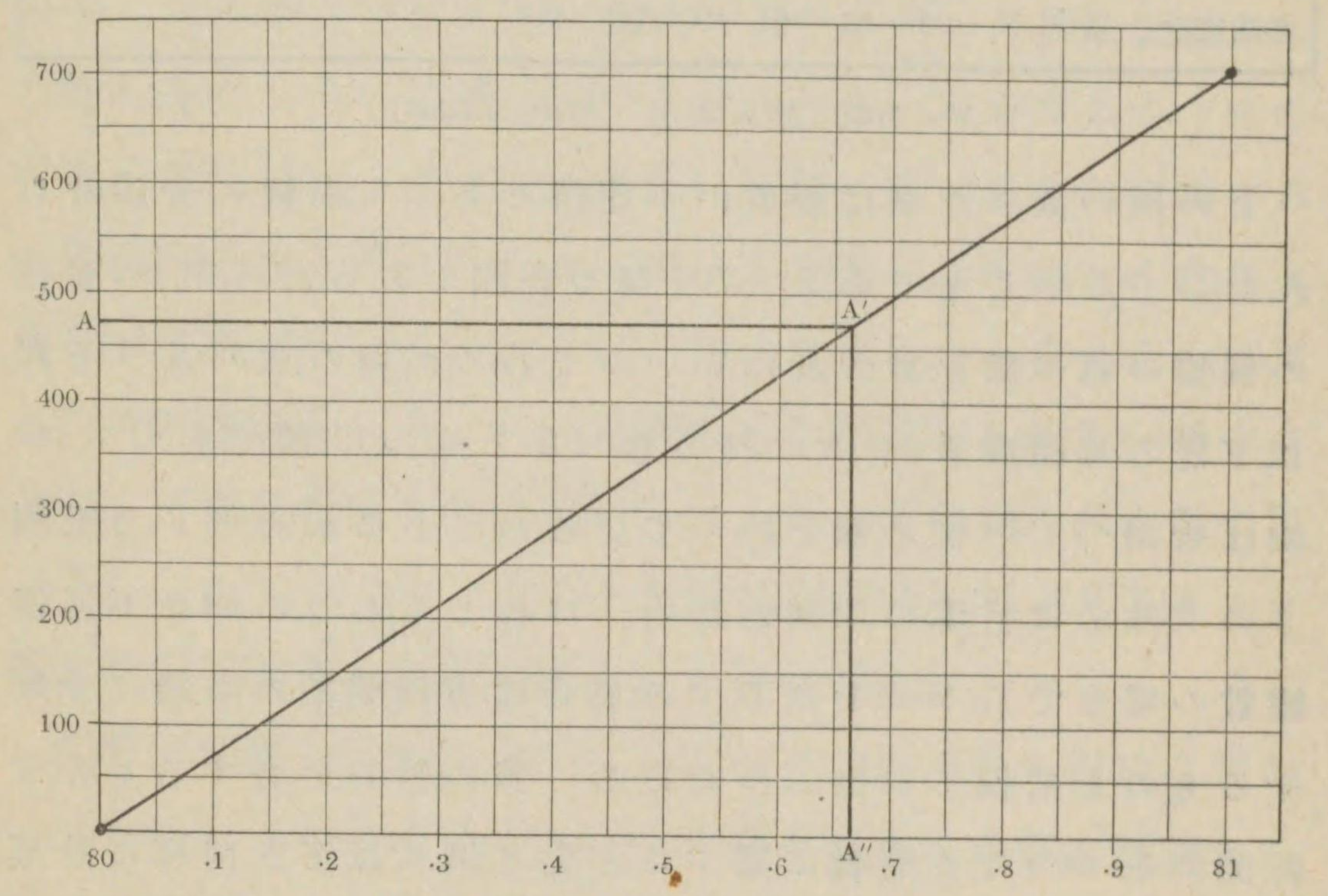
中央値の精密な値を圖計算で算出せんとすれば,勢ひ尨大な方眼紙を要し,不便甚しい. かくの如き時は必要な所だけを,圖計算で行ふ様にすればいい.

例 7 に計算を行つた如く先づ累積度數を求める. 指數 80 以下の累積度數 2,527 となつた. 然るに,81 以下の累積度數を求むれば $\frac{N}{2}$ を超過するが故に,中央値は區間 80—81 の何らかの値を取らなければならぬ. これを求めるのに圖計算を行ふのである.

(i) 先づ,横線に 80 より 81 までの値を精密に取る(第 19 圖).

(ii) 階級 80—81 の度數 709 なるが故に,81 の垂線上に 709 を取り,この點と 80 上の 0 と結ぶ直線を書く.

(iii) 次に, $\frac{6000}{2} - 2527 = 473$ を縦線上に取り A とす, X 軸に並行線, AA′ を畫き,累積曲線と A′ にて交らしむ.



第 19 圖 中央値の計算圖 (例 9).

(iv) 次に, A′ より Y 軸に並行線 A′A″ を描き, X 軸と交る點を A″ とし,點 A″ の目盛 80·67 を讀み,中央値とす. (また一層精密に略 80·667 を讀むことが出來るが,第 19 圖は縮圖のため mm. 以下を省略したから,漸く 80·67 を讀みうるのみである.)

かく計算せる中央値は,例 7 の算術的方法と同一程度の精密さがある.

例 10. Butten (鰈の一種)につき,尾鰭の棘の數を計へて,次の分

布を得た．これよりMed.を計算せよ．

第 20 表

Buttenの尾鰭の棘數の分布（Med.の計算）

鰭の棘數	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61
度　數	5	2	13	23	58	96	134	127	111	74	37	16	4	2	1
累積度數	5	7	20	43	101	197	331	458							

$N=703$,　$M=53{\cdot}67$,　Med. $=53{\cdot}66$.

この例の異つた點は,變數が不連續であつて,階級に分類せられたのでないことである．703個の觀測であるから第352番目の個體の棘の數を計へればいいのであるが,棘の數53及びそれ以下なる累積度數331,次に棘の數54なるもの127體であつて,127個は皆揃つて54個の棘を持つて居るのであるから,例6乃至例9の方法では計算の仕樣がない．けれども,かくの如き不連續變數の場合でも,Med.を計算する場合は,連續變數の階級に分類せるものと同樣に取扱ふのである．棘の數54を有するものは區間53·5—54·5なる階級に屬すると定め例6以下と同樣に計算するのである．Med.の値は53·66となつた．

12. 中央値の特性

中央値は平均として必要な多くの條件を充たして居る．

確定的に計算することが出來,理解に便利で,計算極めて容易である．又すべての觀測値をもととして居る．何となれば一個の觀測値を增加すれば,極めて僅微ではあらうが,中央値はこの影響を受けて變動するからである．けれども全體の個體の計測は必ずしも必要でない．一列にならべた觀測値中,中央の

大さを有するものを計測すればそれがMed.である．又,誤差又は過失により,甚しく大なる又は小なる觀測値が加はつても,中央値の變化すること極めて僅かであつて,偏差小なる一個體の加はつたと同一の結果をもたらすのみである．

算術的平均に比して著しく都合のいい點は,計測値に無限大の有る場合(例へば或る面の曲率半徑を計測して度數分布となす時,場合によつて平面又はそれに甚だ近い場合がある．その時曲率半徑は∞となる)又は,階級に分類されて居るが,兩端にある階級は境界が明にせられて無い場合(第一章例4),かくの如き場合には,算術的平均の計算が不可能であるから,中央値を計算した方が都合がいい．かく算術的平均の用ひられない場合は勿論,その他,算術的平均を補助すると云ふ意味で,これと併せ用ひられることがある．殊に,不相稱型分布に於て,MとMed.とを併せ用ひて便利なことがある(第四章參照)．

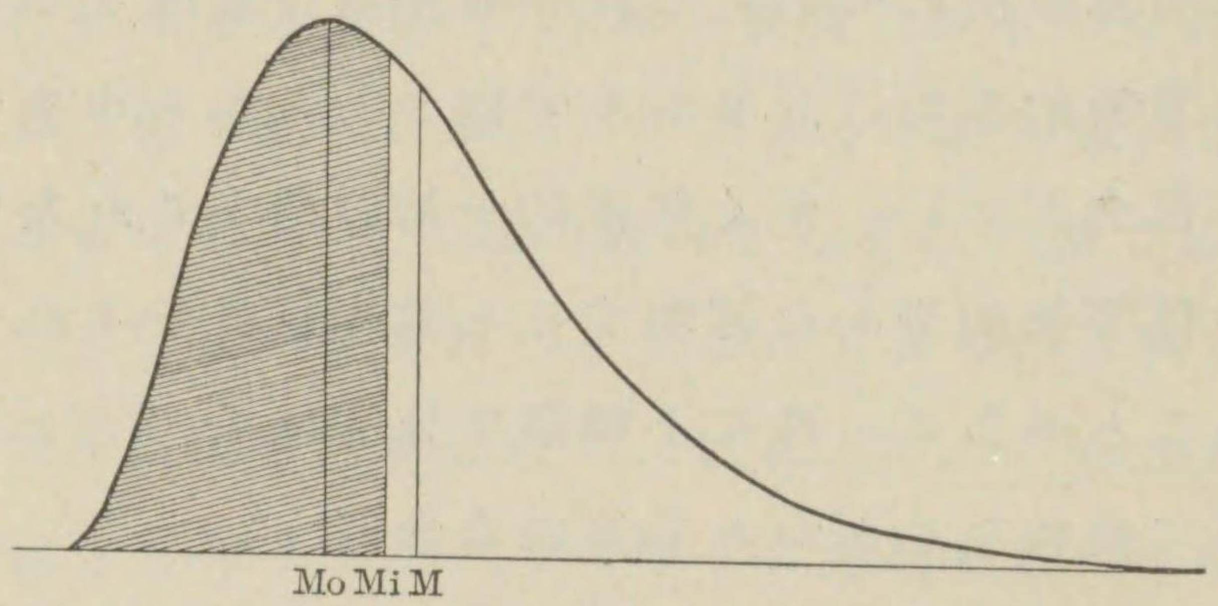
度數分布圖形に於て,Med.に相當して一の垂線を引けば,圖形の面積(度數多角形又はHistogramの總面積,或は曲線と横軸との間の面積)はこの垂線によつて等分せらる．即ち,第20圖に於ける陰影を施せる部分の面積は,然らざる部分の面積に等しい．これは中央値の定義より容易に證明しうるのである．しかし中央値の代數的取扱が出來ないことは,救ふべからざる短所であつて,上記の長所があるに拘らず,一般に廣く用ひられない所以である．

第四項 形 態

13. 形態の定義

定義. 形態 (Mode; Die Mode, Der Modalwert) とは分布密度の最大なる變數の値を云ふ.

この定義は極めて抽象的だから,一寸理解に困難であらうが,分布を度數曲線に表はした時,その曲線の最高點(極大値)に相當する變數の値である. 第20圖 Mo が形態である. Mode とは流



第20圖 不相稱なる度數分布における三つの平均の關係を示す.
(Mo は形態, Mi は中央値, M は算術的平均.)

行の意であるから,流行値と稱して居る人もある.

形態の定義は屢誤り傳へられて居る. 形態とは觀測値の最も流行する値ではあるが,最も流行する階級中央の値では決してない. 故に,度數分布表を作つて,その最大度數の階級の中央の値を以て,形態を定義する人があるが,これは誤りである. 同一觀測數列から度數分布表を作るに,階級の幅や階級の境界の値は各自任意に決定しても差支はない. 今この幅や境界の値を變更すれば,最大度數の階級の中央の値もこれに從つて變動

するものである. かく任意に變更しうる値が形態ならばこれは平均として價値のないものとなる(平均の備ふべき要件の(i)參照). 形態はこんな値ではない.

例 11. Johannsen は豆の一種 (Schwarze belgische Kruppbohnen) の種子1,522個を計測し,階級の境界を二樣に選んで,度數分布表を作製したのが,第21表甲及び乙である. 最大度數の階級中央の値は甲表では 12·50, 乙表では 12·25 である. この二つの値に可なり大なる差異があるが,これは計算の誤りから出來たものではなく,階級境界の選擇の相違から出來た差異である. 階級幅が大になればこの差益甚しくなる. これを見ても諸書に散見

第 21 表

Schwarze belgische Kruppbohnen の長徑の度數分布表.

甲 表

階級	度數
9—10	7
10—11	67
11—12	466
12—13	761
13—14	201
14—15	15
15—16	4
16—17	1

$N = 1522$, $M = 12{\cdot}25$, $\sigma = 0{\cdot}767$, Med. $= 12{\cdot}29$, Mo $= 12{\cdot}37$.

乙 表

階級	度數
8·75—9·75	2
9·75—10·75	43
10·75—11·75	314
11·75—12·75	809
12·75—13·75	316
13·75—14·75	30
14·75—15·75	6
15·75—16·75	2

$N = 1522$, $M = 12{\cdot}25$, $\sigma = 0{\cdot}767$, Med. $= 12{\cdot}25$, Mo $= 12{\cdot}25$.

する様に,最大度數の階級の中央の値を以て形態とするのは誤りであることが判明するであらう.

吾々の知りたい形態はかくの如き不定の値ではなく,各階級の幅を無限に小にし,各觀測數を無限に多く取りたる時に,最も屢出現する階級の中央の値である. しかし無限に多く觀測することも,階級幅を無限に小にすることも不可能であるから,與へられた度數分布表より,一定の法式で理論的の形態を計算するのが普通である. その嚴密な方法はやや高等なる數學の知識を要するので,續編に述べることとし,簡單なる形態の計算法を次に述べることとする.

14. 形態の近似計算法

輕度の不相稱なる度數分布に於ては,

$$\mathrm{Mo} = M - 3(M - \mathrm{Med.}),$$

なる式が成立する (Pearson).

これより

$$\mathrm{Mo} = 3\mathrm{Med.} - 2M. \quad \cdots\cdots\cdots\cdots\cdots (\mathrm{II.}\ 12)$$

この公式より計算せる Mode は,正確に計算せる Mode とは,可なりよく一致する.

例 12. 第16表(例 3)に於ける牛乳搾取量の度數分布より Med. を計算し,これより間接に Mo を計算せよ.

表の第 5 列は累積度數で,これより Med. は例 6 以下の通り計算せらる.

$$\mathrm{Med.} = 642{\cdot}70,$$

$$M = 654{\cdot}36,$$

$$\mathrm{Mo} = 3\mathrm{Med.} - 2M = 619{\cdot}38.$$

形態 Mo を M と Med. とより計算するならば, Mo は第二次的意義を有するのみとなり,重大視せられなくなる. 故に貴重なる研究に於て不相稱なる場合は Mo の嚴密なる計算を必要とする. 又形態は偶然の影響を受けて値の狂ふこと甚しく,又分布の性狀を表はすに甚しく鈍感である. 例へば第22表の分布 I と II とは全く分布の內容を異にせるに拘らず,同一に近い形態を有して居る. 斯く種々なる不都合な點があるので, Med. の如く一般的には用ひられないが,不相稱型の理論的硏究には必要なものである.

第 22 表

同一に近い形態を有し且つ內容の著しく相違せる二分布の假定例

階級	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}
分布 I	1	3	10	15	7	4	2							
分布 II	1	7	11	15	14	13	12	10	9	7	5	3	2	1

算術的平均 M と,中央値 Med. と,形態 Mo との間には次の關係がある.

(i) 相稱の分布型では,中央値及び形態は算術的平均に一致する.

(ii) 不相稱型では(J-型分布を含む), Mo は曲線の最も高き所(極大)に相當する變數の値で, Med. はこれより緩傾斜の脚に偏し, M は同一脚にて一層脚端に偏在して居る(第20圖).

(iii) 輕度の不相稱型では,次の公式が成立する(Pearson).

$$M - \mathrm{Mo} = 3(M - \mathrm{Med.}). \cdots\cdots\cdots\cdots (\mathrm{II.}\ 12)$$

第五項　幾何的平均

15. 幾何的平均

定義. N 個の觀測值の**幾何的平均**(Geometric mean; Der geometrische Mittelwert)とは觀測值の全ての積の N 乘根である.

觀測值を $V_1, V_2, V_3, \cdots, V_N$ とすれば,幾何的平均 G は

$$\left.\begin{aligned} G &= \sqrt[N]{V_1 . V_2 . V_3 . \cdots . V_N} \\ &= (V_1 . V_2 . V_3 . \cdots . V_N)^{\frac{1}{N}} \end{aligned}\right\} \cdots\cdots\cdots (\mathrm{II.}\ 13.\ \mathrm{a})$$

又對數を以て表はす時は,幾何的平均の對數は各觀測值の對數の算術的平均である(對數の底は任意).

$$\left.\begin{aligned} &\text{即ち,}\quad \log G = \frac{1}{N}\left\{\log V_1 + \log V_2 + \log V_3 + \cdots \log V_N\right\}, \\ &\text{又は}\quad \log G = \frac{1}{N}\sum(\log V). \end{aligned}\right\}(\mathrm{II.}\ 13.\ \mathrm{b})$$

もし度數分布表より計算するならば,次の如くである.

$X_1, X_2, \cdots, X_K$ を階級中央の値とし,

$f_1, f_2, \cdots, f_K$ をその度數とすれば,

$$\left.\begin{aligned} G &= \sqrt[N]{X_1^{f_1} . X_2^{f_2} . X_3^{f_3} . \cdots . X_K^{f_K}} \\ \text{又は}\quad \log G &= \frac{1}{N}\{f_1 \log X_1 + f_2 \log X_2 + \cdots\}, \\ &= \frac{1}{N}\sum_{i=1}^{K}(f_i \log X_i), \end{aligned}\right\} (\mathrm{II.}\ 13.\ \mathrm{c})$$

$$\text{但し}\quad N = \sum_{i=1}^{K} f_i .$$

計算には對數の式を用ふ.

16. 幾何的平均の性質

幾何的平均は代數的取扱が出來る點で至極便利であるが,一般的に平易でなく,理解しがたく,餘りに理論的,抽象的であつて,平均値として適當でない. 又,觀測値中に一個でも0が有る時は,他の觀測値の如何に拘らず,幾何的平均は0となり; 觀測値のうちに奇數個の負數がある時は,幾何的平均は實數として計算することが出來ない. かく,不都合があるのみならず,觀測値の中に並外れて大なる値,又は小なる値が有る時は,その爲に幾何的平均は變化を受くることが大であるから,普通用ひられない. しかし人口の如く,幾何級數的に增加するものには,この平均を用ひて便利なことがある.

幾何的平均の性質中最重要なるは,次の如く代數的取扱の可能なることである.

(i) K 群の觀測値が有り,その各の幾何的平均が計算せられてある時,これらを合同せる時の幾何的平均は次の如く計算せらる.

各群の觀測數を　$N_1, N_2, N_3, \cdots, N_K$,

又その幾何的平均を　$G_1, G_2, G_3, \cdots, G_K$,

これを合同せる時の幾何的平均, G は,

$$\left.\begin{aligned} G &= \sqrt[N_1+N_2+\cdots]{G_1^{N_1} . G_2^{N_2} . G_3^{N_3} . \cdots . G_K^{N_K}}, \\ \text{又}\quad \log G &= \frac{1}{N_1+N_2+\cdots+N_K}(N_1 \log G_1 + N_2 \log G_2 \\ &\quad + \cdots N_K \log G_K), \\ \text{又}\quad &= \frac{\sum_{i=1}^{K}(N_i \log G_i)}{\sum_{i=1}^{K} N_i}. \end{aligned}\right\} (\mathrm{II.}\ 13.\ \mathrm{d})$$

(ii) 觀測數列 $X_1, X_2, \cdots, X_N$ ありて,幾何的平均 G_1,

又, $Y_1, Y_2, \cdots, Y_N$ の幾何的平均 G_2 ありとす.

今指數 $I_1 = \dfrac{X_1}{Y_1} \times 100,\ I_2 = \dfrac{X_2}{Y_2} \times 100,\ \cdots\cdots,\ I_N = \dfrac{X_N}{Y_N} \times 100$ をとれば指數の幾何的平均, G_1 は

$$G_1 = \frac{G_x}{G_y} \times 100$$

である.

<u>證明.</u>

$$\sum_{i=1}^{N} (\log I_i) = N \log 100 + \sum_{i=1}^{N} \log X_i - \sum_{i=1}^{N} \log Y_i,$$

各項を N にて除す時は,

$$\log G_1 = \log 100 + \log G_x - \log G_y$$

$$= \log\left(\frac{G_x}{G_y} \times 100\right).$$

即ち $G_1 = 100 \dfrac{G_x}{G_y}.$

100の代りに他の常數を以てするも同一である.

(iii) 三觀測數列 X, Y, Z あり,指數 $\dfrac{Y}{X}$ の幾何的平均 $G_{\frac{y}{x}}$ 又 $\dfrac{Z}{X}$ の幾何的平均 $G_{\frac{z}{x}}$ とす.

然る時は
$$\frac{G_{\frac{y}{x}}}{G_{\frac{z}{x}}} = G_{\frac{y}{z}}$$

である. 證明は(ii)に準ずれば容易である讀者これを試みられよ.

(iv) h 個の觀測群あり,各群 m 個の觀測數を有す. 今,各觀測群より相應せる觀測値を1個づつ取り出してその積を作る時は,その積よりなる數列の幾何的平均, G は

$$G = G_1 . G_2 . \cdots\cdot . G_h$$

である. 但し $G_1, G_2, G_3, \cdots, G_h$ は各群の觀測値の幾何的平均である.

<u>證明.</u>

第一群の觀測値, $V_1', V_2', V_3', \cdots, V_m'$,

第二群の觀測値, $V_1'', V_2'', V_3'', \cdots, V_m''$,

……………………………………

第 h 群の觀測値, $V_1^{(h)}, V_2^{(h)}, V_2^{(h)}, \cdots, V_m^{(h)}$,

相對應せる觀測値を一個づつ取り出してその積を作る時は,その積は次の通りとなる.

$$V_1' . V_1'' . V_1''' . \cdots\cdot . V_1^{(h)},$$

$$V_2' . V_3'' . V_2''' . \cdots\cdot . V_2^{(h)},$$

……………………

$$V_m' . V_m'' . V_m''' . \cdots, V_m^{(h)},$$

これ等の積の幾何的平均は

$$\sqrt[m]{V_1' . V_2' . \cdots\cdot . V_m' . V_1'' . V_2'' . \cdots\cdot . V_m'' \cdots\cdot . V_1^{(h)} . V_2^{(h)} . \cdots\cdot . V_m^{(h)}}$$

$$= G_1 . G_2 . G_3 . \cdots\cdot . G_h .$$

(v) **幾何的平均と算術的平均との關係**[1]

<u>觀測値の偏差 x が算術的平均 M に比して著しく小なる時は,幾何的平均 G は近似的に</u>

$$G \fallingdotseq M\left(1 - \frac{\sigma^2}{2M^2}\right) \cdots\cdots\cdots\cdots\cdots\cdots (\text{II. } 14)$$

<u>とすることが出來る.</u>

[1] この節は標準偏差の項を讀んで後にすべし.

證明. 定義により

$$\log G = \frac{1}{N}\{\log V_1 + \log V_2 + \cdots\cdot \log V_n\},$$

但し log は今は自然對數を取る.

$V_1, V_2, \cdots, V_N$ の算術的平均は M であるから, M よりの各觀測値の偏差を $x_1, x_2, \cdots\cdot$ とすれば,

$$N\log G = \log(M + x_1) + \log(M + x_2) + \cdots\cdot + \log(M + x_N)$$

$$= N\log M + \log\frac{M + x_1}{M} + \log\frac{M + x_2}{M} + \cdots\cdot + \log\frac{M + x_N}{M}$$

$$= N\log M + \log\left(1 + \frac{x_1}{M}\right) + \log\left(1 + \frac{x_2}{M}\right) + \cdots\cdot + \log\left(1 + \frac{x_N}{M}\right).$$

しかるに, $\log\left(1 + \frac{x_1}{M}\right)$ 以下に於ける $\frac{x}{M}$ は何れも1より小であるから,これを $\frac{x}{M}$ の級數に展開することが出來る. 故に,

$$\log G = \log M + \frac{1}{N}\left\{\frac{x_1}{M} - \frac{x_1^2}{2M^2} + \frac{x_1^3}{3M^3} - \cdots\cdot\right.$$

$$+ \frac{x_2}{M} - \frac{x_2^2}{2M^2} + \frac{x_2^3}{3M^3} - \cdots\cdot$$

$$+ \frac{x_3}{M} - \frac{x_3^2}{2M^2} \cdots\cdots\cdots\cdots$$

$$\left.\cdots\cdots\cdots\cdots\cdots\cdots\cdots\right\}$$

即ち, $$\log G = \log M + \frac{1}{N}\left\{\frac{\Sigma x}{M} - \frac{\Sigma x^2}{2M^2} + \frac{\Sigma x^2}{3M^3} - \cdots\cdot\right\}.$$

$\Sigma x = 0$ であり, $\Sigma x^2 = N\sigma^2$ である.

而して $\frac{1}{N}\left(\frac{\Sigma x^3}{3M^3} - \cdots\cdot\right)$ を ε にてあらはすに, x が M に比し甚しく小なる時は, ε は無視することが出來る値である.

即ち $$\log G = \log M - \frac{\sigma^2}{2M^2} + \varepsilon$$

$$= \log\left\{M.\, e^{\left(-\frac{\sigma^2}{2M^2} + \varepsilon\right)}\right\},$$

$$G = M.\, e^{-\frac{\sigma^2}{2M^2} + \varepsilon}.$$

而して, $e^{-\frac{\sigma^2}{2M^2} + \varepsilon}$ は $\left(-\frac{\sigma^2}{2M^2} + \varepsilon\right)$ の冪級數に展開することが出來る.

$$e^{-\frac{\sigma^2}{2M^2} + \varepsilon} = 1 - \frac{\sigma^2}{2M^2} + \varepsilon + \frac{1}{2!}\left(-\frac{\sigma^2}{2M^2} + \varepsilon\right)^2 + \cdots\cdot$$

$$= 1 - \frac{\sigma^2}{2M^2} + \varepsilon + \frac{\sigma^4}{8M^4} - \frac{\sigma^2}{2M^2}\varepsilon + \frac{\varepsilon^2}{2} + \cdots\cdot,$$

ε は無視していい程小である,又 $\frac{\sigma^4}{8M^4}$ 以下も無視していい. 故に,

$$G \fallingdotseq M\left(1 - \frac{\sigma^2}{2M^2}\right). \quad \cdots\cdots\cdots\cdots\cdots\cdots(\text{II. }14)$$

系. 幾何的平均は常に算術的平均よりも小なり.

何となれば $\frac{\sigma^2}{2M^2}$ が正であるから.

(II. 14)式及び其の系は, $\frac{x}{M}$ が何れの x に對しても1より小なる場合であり,且つ $\frac{\Sigma x^3}{M^3}$ 又は高次の項は小にして0と見なして差支なき場合にのみ,適用せらるることを忘れてはならぬ. M が甚小なる時は用ふることが出來ないのである.

第六項 調和平均

17. 調和平均の定義

定義. 各觀測値 $V_1, V_2, V_3, \cdots, V_N$ の逆數の算術的平均の逆數を**調和平均** (Harmonic mean; Das harmonische Mittel) と云ふ.

即ち H を調和平均とすれば,

$$
\left.\begin{aligned}
H &= \frac{1}{\frac{1}{N}\left(\frac{1}{V_1}+\frac{1}{V_2}+\cdots\cdot+\frac{1}{V_N}\right)}, \\
\text{或は,}\quad \frac{1}{H} &= \frac{1}{N}\left(\frac{1}{V_1}+\frac{1}{V_2}+\cdots\cdot+\frac{1}{V_N}\right) \\
&= \frac{1}{N}\sum\left(\frac{1}{V}\right).
\end{aligned}\right\}\cdots\cdots(\text{II. }15)
$$

物價の統計などでは屢調和平均が用ひられる．例へば米1升の價を時期を異にして觀測し,圓を以て表はし $V_1, V_2, V_3, \cdots\cdot, V_n$ とす．今その調和平均を求むるに $\frac{1}{V_1}+\frac{1}{V_2}+\cdots\cdot+\frac{1}{V_n}$ の各項は1圓につき米幾升を買ひうるかを表示して居るから，$\frac{1}{H}$ は1圓につき買ひうる米の量の平均である．生物統計學的には,1分間に於ける脈搏數を $V_1, V_2, \cdots\cdot, V_n$ であらはす時,一脈搏に要する平均時間は調和平均の逆數である．故に,かくの如き場合調和平均を出すのは必ずしも無意義ではない．しかし,斯く特別の場合以外は用ひられることがない．

18. 調和平均と算術的平均との關係[1]

$$V_1 = M + x_1,\quad V_2 = M + x_2,\quad V_3 = M + x_3,\quad \cdots\cdot$$

であるから

$$
\begin{aligned}
\frac{1}{H} &= \frac{1}{N}\left\{(M+x_1)^{-1}+(M+x_2)^{-1}+\cdots\cdot\right\} \\
&= \frac{1}{MN}\left\{\left(1+\frac{x_1}{M}\right)^{-1}+\left(1+\frac{x_2}{M}\right)^{-1}+\cdots\cdot\right\}.
\end{aligned}
$$

$\frac{x_1}{M}$, $\frac{x_2}{M}$, $\cdots\cdot$ が何れも1に比し小である時は,この括弧の中は級數に展開することが出來る．即ち,

$$
\frac{1}{H} = \frac{1}{MN}\left\{1-\frac{x_1}{M}+\frac{x_1^2}{M^2}-\frac{x_1^3}{M^3}+\cdots\cdot\right.
$$

(1) この節も標準偏差を理解して後に讀むべし．

$$
\begin{aligned}
&+1-\frac{x_2}{M}+\frac{x_2^2}{M^2}-\frac{x_2^3}{M^3}+\cdots\cdot \\
&+1-\frac{x_3}{M}+\frac{x_3^2}{M^2}-\frac{x_3^3}{M^3}+\cdots\cdot \\
&\left.\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\right\} \\
&= \frac{1}{NM}\left\{N-\frac{\Sigma x}{M}+\frac{\Sigma x^2}{M^2}-\frac{\Sigma x^3}{M^3}+\cdots\cdot\right\}
\end{aligned}
$$

$\frac{1}{N}\Sigma x = 0$, $\frac{1}{N}\Sigma x^2 = \sigma^2$ にして $\frac{\Sigma x^3}{NM^3}$ 以下の項はこれを ε にて表はす．ε は $\frac{x}{M}$ 甚小なる時は省略することが出來る數である．

$$\frac{1}{H} = \frac{1}{M}\left(1+\frac{\sigma^2}{M^2}+\varepsilon\right),$$

$$H = M\left(1+\frac{\sigma^2}{M^2}+\varepsilon\right)^{-1}.$$

$\left(1+\frac{\sigma^2}{M^2}+\varepsilon\right)^{-1}$ も亦無限級數に展開することが出來る,故に

$$H = M\left\{1-\frac{\sigma^2}{M^2}-\varepsilon+\left(\frac{\sigma^2}{M^2}+\varepsilon\right)^2-\cdots\cdot\right\} = M\left(1-\frac{\sigma^2}{M^2}+\varepsilon'\right)$$

ε' は $\frac{x}{M}$ の三乘羃以上の項よりなる故に $\frac{x}{M}$ が甚小なる時は ε' は0と見なすことが出來る．故に

$$H \doteqdot M\left(1-\frac{\sigma^2}{M^2}\right)$$

定理． 算術的平均に比して觀測値の偏差 x が著しく小なる時は,調和平均 H は

$$H = M\left(1-\frac{\sigma^2}{M^2}\right)\quad\cdots\cdots\cdots\cdots\cdots\cdots(\text{II. }16)$$

である．但し,この式も各偏差は皆算術的平均 M より小であること,又 $\frac{\Sigma x^3}{M^3}$, $\frac{\Sigma x^4}{M^4}$ 以下が甚だ小であつて,0と假定して差支ない場合にのみ用ひられる．若し然らざる時は,公式(II. 16)が成立しない．

第七項 重みづけられたる平均

19. 重みづけられたる平均の定義

重みづけられたる平均は算術的平均の更に一般的なる平均である. 例をあげて說明すれば,或る學校で點數の平均を算出するに數學の得點を 3 倍し,語學の得點を 2 倍し,國語,漢文,物理,化學の點をそれぞれ 1 倍し,これらの總和を 9 で除したとすれば,これは單なる算術的平均でない. かくの如き平均は**重みづけられたる平均** (Weighted mean) である.

定義. 觀測値の各に或る係數を乘じ,その積の和を係數の和を以て除したるものを,重みづけられたる平均と云ふ.

係數を $w_1, w_2, w_3, \cdots, w_n$ とすれば,重みづけられた平均, M' は

$$M' = \frac{w_1 V_1 + w_2 V_2 + \cdots w_n V_n}{w_1 + w_2 + \cdots + w_n} = \frac{\Sigma w V}{\Sigma w} \quad \cdots\cdots\cdots(\text{II. } 17)$$

$w_1, w_2, \cdots, w_n$ を**各觀測値の重み** (Weight) と云ふ. w は 1 より大なることも小なることも有り得るが,常に正である.

若し $w_1 = w_2 = w_3 = \cdots = w_n = k$ とすれば,即ちすべての重みが等しい時は, M' は算術的平均 M となる. 故に重みづけられたる平均は一般化せられたる算術的平均である.

重みは如何にして決定するかと云ふに,推定又は獨斷によつて決定することもあるが,各觀測値の多少によつて決定することもある.

元來重みづけられたる平均は觀測値が不公平に平均に關與しない爲のものである. 例にあげた場合の樣に語學,數學,國語等の重要性に應じて,重みを附けるのであるから,その重要性を正しく公平に推斷しなければ,却て甚しく不公平となる處がある. 故に生物統計學等では,推定又は獨斷により重みを決定することは避くべきである.

例へば,日本人男子の身長を r 區に分ちて計測し,その各區域の人口を $n_1, n_2, n_3, \cdots, n_r$ とし,各地域に於ける身長の算術的平均を $M_1, M_2, M_3, \cdots, M_r$ とすれば,日本人全體の平均身長は $n_1, n_2, \cdots, n_r$ を重みとする重みづけられたる平均 M' を求むればいい.

これは單に重みの等しい算術的平均:

$$M = \frac{M_1 + M_2 + M_3 + \cdots M_r}{r}$$

よりは,より正確なる日本人の平均身長である.

しかし,各區域の眞の人口 $n_1, n_2, n_3, \cdots, n_r$ を用ふれば,大きな數を取扱はねばならぬ故に,この人口の比を用ふるのが一層便利である.

$$n_1 : n_2 : n_3 : \cdots\cdots = \bar{n}_1 : \bar{n}_2 : \bar{n}_3 : \cdots\cdots : \bar{n}_r$$

とし,簡單なる數 $\bar{n}_1, \bar{n}_2, \cdots, \bar{n}_r$ を重みとするがいい.

この人口の比を用ふる時は,各區域の眞の人口が判明せずとも,近似的比が明になりさへすれば計算が出來る.

若し重みを觀測總數とする時は,重みづけられたる平均は公式 (II. 8) となる. その他各地方にて觀測せられたる値より,出生率,死亡率,結婚率等を計算せんとするには,重みづけられたる平均を計算するのが正しい.

例 13. 第18表(例 5)の數を用ひ,北部朝鮮人の身長平均を計算せよ. 但し各道の人口は次の通りとす(單位1000人).

平安南道	1,234,	2
平安北道	1,372,	2
咸鏡南道	1,349,	2
咸鏡北道	629,	1

$$M' = \frac{166{\cdot}77 + 2 \times 166{\cdot}18 + 2 \times 165{\cdot}46 + 2 \times 163{\cdot}50}{7}$$

$$= 165{\cdot}29$$

演習問題

1. 第一章,例1,油菊の分布より,M及び Med. を計算し,その殆ど相一致することを確めよ.

2. 第一章,例2,兵士の身長分布よりM及び Med. を求め,その殆ど一致することを見,相稱分布型では Med. を計算するには及ばざることを確めよ.

3. 第一章,例6,英國人男子の體重分布表より,M及び Med. を計算し,その一致せざることを確めよ. 一般に不相稱型ではMと Med. とは一致せざるが故に,不相稱型では Med. も無意義でない.

4. 次の分布表の算術的平均を求む.

第 23 表

變數	0	1	2	……	$(n-1)$	n
度數	q^n	$nq^{n-1}p$	$\frac{n(n-1)}{1\cdot 2}q^{n-2}p^2$	……	nqp^{n-1}	p^n

但し$q+p=1$とす. 附錄數學公式を用ひよ.

5. 或る年の國勢調査にて人口P_o,滿n年後の調査に於てP_nであつたとせよ. その中間滿$\frac{n}{2}$年後の人口は,P_oとP_nとの幾何平均なることを證明せよ.

解. 初めの人口P_o,滿1年後の人口rP_o,滿2年後の人口r^2P_o,滿$\frac{n}{2}$年後の人口は$r^{\frac{n}{2}}P_o$,而して$P_n = r^nP_o$,

即ち
$$P_{\frac{n}{2}} = \sqrt{P_oP_n}$$

但し,今は人口の增加率を常數と假定したのであることを忘れてはならぬ.

6. 相稱分布又は輕度の不相稱なる分布で,σに比しMが大なる價であれば,

$$M > G > H$$

なることを證明せよ. (但し第三章を理解して後に解くを可とす.)

參考文獻

一般的參考書としては Johannsen, Yule, Pearl, Czuber 等, その他 Mode に關するもの:—

1. Pearson, Karl :- Skew Variation in Homogeneous Material. *Phil. Trans. Roy. Soc. Serie A, Vol. 186, 1895.*
2. Yule, G. U. :- Note on the History of Pauperism in England and Wales, etc. *Jour. Roy. Stat. Soc. Vol. 59, 1896.*
3. Pearson, K. :- On the Modal-Value of an Organ or Character. *Biometr. Vol. I, 1902.*

Geometric mean に關するもの:-

4. Galton, Francis :- The Geometric Mean in Vital and Social Statistics. *Proc. Roy. Soc. Vol. 29, 1879.*
5. McAlister, Donald :- The Law of the Geometric Mean. *ibid.*
6. Žižek, Franz :- Die statistische Mittelwerte.

岡崎文規譯:- 統計的中數値論. 東京有斐閣,大正15年.

原著は讀む暇が無かつたが,譯文は相當頁數に達して居る. 廣く平均について說明し公式を用ひないが爲に,記述抽象的で讀者に不必要な困難を强ひて居る.

第三章
散　布　度

第一項　散布度

1. 散布度

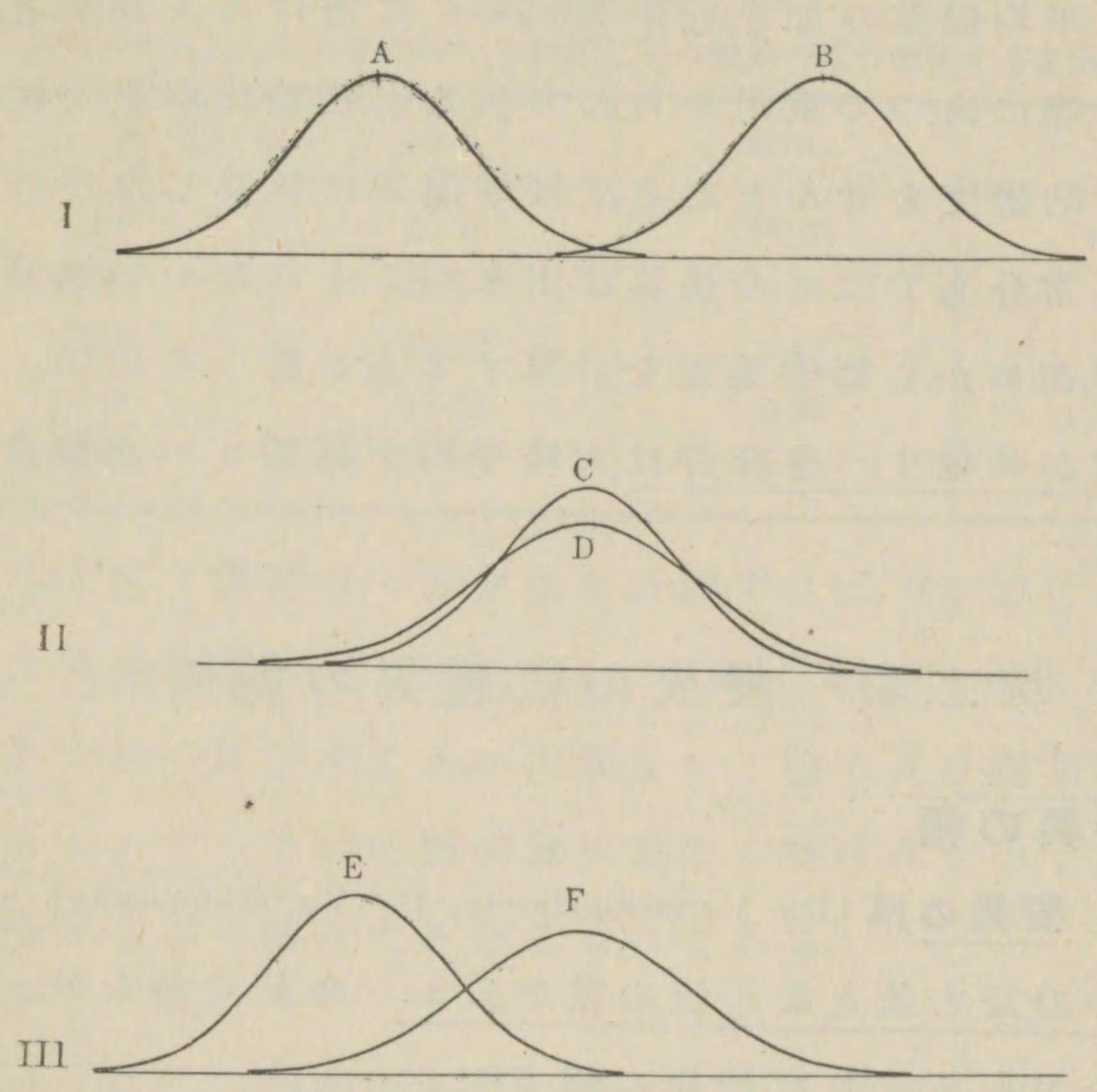
度數分布表で變異を概觀することが出來るけれども,更に進んで,簡單な數字で分布を代表する數値が必要である. 種々なる平均値はその一種であつた. しかし,平均値は分布の中心の位置を指示するのみで,分布の密度を表示して居ない. 第21圖Iに於ける二分布曲線,AとBとは分布の密度が同一である,即ち分布曲線は等しいが,中心の位置,即ち平均が相違して居る. 又IIに於けるCとDとを比較すればこれに反して,平均は同一であるが,分布の密度が相違して居る. DはCに比し分布が疎で,曲線は低く且つ廣範圍に分布して居る. IIIに於けるEとFとは分布の中心も,密度も,共に相違して居る例である. これを見ても,分布の密度を研究することは,平均値に次いで必要なることが知られるであらう.

散布度とは分布の密度を表はす數値である.

平均に種々ある如く,種々の散布度がある. これに

(1) 變異の幅(範圍),

(2) 四分の一偏差,



第21圖　異なる分布の比較.

I. 平均を異にせる二分布.　II. 散布度のみを異にせる二分布.
III. 平均及散布度を異にせる二分布.

(3) 百分法,

(4) 標準偏差,

(5) 平均偏差,

(6) 比較散布度

等がある.

このうちで,1—5は**比較散布度**に對して,總稱して**絕對散布度**と云ふ. これらの中標準偏差が最も普通に用ひられ,且つ分布の理論的研究には缺くべからざるものである. 他の散布度は代數的取扱が困難乃至不可能であるが爲に,理論的研究には適

しない. 平均偏差の如き,近來 München に於ける人類學者,民族衛生學者等に由つて重用せられて居るが,代數的取扱が困難で,從て理論的研究をせんとならば,標準偏差に換算しなければならぬ. 正常分布ではこの換算は出來ぬことはないが,換算する位ならば初めから標準偏差を計算する方が勝つて居る.

散布度が具備すべき條件は,大略平均が具備すべき條件と同一である.

第二項　變異の幅(變異の範圍)

2. 變異の幅

定義 I. **變異の幅** (Die Variationsbreite, Die Variationsweite) とは觀測値の中の最大値と最小値の差を云ふ. 今最大値を Max. とし,最小値を Min. とすれば,變異の幅 B は

$$B = \text{Max.} - \text{Min.}$$

である.

この最大値と最小値と,或は變異の幅は共に屢用ひられる數であるが,經驗によれば變異の幅は甚だ不正確な數であつて,(i) 計測總數の爲に影響せられること大であり, (ii) 偶然の影響を受けることも亦大であつて不適當な數値である.

例 1. Johannsen が Braune Prinzessbohnen を計測し, 120 個, 2,500 個, 5,000 個等,計測數を增加するに從ひ,變異の幅を記載したのである. さうすると Max. も Min. も變化し變異の幅も第24表 B 欄の如く,初め 4·75 から 9·00 まで變化して居る.

觀測數 10,000 に至り B が初めて一定したかの觀があるが,

第 24 表

Braune Prinzessbohnen を計測しその變異の幅の變化するを示す.

觀測總數	Max.	Min.	B
120	15·50	10·75	4·75
2,500	16·25	8·25	8·00
5,000	17·00	8·25	8·75
10,000	17·25	8·25	9·00
12,000	17·25	8·25	9·00

しかし,これも偶然で,一層多數を計測すれば, B が更に大となることは見易き道理で, 10,000 乃至 12,000 の觀測でも未だ B が一定になつたと見なすことが出來ない. 卽ち B は觀測數に影響されること大であつて,觀測數少數なる時は B も小であるが,大なれば B も亦一般に大となるものである.

變異の幅は,又,偶然の影響を受くること大である. 卽ち,偶然に甚大なる一個,又は偶然甚小なる一個が交つて居たが爲に,この Max. 又は Min. は著しき影響をうけるものである.

例 2. Johannsen は Braune Prinzessbohnen 7,500 を 2,500 個づつの三つの群に分ち,その最大及び最小値を求めた所,第25表の如き結果となつた.

第 25 表

Braune Prinzessbohnen の變異の幅を示す.

	N	Max.	Min.	B
I　群	2500	16·25	8·25	8·00
II　群	2500	17·00	8·25	8·75
III　群	2500	17·00	9·75	7·25

何れも2,500個と云ふ多數を計測して居ながら,B は7·25, 8·00, 8·75 と甚しき開きがある. B の浮動のない確かな數を得る爲には,一層多數を計測しなければならない. 然るに,普通2,500個の計測を行ひうるのは特別な場合であつて,觀測數十乃至數百位で滿足しなければならないことが多いから,一般には,B は更に一層甚しき浮動があるであらう. その他次の缺點がある.

變異の幅はすべての觀測値を基礎としない. 唯,兩極端の二つの値を以て代表せしむるのみで,與へられたる變異の幅以内での散布の狀態には無頓着な數である. 故に變異の幅が同一であつても,必ずしも,分布が類似して居ると考へられないのである.

代數的取扱が不可能であることも,不利な點である. 要するに,變異の幅は最悪の散布度であつて,使用にたえない[1].

第三項 標準偏差

3. 標準偏差の定義

定義 II. 各觀測値の平均との差を偏差 (Deviation ; Die Abweichung) と云ふ. 但し,平均には算術的平均の外,稀に中央値を用ふることもある.

觀測値を, $V_1, V_2, V_3, \cdots\cdots, V_N$,

算術的平均を M,

偏差を $a_1, a_2, \cdots\cdots, a_N$ とすれば,

[1] Johannsen は Ein zuverwerfendes unbrauchbares Mass der Variabilität であると云つて居る.

$$a_1 = V_1 - M,\quad a_2 = V_2 - M,\quad a_3 = V_3 - M,\quad \cdots\cdots$$

である.

定義 III. 偏差の平方の總和を觀測數で除したるものの平方根を標準偏差と云ふ (Standard deviation ; Die Standardabweichung, Die stetige Abweichung ; Die mittlere quadratische Abweichung)

標準偏差を σ とすれば,

$$\left.\begin{aligned} \sigma &= \sqrt{\frac{1}{N}\{a_1^2 + a_2^2 + \cdots\cdots + a_N^2\}}, \\ \text{又}\quad \sigma &= \sqrt{\frac{\Sigma a^2}{N}}, \end{aligned}\right\} \cdots\cdots(\text{III. } 1)$$

σ は正なる價を取る. もし,正負共に取る必要のある時は,特に, $\pm\sigma$ と記號するのが普通である.

度數分布が與へられて居る時は,計算はやや簡單となる.

$x_1 = X_1 - M,\quad x_2 = X_2 - M,\quad \cdots\cdots$ とすれば,

$$\left.\begin{aligned} \sigma &= \sqrt{\frac{1}{N}(x_1^2 f_1 + x_2^2 f_2 + x_3^2 f_3 \cdots\cdots + x_K^2 f_K)}, \\ &= \sqrt{\frac{\Sigma(x^2 f)}{N}} = \sqrt{\frac{\Sigma(x^2 f)}{\Sigma f}} \end{aligned}\right\} \cdots\cdots(\text{III. } 2)$$

例 3. 前章例 1 に於ける頭骨の全額面角の標準偏差を計算せよ.

解. $M = 84{\cdot}38°$ であるから,各階級の中央の値とこの M との差を求め,これを第四列に記入した. 即ち,x の列である(第14表).

次に x^2 を計算し,第五列に記入し, fx^2 を第七列に記入し,最後に第七列の値の和 1171·9 を求める.

$$\Sigma(fx^2) = 1171{\cdot}9,$$

$$\therefore\ \sigma = \sqrt{\frac{1171{\cdot}6}{93}} = 3{\cdot}55$$

となる.

4. 標準偏差の簡單なる計算法 I.

x_1, x_2, x_3, ···· 等は普通簡單な數でなく,三桁又は時には四桁の數であるから,その平方を計算するのは骨が折れる. そこで,原點 A を任意に定め,各階級中央の値の A よりの偏差を求める. 即ち,

$$\xi_1 = X_1 - A,\ \ \xi_2 = X_2 - A,\ \ \xi_3 = X_3 - A,\ \cdots\cdots\ \text{とし}$$

$$\left.\begin{aligned} s &= \sqrt{\frac{1}{N}\{f_1\xi_1^2 + f_2\xi_2^2 + \cdots\cdot\}} \\ &= \sqrt{\frac{\Sigma(f\xi^2)}{N}} \end{aligned}\right\} \cdots\cdots\cdots\cdot(\text{III. 3. } a)$$

とすれば, s を**平均平方根偏差** (root-mean-square deviation) と云ふ. 若し,原點 A を算術的平均に取る時は,平均平方根偏差は標準偏差に一致する. 又,若し原點 A を算術的平均以外の所に選んだ場合,容易に s を σ に換算することが出來る.

原點 A を何れかの階級の中央の値(なるべく分布の中心に近き或る階級中央の値)に取る時は,計算がことに容易である. ξ はすべて小なる整數となるからである. s と σ との關係は次の通りである.

$$X - M = x,\quad X - A = \xi,$$

故に, $\xi = x + M - A,$

$$\xi = x + d,\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots(\text{A})$$

但し, $d = M - A$ とす.

A 式は詳しく記述すれば次の如くなる.

$$\xi_1 = x_1 + d,\ \ \xi_2 = x_2 + d,\ \ \xi_3 = x_3 + d,\ \cdots\cdots\cdots(\text{B})$$

さて, B式を用ひて,

$$\begin{aligned} & f_1\xi_1^2 + f_2\xi_2^2 + f_3\xi_3^2 + \cdots\cdot + f_K\xi_K^2. \\ &= f_1(x_1+d)^2 + f_2(x_2+d)^2 + f_3(x_3+d)^2 + \cdots\cdot + f_K(x_K+d)^2, \\ &= f_1x_1^2 + f_2x_2^2 + f_3x_3^2 + \cdots\cdot + f_Kx_K^2 \\ &\quad + 2d(f_1x_1 + f_2x_2 + f_3x_3 + \cdots\cdot + f_Kx_K) \\ &\quad + d^2(f_1 + f_2 + f_3 + \cdots\cdot + f_K). \end{aligned}$$

然るに前章第 8 節定理 I より

$$f_1x + f_2x_2 + f_3x_3 + \cdots\cdot + f_Kx_K = 0$$

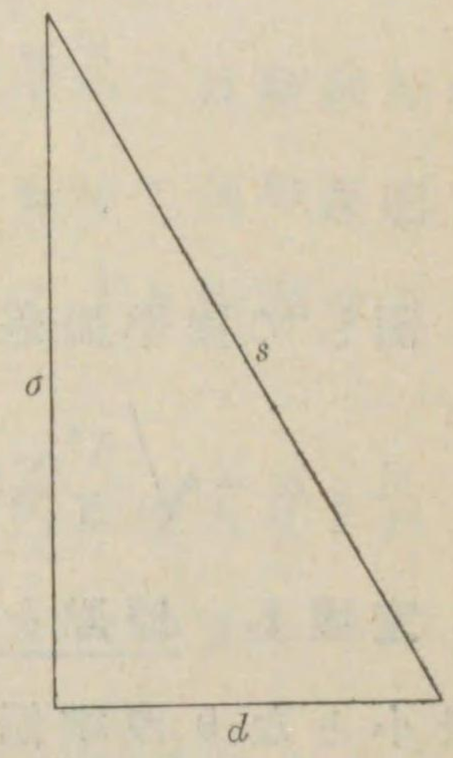
であるから,

$$\sum_{i=1}^{K} f_i\xi_i^2 = \sum_{i=1}^{K} f_ix_i^2 + d^2\sum_{i=1}^{K} f_i,$$

即ち, $Ns^2 = N\sigma^2 + Nd^2,$

$$\left.\begin{aligned} \sigma^2 &= s^2 - d^2, \\ \text{又,}\quad \sigma &= \sqrt{s^2 - d^2}\,. \end{aligned}\right\} \cdots\cdots\cdots\cdots\cdots(\text{III. 3. } b)$$

σ と s と d との關係は,直角三角形の三邊の關係である. 底邊に d をとり,斜邊に s をとれば,他の一邊は σ となる. 故に, s と d とを知りて σ を計算するに,直接(3.b)式によることも出來るが,ミリメーター紙で近似的に圖計算を行ふことも出來る.



第 22 圖 σ と s と d との關係を示す.

例 4. 前章例 2 に於ける日本人男學生の頭長幅指數の分布表より,標準偏差を計算せよ(第15表).

解. σ の計算は例 1 の如く行へるけれども,ここでは簡便法によることとし,先づ s^2 を計算する.

原點 A に算術的平均を計算せる時の原點 80·5 を用ふれば便利である. ξ は第三列に,ξ^2 は第六列に, $f\xi^2$ は第七列に記入した. $\Sigma f\xi^2$ は第七列の値の總和である.

$$s^2 = \frac{81190}{6000} = 13{\cdot}532,$$

而して, $d = M - A = 80{\cdot}81 - 80{\cdot}50 = 0{\cdot}31$

$d^2 = 0{\cdot}096$ であるから

$\sigma = \sqrt{13{\cdot}436} = 3{\cdot}666.$

5. 標準偏差の簡便なる計算法 II.

階級幅 1 なる時, s (平均平方根偏差)の計算には,公式 (3. a) 及び(3. b)を用ふ. 若し,階級幅が 1 ならずして λ なる時は,

$$\xi_1 = \lambda\zeta_1, \quad \xi_2 = \lambda\zeta_2, \quad \xi_3 = \lambda\zeta_3, \cdots\cdots$$

一般的に $\xi = \lambda\zeta$ と置く時は ζ は 0, 1, 2, $\cdots\cdot$ 等整數である

$$\therefore \quad s^2 = \frac{1}{N}(f_1\lambda^2\zeta_1{}^2 + f_2\lambda^2\zeta_2{}^2 + f_3\lambda^2\zeta_3{}^2 + \cdots\cdot)$$

$$s^2 = \frac{\lambda^2}{N}(f_1\zeta_1{}^2 + f_2\zeta_2{}^2 + f_3\zeta_3{}^2 + \cdots\cdot),$$

卽ち, $$s^2 = \frac{\lambda^2}{N}\Sigma f\zeta^2.$$

而して,標準偏差は $\sigma = \sqrt{s^2 - d^2}$

又 $$\sigma = \sqrt{\frac{\lambda^2}{N}\Sigma f\zeta^2 - d^2} = \lambda\sqrt{\frac{1}{N}\Sigma f\zeta^2 - \left(\frac{\Sigma f\zeta}{N}\right)^2} \cdots \text{(III. 4)}$$

定理 I. 原點を算術的平均に選びたる時,平均平方根偏差が最小となり,標準偏差に一致す.

證明は極めて容易であるから讀者自ら試みられよ.

例 5. 前章例 3 に於ける牛乳搾取量の度數分布表から,標準偏差を計算せよ(第 16 表).

解. 例 4 の如く,ξ と ξ^2 を求めていいけれども,階級幅が大であるから, ξ^2 は可なり大なる數となる. 故に, λ を階級幅とし, $\xi = \lambda\zeta$ とす. ζ は簡單な整數となるを以て,ζ^2 も亦極めて容易に計算することが出來る. ζ^2 は第六列に,$f\zeta^2$ は第七例にあり.

$$\sum f\zeta^2 = 80623,$$

$$\frac{\Sigma f\zeta^2}{N} = 12{\cdot}127,$$

$$s^2 = 12{\cdot}127\lambda^2,$$

$$\sigma^2 = s^2 - d^2 = 12{\cdot}127\lambda^2 - c^2\lambda^2, \left(\text{但し } c = \frac{\Sigma f\zeta}{N}\right),$$

$$= \lambda^2(12{\cdot}127 - 0\ 413^2),$$

$$\sigma = 50 \times \sqrt{11{\cdot}916},$$

$$= 172{\cdot}6.$$

σ の甚しく大なる點に注意せよ.

6. Sheppard の修正式

正常分布,又は輕度の不相稱分布の算術的平均(1) は階級幅の大小によつて影響せらるることは少いので有つて,何等修正を要しないが,標準偏差の計算は,階級幅の大さにより影響せらるること大である.

度數分布表より σ を計算するには,一階級內に編入せられる觀測値が,悉く階級中央の値を取つたと假定したのであるが,こ

(1) J-型分布又は强度の不相稱型では,階級粗大なる時,算術的平均に大なる影響があるから,その時は階級幅をなるべく小に取る.

れは嚴密に云ふならば正しくない. Mの計算では,正の偏差の階級と,負の偏差の階級と誤差互に相殺したのであるが,標準偏差では平方を取るが故に,誤差は相殺しない. 故に,階級幅大に過ぎる時は,著しき誤差を生じ,嚴密なる意味での標準偏差は,常に,修正をしなければならないのである. これには Sheppard の修正式と云ふのを用ひる.

$\bar{\sigma}$を修正せる(正しき)標準偏差,λを階級幅とすれば,

$$\bar{\sigma}=\sqrt{\sigma^2-\frac{\lambda^2}{12}} \quad \cdots\cdots\cdots\cdots\cdots\cdots\text{(III. 5. a)}$$

又,この式に公式(3. b)を代入すれば,

$$\bar{\sigma}=\sqrt{s^2-d^2-\frac{\lambda^2}{12}} \quad \cdots\cdots\cdots\cdots\cdots\text{(III. 5. b)}$$

證明は省略する.

σ^2に比し$\dfrac{\lambda^2}{12}$が甚だ小なる時は,σに修正を行はずとも大した差異がない. けれども,修正を施したか否かは報告書に明記して置かねばならぬ. 一般に階級數が少い場合は,Sheppard の修正を行ひ,階級數多き場合,例へば,分布表の階級數25以上もある様な時には,修正を行ふ必要はないと心得て居てよい.

例 6. 例5の標準偏差に Sheppard の修正を施せ.

例3では, $\sigma=172{\cdot}63,\ \lambda=50,$

$$\bar{\sigma}=\sqrt{172{\cdot}63^2-\frac{50^2}{12}},$$

$$=\sqrt{29791-208}=\sqrt{29583},$$

$$\bar{\sigma}=172{\cdot}0$$

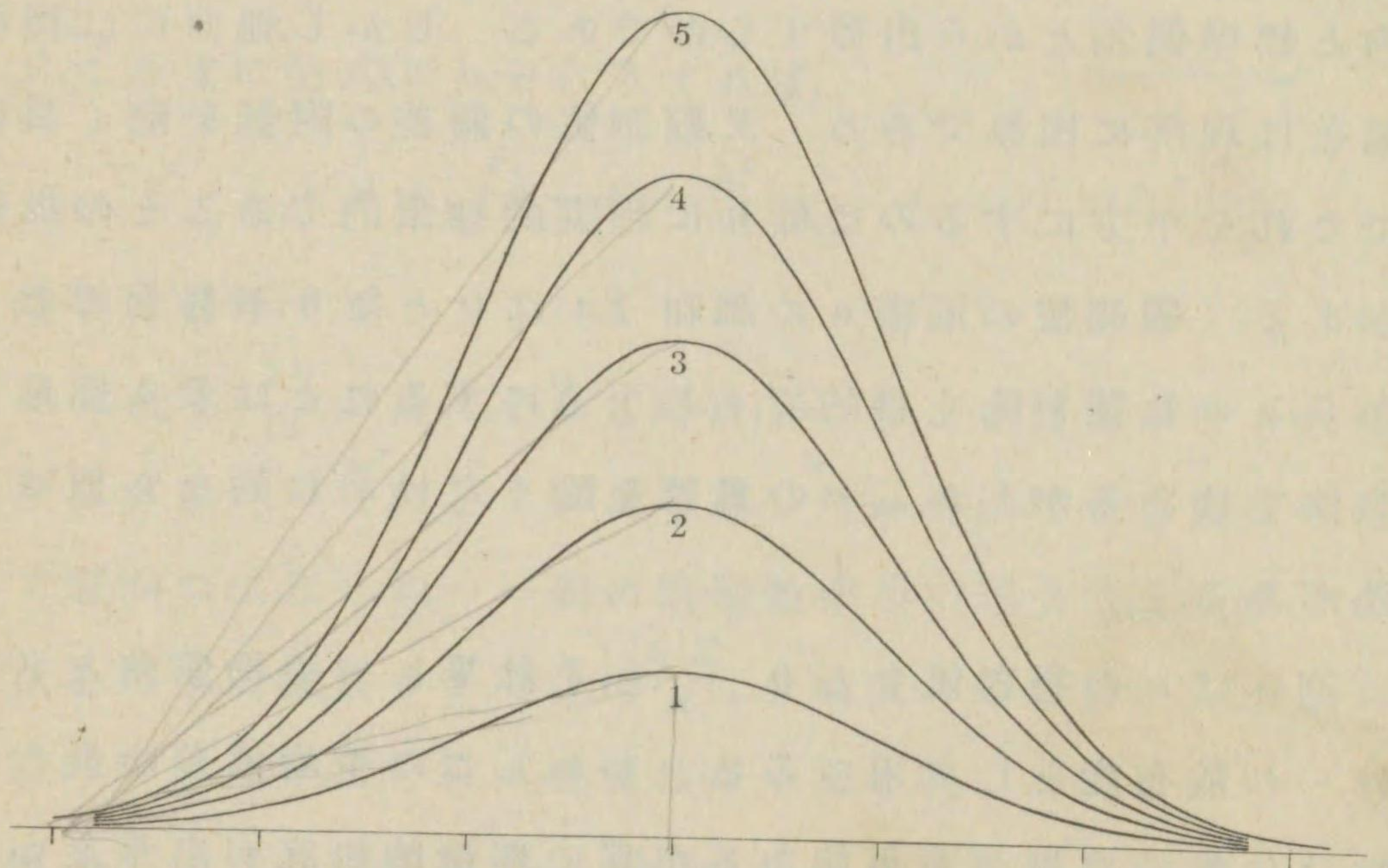
第16表には$\bar{\sigma}$の一層簡便なる計算が施されて居る.

7. 標準偏差の特性 I.

標準偏差なる名は Pearson のつけたもので,昔は平均誤差とか,中數誤差とか名づけて居たものである. 平均の具備すべき條件を,算術的平均が最もよく備へて居る如く,同様の條件を散布度のうち標準偏差が,最もよく具備して居る. 即ち,すべての觀測値を基礎として計算し,計算が容易で,偶然の影響も少く,且つ代數的取扱が可能である. この代數的取扱が可能なることは最も必要な條件であつて,之あるが爲に,理論的研究は算術的平均と標準偏差とから出發するのである. しかし,他面には,標準偏差は理解に困難である. 又,觀測値の偏差の附號を除く目的でこれを平方にするのは,餘りに理論的抽象的であるとの批難がある. 觀測値の偏差aの總和Σaは0となり,不都合であるから,この負號を除く目的で,aを平方にすることは最も簡單な方法ではあるが,しかし,aの負號を除くだけが目的なら,他に方法がある.

例へばaの絶對値をとり$\dfrac{\Sigma|a|}{n}$を計算して平均偏差と名づけ,一の散布度として用ひらる. しかしこの平均偏差の最大缺點は,代數的取扱が不可能なるが爲に,理論的研究が出來ない點にある. 誤差論でも,古來$\dfrac{1}{N}\Sigma a^2$が用ひられ,生物學でも Galton 以來標準偏差が用ひられて居るのも決して偶然でもなく,古來の學者が數を玩んだものでも無いのである. 又,標準偏差の計算も簡便なる方法を用ふれば決して困難でない. ことに計算器を用ふる人には極めて容易で,平均偏差と同程度に寧ろ却て迅速に計算しうるのである.

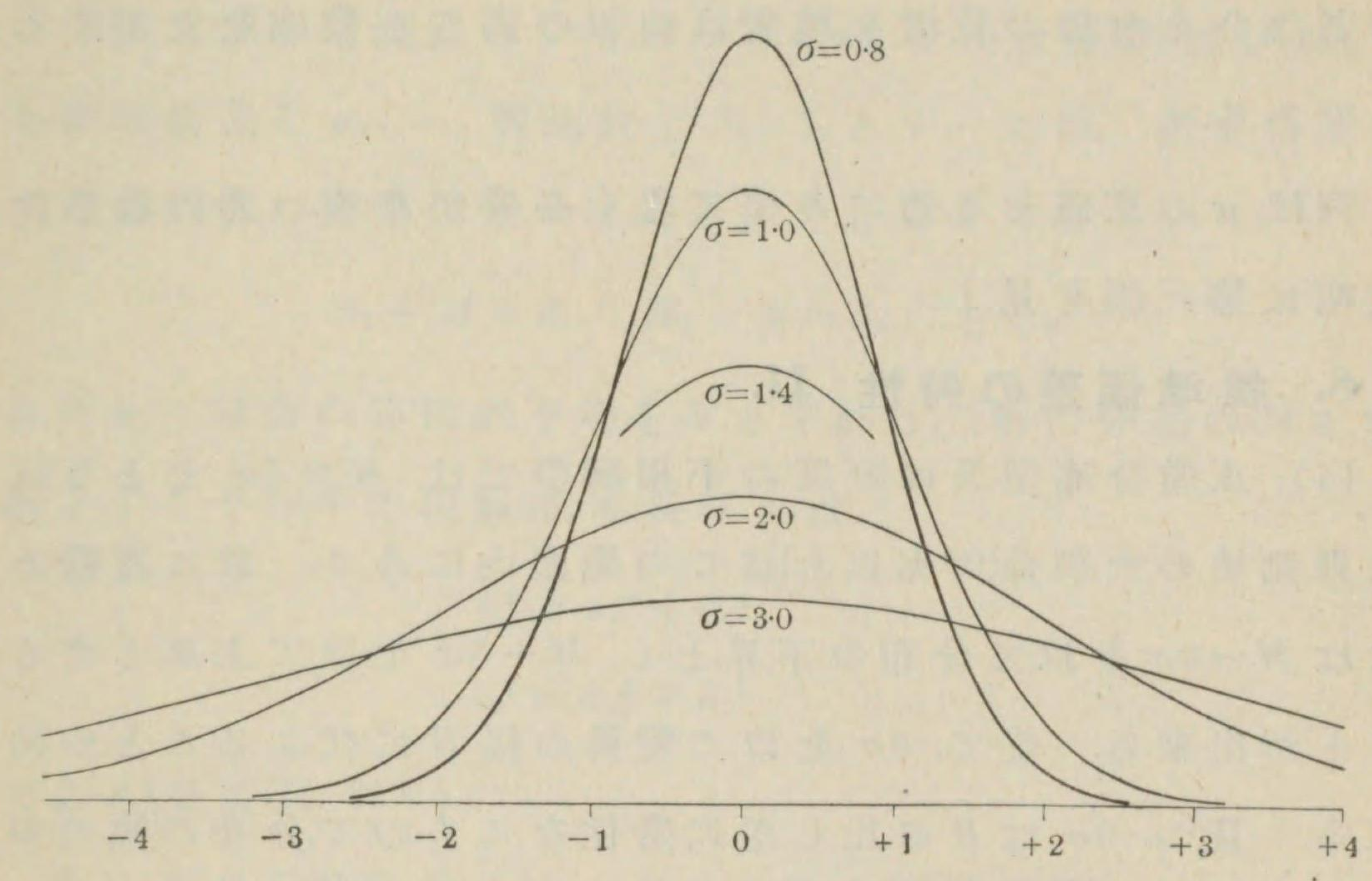
正常なる分布を代表する數値が三つある. 觀測總數と算術的平均と標準偏差である. このNとMとσとのうち,若しNが異なれば,度數曲線は第23圖の如く變化する. 1なる度數曲線に比して,觀測總數が2倍となれば圖の2;3倍となれば圖の3;4倍,5倍せるものは4,5の如き分布曲線となる. 一般に,Nがr倍になれば,各度數はr倍になる. 又Nが變つてもMとσとに大なる變化はない,唯,その信賴度がかわるのみである.(信賴度については第三編を見よ.)



第23圖　Nを異にする種々の正常分布曲線.

次に,算術的平均が變れば,分布の中心の位置がかはる. 從て分布曲線も全體として,右または左に偏することになる(第21圖, I. III.). 又σが變れば,散布の程度が變る. 若し,觀測總數を同一にして[1],二つの分布を比較すれば第24圖の如く,σの小なるも

(1) 二つの分布の分布型を比較するには,觀測數を同一にしなければならぬ. 即ち,共に千分比又は百分比分布となして比較するのである.



第24圖　σを異にせる種々の正常分布曲線.

のでは觀測値は平均に近く密集し,度數の密度は中央に於て高く,兩側に至る時は急に低くなり,やがてx軸と一致するが,σ大となれば,觀測値の多くは平均より遠くに散在し,中央の度數漸次小となるに至るものである. 第24圖に於ける横線上の數字は,$\sigma=1$なる曲線の$\pm\sigma\pm2\sigma$等の位置を示したもの,又曲線の數字はその曲線のσの値である. 故に,σは又觀測値の不正確さを表はすのであるから,$\frac{1}{\sigma}$を以て**正確度**と稱する人もある. 又, Lexisの如く,$\frac{1}{\sqrt{2}\sigma}$を以て正確度(Precision)とし,hを以て表はすこともある. $h=\frac{1}{\sqrt{2}\cdot\sigma}$, $\sqrt{2}\sigma=\frac{1}{h}$を以てModulus (Airy)と云ふこともある.

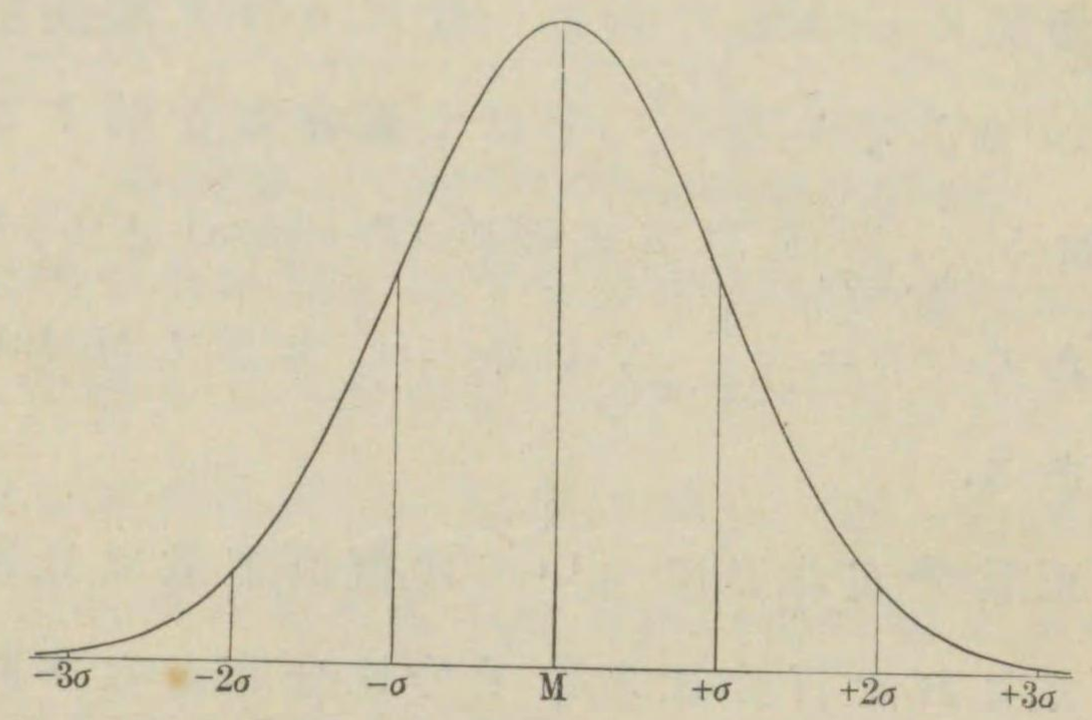
要するに正常分布の形については,觀測總數は比較的必要なものではなく,これを百分比又は千分比分布となす時は,その算術的平均の位置と標準偏差とが分布を代表するに至る. 特に

前者は分布曲線の位置を,後者は曲線の高さと散布を支配するのである.

尙ほ,σ の意義をここにあげて置く必要がある. 其の數學的證明は第三編を見よ.

8. 標準偏差の特性 II.

(i) 正常分布型又は輕度の不相稱型では, $M\pm3\sigma$ をとる時は觀測値の大部分(99%以上)はこの範圍內にあり. 故に,實際上には $M-3\sigma$ を以て分布の下界とし, $M+3\sigma$ を以て上界とすることが出來る. 從て, 6σ を以て變異の幅 B に代ふることが出來る. 且つ, 6σ は B に比し遙に常住なるもので,分布の幅の如く甚しく變動するものでなく,好都合である. 若し,完全なる正常分布ならば,觀測値の 99·7% は $M\pm3\sigma$ の範圍內にある(第三編). この事實は又 σ の概略の檢算に用ふることが出來る. 卽ち, $M\pm3\sigma$ を取り,度數の約 99·5% が,(觀測値小なる時は100%が)その範圍內にありや否やを檢するのである. 但し,著しき不相稱なる分布及び階級甚だ少き分布には適用できない.



第 25 圖　正常分布曲線とその標準偏差.

(ii) 二群の觀測値ありて,その算術的平均は各 M_1 及び M_2, その標準偏差を σ_1, σ_2, 觀測數を N_1, N_2 とす. この二群を合同せる時,合同群の標準偏差 σ は次の如く計算せらる.

$$M_1-M=d_1,\quad M_2-M=d_2\quad \text{とし,}$$

合同せる場合の算術的平均を M とすれば,二群の分布の M を原點とする平均平方根偏差, s_1 及び s_2 は,

$$s_1^2=\sigma_1^2+d_1^2,$$
$$s_2^2=\sigma_2^2+d_2^2$$

である(公式III. 3. b.).

故に,この二群を混じて一とせる時の標準偏差 σ は,

$$\sigma=\sqrt{\frac{N_1(\sigma_1^2+d_1^2)+N_2(\sigma_2^2+d_2^2)}{N_1+N_2}}\cdots\cdots(\text{III. 6. a})$$

又若し, $N_1=N_2$ なる時は, $d_1=d_2$ なるを以て,

$$\sigma=\sqrt{\frac{\sigma_1^2+\sigma_2^2+2d_1^2}{2}}.\quad\cdots\cdots\cdots\cdots\cdots\cdots(\text{III. 6. b})$$

若し,三群以上の合同をなす時は,

$$\left.\begin{aligned}&\sigma=\sqrt{\frac{\sum_{i=1}^{m}N_i\sigma_i^2+\sum_{i=1}^{m}N_id_i^2}{\sum_{i=1}^{m}N_i}},\\ \text{或は}\quad&\sigma=\sqrt{\frac{N_1(\sigma_1^2+d_1^2)+N_2(\sigma_2^2+d_2^2)+\cdots\cdots}{N_1+N_2+N_3+\cdots\cdots}}.\end{aligned}\right\}\cdots(\text{III. 6. c})$$

第四項　平均偏差

9. 平均偏差

これも，Galton 以來用ひられて居るのであるが，近來 Lenz その他 München の人々によつて盛に使用せられて居るものである．

定義 IV. 各觀測値の，平均よりの偏差を a とし，a の絕對値の算術的平均を平均偏差と云ふ．(Mean deviation; Die durchschnittliche Abweichung, Die arithmetische Abweichung).

算術的平均よりの各觀測値の偏差を $V_1-M=a_1,\ V_2-M=a_2,$ .... とし，ε を平均偏差とすれば，

$$\left.\begin{aligned} \varepsilon &= \frac{1}{N}\{|a_1|+|a_2|+|a_3|+\cdots\cdots+|a_n|\}, \\ \text{或は,}\quad \varepsilon &= \frac{1}{N}\sum_{i=1}^{N}|a_i|, \\ \text{或は,}\quad \varepsilon &= \frac{\Sigma|a|}{N}, \end{aligned}\right\}\cdots(\text{III. }7)$$

又度數分布表となつて居る時は，

$$\left.\begin{aligned} \varepsilon &= \frac{1}{N}\{|f_1a_1|+|f_2a_2|+|f_3a_3|+\cdots\cdots+|f_Ka_K|\}, \\ \varepsilon &= \frac{\sum_{i=1}^{K}|f_ia_i|}{N} = \frac{\sum_{i=1}^{K}|f_ia_i|}{\sum_{i=1}^{K}f_i}. \end{aligned}\right\}(\text{III. }8)$$

平均偏差として近來 Lenz 其の他の人の使用して居るのは，算術的平均よりの偏差 $|a|$ の算術的平均であるが，以前屢用ひられたのは，次の定義によるものがあつた．**平均偏差**は，各觀測値の，中央値よりの偏差 b の絕對値の算術的平均である．

Med. を中央値，e を平均偏差とし，又 $V_1-\text{Med.}=b_1,\ V_2-\text{Med.}=b_2,$ .... とすれば

$$\left.\begin{aligned} e &= \frac{1}{N}\{|b_1|+|b_2|+|b_3|+\cdots\cdots+|b_n|\}, \\ &= \frac{\sum_{i=1}^{N}|b_i|}{N} \end{aligned}\right\}\cdots\cdots(\text{III. }9)$$

度數分布表からは次の如く計算する．

$$\left.\begin{aligned} e &= \frac{1}{N}\{|f_1b_1|+|f_2b_2|+\cdots\cdots\}, \\ &= \frac{\sum_{i=1}^{K}|f_ib_i|}{N} = \frac{\sum|fb|}{N}, \end{aligned}\right\}\cdots\cdots\cdots(\text{III. }10)$$

平均平方根偏差は算術的平均を原點とした時，卽ち標準偏差に一致する時が，最小であつた．しかるに，

定理． 平均偏差は，原點を中央値に取りて計算したる時最小の値をとる．

故に，平均偏差を計算するには，中央値を原點として計算するのが最も合理的である．しかし，近來は多く算術的平均を原點として計算して居る．これ平均値のうち算術的平均が最も普通に用ひられるのと，M を原點とせる ε と，Med. を原點とせる e とは一般に大なる差違がなきとによる．

定理の證明．今，觀測數を N とし，小なる値の方より m 個の觀測値を越えたる所に一の原點を取り，計算せる平均偏差を $\varDelta$ とせよ．次に $(m+1)$ 個を越えたる所に原點を取れば m 個の偏差は c だけ大となり，$(N-m)$ 個の偏差は c だけ小となる．故に

新しき偏差$\varDelta'$は,

$$\varDelta' = \varDelta + \frac{mc-(N-m)c}{N},$$

即ち,　　$$\varDelta' = \varDelta + \frac{c}{N}(2m-N).$$

そこで,初めmが小である時は$2m-N$は負であるから,原點の移動するに從ひて,$\varDelta'$は$\varDelta$より小となる. 然し$2m=N$となる時は$\varDelta'=\varDelta$となる. 又$2m-N$が正となるや,此度は反對に$\varDelta'$は$\varDelta$より大となる. 即ち,$2m=N$なる時$\varDelta$が最小である. 換言すれば原點が中央値にある時,平均偏差が最小である. Nもし偶數ならば,第$\dfrac{N}{2}$番目の觀測値と,$\left(\dfrac{N}{2}+1\right)$番目の觀測値との間の何處に原點を選んでも,$e$は常に等しい. しかし,$N$が奇數なる時は,$\dfrac{N+1}{2}$番目のものに原點を選んだ時に$\varDelta$最小である.

10. 平均偏差の計算法

平均偏差は標準偏差の如く簡便計算が容易でないが,決して出來ないわけではない.

例 7. 前章例1に於ける朝鮮人全顔面角の平均偏差を計算せよ.

第26表はεの計算の順序を示したものであつて,第3列は$|X-M|=|x|$を記入した. 又第4列は$|fx|$であるから,その總和279·86を求め,これを93等分する. 即ち$\varepsilon=3{\cdot}009$となる. 計算器を用ふるならば極めて容易であるが,若しそうでなければ第5列 I. II. の如く計算してもいい.

第 26 表

平均偏差の計算法. ($M = 84{\cdot}38.$)

X	f	$\lvert x\rvert$	$\lvert fx\rvert$	I	II
77	2	7·38	14·76		$7\times2=14$
78	3	6·38	19·14		$6\times3=18$
79	4	5·38	21·52		$5\times4=20$
80	6	4·38	26·28		$4\times6=24$
81	8	3·38	27·04	45	$3\times8=24$
82	7	2·38	16·66		$2\times7=14$
83	9	1·38	12·42		$+\ 1\times9=9$
◎84	6	0·38	2·28		123
85	11	0·62	6·82		
86	6	1·62	9·72		$1\times11=11$
87	10	2·62	26·20		$2\times6=12$
88	9	3·62	32·58		$3\times10=30$
89	5	4·62	23·10	48	$4\times9=36$
90	6	5·62	33·72		$5\times5=25$
91	—	—	—		$6\times6=36$
92	1	7·62	7·62		
					$+\ 8\times1=8$
計	93	—	279·86		158
$\Sigma\lvert fx\rvert=279{\cdot}86,$ $\varepsilon=3{\cdot}009.$				$\varepsilon=\frac{1}{93}(123+158-1{\cdot}14)$ $=\frac{279{\cdot}86}{93}=3{\cdot}009.$	

(1)　先づ算術的平均の屬する階級を求め,その階級の中央の値を原點として,偏差の總和$\Sigma\varDelta'$を計算せよ. (第5列 II).

(2)　次に算術的平均より小なる觀測値の數を求め,これをN_1とし,又平均より大なる觀測値の數を求め,これをN_2とせよ.

$N_1+N_2=N$である. 平均と原點との差を$A-M=d$とすれば,(dは正又は負となることあり,)ここでは$d=-0{\cdot}38$である. N_1個につきてはdだけ多く計算しすぎ,N_2につきてはdだけ

少く計算して居るのである．故に，$(N_2-N_1)d$ だけ $\Sigma\varDelta'$ に加へなければならぬ．

$$\varepsilon=\frac{\Sigma\varDelta'+(N_2-N_1)d}{N}. \quad \cdots\cdots\cdots\cdots(\text{III. }11)$$

第5列の下方にこの運算が施されてある．

例 8. 前章例2に於ける日本人男學生の頭長幅指數について ε を計算せよ(第15表)．

(i) 算術的平均の存在する階級は 80—81 であるから，その中央の値 80·5 を原點とし，偏差 $\varDelta'$ の總和を求むるに，第15表第四列より，

$$\Sigma\varDelta'=7658+9502=17160,$$

(ii) 算術的平均 80·81 以下の値を取る度數，　$N_1=3236$

80·81 以上の値を取る度數，　$N_2=2764$

但し，80—81 の階級はその中央の値 80·50 であるから，その級度數 709 は N_1 に繰入れる．

(iii)
$$A-M=d=+80\cdot5-80\cdot81$$
$$d=-0\cdot31,$$

(iv)
$$\begin{aligned}\varepsilon&=\frac{1}{6000}\left\{\sum\varDelta'+(N_2-N_1)d\right\},\\&=\frac{1}{6000}\{17160+472\times0\cdot31\},\\&=\frac{1}{6000}(17160+146\cdot32),\\&=\frac{17306}{6000}=2\cdot884,\end{aligned}$$
$$\varepsilon=2\cdot884.$$

特に，(N_2-N_1) と d との附號に注意して計算すべし．尙ほ熟練する迄は例7前半の方法により計算する方が安全である．

例 9. 同一の分布表より，Median を原點とする平均偏差 e を求めよ．M に代ふるに Med. を以てすればいい．

(i) 先づ原點を 80·5 とし，$\Sigma\varDelta'=17160$，

(ii) Med. は 80·67 であるから N_1，N_2 は例8の通り，又

$$N_1=3236,\quad N_2=2764,\quad N_2-N_1=-472$$

又
$$d=A-\text{Med.}=-0\cdot17,$$

(iii)
$$\begin{aligned}e&=\frac{1}{6000}\left\{\sum\varDelta+(N_2-N_1)d\right\},\\&=\frac{1}{6000}(17160+472\times0\cdot17),\\&=\frac{1}{6000}(17240\cdot24)=2\cdot873,\end{aligned}$$

ここで注意すべきことは，$e=2\cdot873$ は $\varepsilon=2\cdot884$ より小なることで，上の定理を滿足する．

11. 平均偏差の特性

平均偏差は計算が容易であり理解し易く，標準偏差の如く技術的數學的であるとの批難は全くない．又，若し病的に大，若くは特別に小なる觀測値が有つても，これが ε に(又は e に)影響する所極めて小であるから，これらの點で標準偏差とちがつて都合がいい(Lenz)．しかし，救ふことの出來ない短所は代數的取扱が不可能なことである[1]．その爲に理論的研究には適しないが，標準偏差と平均偏差との間には，正常分布では，

$$\left.\begin{aligned}\sigma&=1\cdot2533\,\varepsilon\quad\text{又は}\quad\sigma=\sqrt{\frac{\pi}{2}}\,\varepsilon\\\varepsilon&=0\cdot79788\,\sigma\quad\text{又は}\quad\varepsilon=\sqrt{\frac{2}{\pi}}\,\sigma\end{aligned}\right\}\cdots\cdots(\text{III. }12)$$

[1] 故に第三編で理論的研究には專ら標準偏差を用ふる．

なる關係がある(第三編參照).

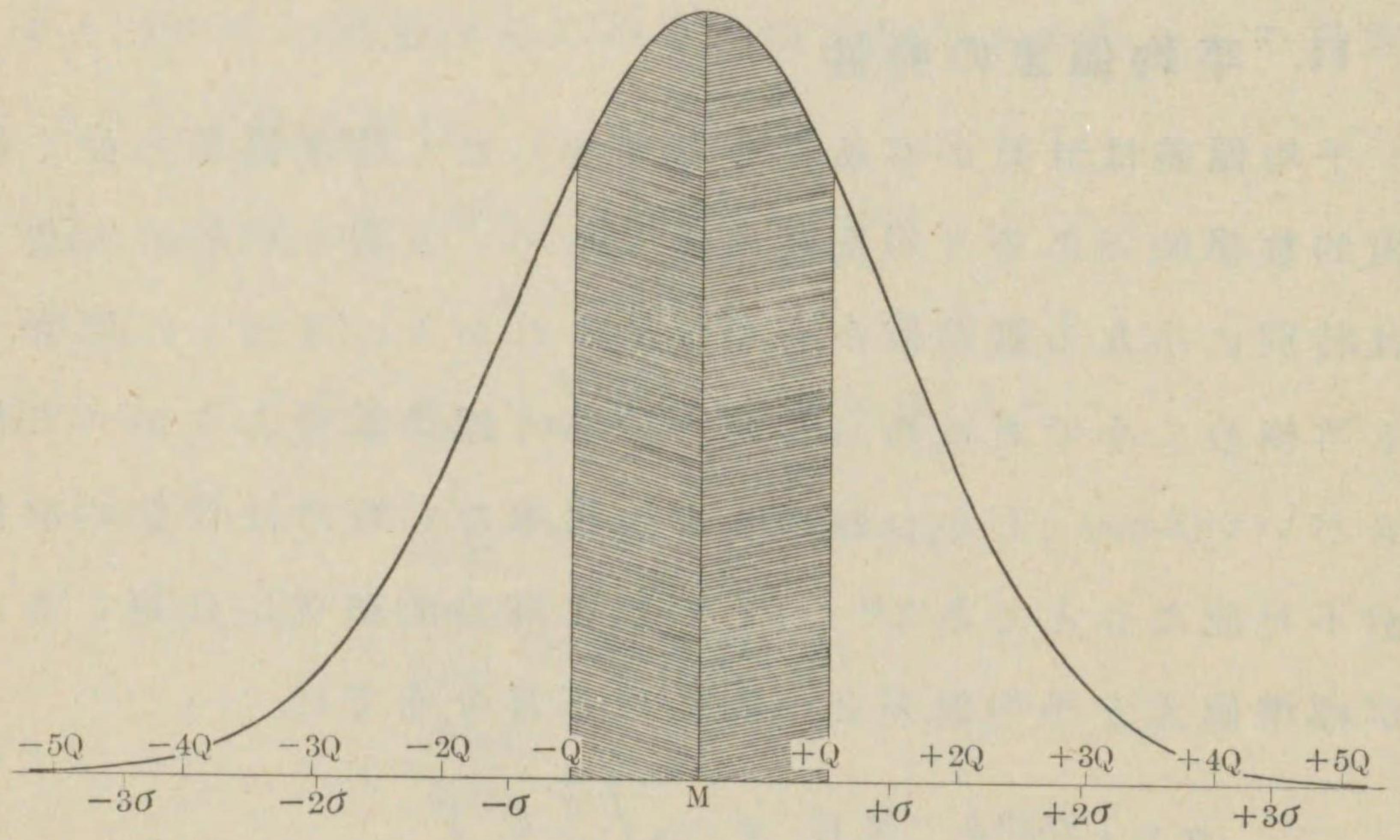
$$\text{或は}\quad 4\sigma \fallingdotseq 5\varepsilon \quad\cdots\cdots\cdots\cdots\cdots(\text{III. }12)$$

とすることが出來る. 又,正常分布又は輕度の不相稱型では,6σが觀測値の99%以上を含むに對して,$7\cdot5\varepsilon$は觀測値の99%以上をふくむ. 何となれば,(III. 12)より,

$$6\sigma \fallingdotseq 7\cdot5\varepsilon$$

だからである.

この式は完全なる正常分布につきての關係であるが,正常分布に近い實測の分布に於ても通用するものと考へていい. 又,この關係があるからεでは理論的研究は出來ないが,σで理論的に計算したものを,εに換算することが出來るのである.



第26圖 正常分布曲線. σ と Q の位置. $3Q \fallingdotseq 2\sigma$ なること, 3σ 又は 4·5Q は殆ど分布の限界であること, M 及び ±Q に引ける垂線は曲線下の面積を四等分することを示す.

第五項 四分の一偏差

12. 四分の一偏差の定義

これも Galton が初めて使用したものであるが,今日では主として標準偏差が用ひられ,四分の一偏差はあまり使用せられない. 唯,確率誤差の計算にはこの四分の一偏差を用ふるが故に,一應詳說して置く必要がある.

定義. すべての觀測値の$\frac{1}{4}$がq_1より小に,$\frac{3}{4}$が之より大なるが如き値q_1を,**第一四分限界**(又**第一四分の一位**)と云ふ (Lower quartile; Die erste Viertelgrenze). 又,觀測値の$\frac{1}{4}$がq_3より大に,$\frac{3}{4}$がこれより小なるが如き値q_3を,**第三四分の一限界**(**第三四分の一位**)(Upper quartile; Die dritte Viertelgrenze), 總稱して**四分一限界**と云ふ. 第二四分の一限界は**中央値**である. 全分布は, q_1 Med., q_3に由つて全く四等分せられ,各の區間に觀測總數の$\frac{1}{4}$が存在する.

第26圖はq_1, q_3, 及びMに垂線を引いたのである. (Med. に代ふるに今假りにMを用ひたが,不相稱分布では正しくは Med. でなければならぬ.) この三個の垂線で曲線下の面積(卽ち度數分布)は四等分せられて居る. 正常分布であれば

$$q_3 - \text{Med.} = \text{Med.} - q_1$$

であるが,實測の場合では必ずしも等しくない. 卽ち,

$$Q_2 = q_3 - \text{Med.},$$

$$Q_1 = \text{Med.} - q_1$$

に於けるQ_1とQ_2とは等しくない. そこでQ_1とQ_2との平均卽ち

$$Q = \frac{q_3 - q_1}{2} \quad \cdots\cdots\cdots\cdots\cdots\cdots \text{(III. 13)}$$

を**四分の一偏差** (Quartile-deviation, Semi-inter-quartile range, Die Quartilabweichung) と名づく.

四分の一偏差のこの計算は,平易で,簡單で,理解し易く,偶然の影響も少く,特別に小なる又は特別に大なる觀測値の爲に影響せらるること極めて僅微であり,觀測數によつて變化することも少く,散布度としてはいいものである.

Johannsen は Braune Prinzessbohnen につきて,觀測數を變更して Q を計算した. それによると,

$N = 120$ なる時, $Q = 1{\cdot}26$;

$N = 2{,}500$ なる時, $Q = 1{\cdot}23$;

$N = 12{,}000$ なる時, $Q = 1{\cdot}24$.

であつた. 度數分布の兩端にある階級の境界が明でない場合,又各階級の幅が不等なる場合(第6表)でも,四分一偏差の計算には差支がないと云ふ便利がある. かく幾多の長所があるが,唯代數的計算の出來ないことは,救ふべからざる缺點である. 尙ほ,一つ重要なる四分の一偏差の性質は,正常分布に於ては,

$$\left.\begin{array}{l} Q = 0{\cdot}67449\sigma, \\ 3Q \fallingdotseq 2\sigma \end{array}\right\} \quad \cdots\cdots\cdots\cdots \text{(III. 15)}$$

と云ふことである. 詳しくは觀測値の正確さの章を見よ.

13. 四分の一偏差の計算法

計算法は中央値の計算法に類する. 先づ全分布の $\frac{1}{4}$ を超えた時の變數の値 q_1 と, $\frac{3}{4}$ を超えたる時の値 q_3 を求めねばならぬ.

例 10. Feuerbohnen と云ふ豆の直徑の分布より Q を計算せよ.

第 27 表

Feuerbohnen の直徑(Q の計算法).

直徑	17—18	18—19	19—20	20—21	21—22	22—23	23—24	24—25	25—26	26—27	27—28	28—29	29—30	30—31	31—32	32—33
度數	5	13	38	41	95	124	152	134	129	100	70	45	38	7	7	2
累積度數	5	18	56	97	192	316	468	602	731	831	901	946	984	991	998	1000

第27表の第一行と第二行とは,普通の度數分布表である. 次に第三列の如く,累積度數の表を順次作つて行くのである. この表は,階級境界の値が與へられて居るのであるから, 18 mm. 以下の個體數 5, 19 mm. 以下の個體數 18, ‥‥ 等を意味する. かく數へると 33 mm. 以下のもの總計 1000 となる. さて, $\frac{1000}{4} = 250$ が q_1 の下にある樣に q_1 を選ぶのである. しかるに 22 mm. 以下の觀測値 192 個あり,又 22—23 なる階級の度數 124 である. この 124 個がこの階級内に平等に分布せりと假定する.

今, $250 - 192 = 58$, 即ち 22 mm. より數へて,第58番目の個體を超えた所に, q_1 が有る筈である. 即ち,

$$q_1 = 22 + \frac{58}{124} = 22{\cdot}468,$$

$$\fallingdotseq 22{\cdot}47.$$

同樣に, q_3 は 750 個の觀測値を超えたる所にある筈である. 而して 26 mm. までには 731 個あるが故に, $750 - 731 = 19$, 即ち次の階級の第19の區劃を超えたる所が q_3 である.

故に, $$q_3 = 26 + \frac{19}{100} = 26{\cdot}190.$$

さて $$q_1 = 22{\cdot}468,$$

$$\mathrm{Med.} = 24{\cdot}239,$$

$$q_3 = 26{\cdot}190,$$

$$\therefore \quad Q_1 = \mathrm{Med.} - q_1 = 1{\cdot}771,$$

$$Q_2 = q_3 - \mathrm{Med.} = 1{\cdot}951.$$

Q_1とQ_2とが等しくないのは,多少不相稱だからである.

$$Q = \frac{Q_1 + Q_2}{2} = \frac{q_3 - q_1}{2} = 1{\cdot}861.$$

例 11. Butten の尾鰭の棘數の分布表から,Qを計算せよ(第20表).

同様に求めて, $q_1 = 52{\cdot}27,$

$$\mathrm{Med.} = 53{\cdot}66, \quad M = 53{\cdot}67,$$

$$q_3 = 55{\cdot}13,$$

$$Q = 1{\cdot}43.$$

玆に注意すべきことが一つある. Butten の棘の數は不連續數であるから,例18と同方法で q_1, q_3 を求められない筈である.けれども連續變數である場合と同様に,例へば棘數53は 52·5—53·5 の階級と同様に考へて計算するのである. 從て,四分の一偏差はこの場合,よほど抽象的意義を帶びて來る.

14. 四分の一偏差の圖計算法

中央値と同じく,四分の一位の計算も圖計算が行はれる.

例 12. 前章例 8 に於ける,頭骨全顔面角の四分の一位, q_1, q_3 を求め,これより四分の一偏差を計算せよ.

中央値と全く同方法で,同一圖を用ひて q_1, q_3 を計算すればいいのである(第18圖, 19圖).

$$\frac{93}{4} = 23{\cdot}25,$$

$$\frac{93}{4} \times 3 = 69{\cdot}75$$

を求め,第18圖に於ける Y 軸に於て, 23·25(B) 及び 69·75(C) を取り, X 軸に並行線 BB' 及び CC' を引き,累積曲線と B', C' にて交らしむ. 次に B', C' より Y 軸に並行線を引きて, X 軸と B'', C'' にて交らしむ. B'', C'' を X の目盛で讀む時は,その値は q_1, q_3 である.

$$q_1 = 81{\cdot}6,$$

$$q_3 = 87{\cdot}3,$$

若し,一層精密な數を欲する時は,前章例 9 (第19圖)と同一の方法で求めることが出來る.

第六項 百分法

15. 百分法

中央値と四分の一限界とは,全分布を四個に等分するものであつて,各區域に觀測總數の四分の一が存在する. これを擴張して全分布を百分し,各區域内に觀測總數の 1%(又は數%)づつ存在する様に區切りすることが出來る. この方法を**百分法**,又は **Galton の百分法**と云ふ (Method of percentiles). 普通には,其の 5 階級を合し 5% の觀測數が含まれる様に,20個の區分に分つこと; 或は10%が各區分に包含せらるる様に十分することの何れかが行はれる. 後者を特に**十分法**と云ひ,全分布を10個の階級に分ち,各階級の度數が何れも觀測數の10%である様に,境

界值を選定することを云ふのである.

度數分布では階級の幅を一定にして,觀測せるままの度數を記入したのである. 然るに,百分法(又は十分法)では,各階級の度數を等しく取るのである. 從て,各階級の境界は計測値から計算によつて定めらる. 故に階級の幅が同一でない.

ここに便利なことは,百分法(十分法を含む)では,度數を記入するに及ばない. 全範圍を,十分法で10階級に分てば,階級に $\frac{N}{10}$ 個の觀測値が存在することが明白であるから,單に階級の境界を記錄すればいいのである. 例へば,十分法に於て,境界値を c_1, c_2, c_3, c_4, c_5, c_6, c_7, c_8, c_9 とすれば, c_5 は Med. に一致するが, q_1 は c_2 と c_3 との間にあり. 比例部分にて q_1 を求めることが出來る. 又, q_3 も c_7 と c_8 との間にある.

5 %づつ取り,20の階級に分てば, c_{10} は Med. であり,又 c_5, c_{15} は q_1, q_3 に相當する. 故に二十階級に分つのも便利である. この方法は Galton 以來屢用ひられたが,今日では標準偏差の簡單にして理論的なるに若かずとて用ひられない.

例 13. 前章例 2 に於ける,日本人男學生の頭長幅指數の分布に十分法を行ひて,之を區分せよ(第15表).

解. 百分法の計算は中央値及び四分の一位の計算を應用すればいい. $N=6000$ であるから600個體が各區分の度數でなければならぬ(第15表第5列).

$$c_1 = 76 + \frac{70}{326} = 76{\cdot}21,$$

$$c_2 = 77 + \frac{344}{434} = 77{\cdot}79,$$

$$c_3 = 78 + \frac{510}{619} = 78{\cdot}82,$$

$$c_4 = 79 + \frac{491}{618} = 79{\cdot}79,$$

$$c_5 = 80 + \frac{473}{709} = 80{\cdot}67,$$

$$c_6 = 81 + \frac{364}{657} = 81{\cdot}55,$$

$$c_7 = 82 + \frac{307}{549} = 82{\cdot}56,$$

$$c_8 = 83 + \frac{358}{497} = 83{\cdot}72,$$

$$c_9 = 85 + \frac{123}{248} = 85{\cdot}50,$$

故に,十分法の階級は,

68.0—76·21—77·79—78·82—79·79—80·67—81·55—82·56—83·72—85·50—**96.0**

となる. これを一見しても明なる通り,分布の中央は比較的密に區分せられて居るが,兩端は區分が疎である. 又兩端の値は觀測値の Max. と Min. とである.

第七項　比較散布度

16. 比較散布度

標準偏差でも,平均偏差でも,四分の一偏差でも,今まで述べたのは皆**絕對散布度**である. しかし,散布の狀態は觀測値の大さに影響せらるること大なるものがある. 例へば鼻高徑を計測して,その散布度を σ であらはすなれば, 3.26 mm. (朝鮮人♂) であるが,頭骨最大長徑の σ は 6·75 mm. で前者より大であつた. これは頭最大長徑の各觀測値が,鼻高徑より甚しく大なるに基

因するものである．更に，身長の標準偏差は 57·8 mm. と云ふ風に大である．そこで標準偏差を算術的平均の百分比であらはすならば，觀測値の大さから來る影響を除くことが出來る．これを**變異係數** (Coefficient of variation ; Der Variationskoeffizient) と云ひ，v を以てあらはす．

定義 I. **變異係數**とは標準偏差を算術的平均の百分比で表はしたものである．[1]

$$v = 100\,\frac{\sigma}{M} \quad\cdots\cdots\cdots\cdots\cdots\cdots\cdots \text{(III. 16)}$$

又平均偏差(ε)を用ふることもあり，四分の一偏差を用ふることもある．

定義 II. **變異示數** (Der Variabilitätsindex, Der Variationsindex) とは平均偏差を算術的平均の百分比にて表はしたるものである．

$$w = 100\,\frac{\varepsilon}{M} \quad\cdots\cdots\cdots\cdots\cdots\cdots\cdots \text{(III. 17. a)}$$

若し，中央値を原點とせる平均偏差 e を中央値の百分比であらはす時は，變異示數 w' は，

$$w' = 100\,\frac{e}{\text{Med.}} \quad\cdots\cdots\cdots\cdots\cdots\cdots \text{(III. 17. b)}$$

定義 III. **四分の一變異係數** (Der Quartilkoeffizient) とは四分の一偏差を算術的平均の百分比として表はしたものである．

これを κ にてあらはせば，

$$\kappa = 100\,\frac{Q}{M} \quad\cdots\cdots\cdots\cdots\cdots\cdots\cdots \text{(III. 18)}$$

變異係數，變異示數，四分の一變異係數等を總稱して，**比較散布度**

(1) 正しくは，標準偏差に Sheppard の修正を施すべきである．即ち $v = \frac{\sigma}{M}$

(又は**關係的散布度**)と云ふ．

平均も絕對散布度も，觀測値と同次であるが，比較散布度は比であつて，單なる數である．故に計測の單位に關係しない．例へば身長を mm. 單位で計測しても，cm. を單位としても，また吋を單位としても，比較散布度には少しも影響せらるることがないのである．故に，比較散布度は全くことなる計測項目の變異を比較するに用ふることが出來る．例へば，身長の變異係數と，頭指數の變異係數とを比較することが出來或は，人の身長の變異係數と，牛の體重の變異係數と比較することも出來る．

角度の比較散布度を計算することは無意義である．角度，例へば全顏面角等の標準偏差は，算術的平均の大さには無關係であるが爲である．全顏面角は獨乙水平面(O. A. E.)を基準として計測するものであるが，他の平面を基準としても計測し得る．而して，基準を變更すれば M は變化するが，σ は依然として變化しないのである．かくの如き計測値に比較散布度を計算するは，意義が無いのみならず誤りである．

演習問題

1. 第一章，例 1 の油菊の度數分布表(第 3 表)より，σ 及び Q を計算し．公式(III. 15)の適合するや否やを檢せよ．

2. 第一章，例 2 の兵士の身長分布(第 4 表)より，σ 及び Q を計算し，同じ式の適合するやを檢せよ．

3. 第一章，例 6 英國人男子の體重分布表(第 8 表)より，σ を計算し，且つ Q_1, Q_2 を計算し，かく不相稱分布では Q_1 と Q_2 とは同一にあらざることを認識せよ．

4. 第一章,例 3 につき σ 及び $\bar{\sigma}$ を小數點以下三位まで計算し,且つ, $M \pm 3\bar{\sigma}$ の範圍内には觀測數の幾%が存在するかを計算し,相稱分布では標準偏差の特性 II, (i) の略正しきことを認識せよ. 但し, $M \pm 3\bar{\sigma}$ なる値が,或る階級内の一値を取る時,その階級内に於ては觀測値が等間隔に分布して居るものと假定せよ.

5. 第一章,例 7 (第 9 表) *Ranunculus* の分布 B について, M, σ, Q_1, Q_2 を計算し, Q_1 と Q_2 の等しからざることを認め, $M \pm 3\sigma$ の範圍内に分布の幾%が存在するかを計算せよ.

6. 第二章,例 4 (第 17 表)金澤地方に於ける日本人頭骨の上顎面指數の分布表 I, II, III. よりそれぞれ σ 及び $\bar{\sigma}$ を計算し, Sheppard の修正法の用ふるに足ることを認識せよ.

7. 第二章の問題 4 の分布に就て,標準偏差を計算せよ. (備考第十一章に解あり.)

8. 1 より N までの整數の標準偏差を求む.

解. 1 より N までには N 個の不連續變數がある. 又度數は何れも 1 である. 故に, 1 より N までの整數の算術的平均 M は數學公式(公. 19),

$$S = \frac{N}{2}(N+1),$$

を應用して,

$$M = \frac{S}{N} = \frac{N+1}{2},$$

又, 0 を原點として偏差の平方の和を求むるに,

$$1^2 + 2^2 + 3^2 + \cdots\cdots + N^2 = S'$$

であるから,數學公式(公. 23)により,和は $\dfrac{N(N+1)(2N+1)}{6}$;

故に,標準偏差を σ とすれば,

$$\sigma^2 = \frac{N(N+1)(2N+1)}{6N} - \frac{(N+1)^2}{4} = \frac{N^2-1}{12}$$

$$\sigma = \sqrt{\frac{N^2-1}{12}}.$$

9. 本章の問題, (1), (2), (3), (4), (5), (6) に於て計算せる σ 又は $\bar{\sigma}$ より變異係數を計算せよ.

參考文獻

本章の一般的參考書として Yule, Johanns n, Czuber, Pearl, の外に;

標準偏差に關するもの:-

1. Pearson, Karl :- Contributions to the Mathematical Theory of Evolution. *Phil. Trans. Roy. Soc., Serie A, Vol. 185, 1894.*

平均偏差に關するもの:-

2. Trachtenberg, M. I.:- A Note on a Property of the Median. *Journ. Roy. Stat. Soc., Vol. 128, 1915.*

平均偏差と標準偏差の優劣問題については:-

3. Lenz, Fritz :- Bemerkungen zur Variationsstatistik u. Korrelationsrechnung u. einige Vorschläge. *Arch. f. Rass.- u.. Gesells.-Biol., Bd. 15, 1923/24.*

4. Weinberg, Wilhelm :- Bravais cd. Lenz, durchschnittliche od. mittlere Streuung. *Arch. f. Rass.- u. Gesells.-Biol., Bd. 18, 1926.*

同誌に Lenz と Weinberg との尙ほ數回の論爭あり.

四分の一偏差及び百分法に關するもの:-

5. Galton, Francis :- Statistics by Intercomparison, with Remarks on the Law of Frequency of Error. *Phil. Meg., Vol. 49, 1875.*

6. Galton Francis :- Natural Inheritance. London, 1889.

比較散布度に關するもの:-

7. Pearson, Karl :- Regression, Heredity, and Panmixia. *Phil. Trans. Roy. Soc., Serie A, Vol. 187, 1896.*

8. Verschäffelt, E. :- Ueber graduelle Variabilität von pflanzlichen Eigenschaten. *Ber. deutsch. bot. Gesells., Bd. 12, 1894.*

その他次の如きものが有るが,多くは歷史的價値あるのみ.

9. Ihering, H. V. :- Zur Einführung von Oscillationsexponenten in die Craniometrie. *Arch. f Anthrop., Bd. 10, 1878.*

10. Stieda, Ludwig :- Ueber die Anwendung der Wahrscheinlichkeitsrech-

nung in der anthropologischen Statistik. *Arch. f. Anthrop., Bd. 14, 1883.*

11. Bartels, P. :- Untersuchungen u. Experimente an 15000 menschlichen Schädeln. *Zts. f. Morph. u. Anthrop., Bd. 7, 1904.*

12. Derselbe :- Ueber die Anwendung feinerer mathematischer Methoden in der anthropologischen Statistik. *Zts. f. Morph. u. Anthrop., Bd. 9, 1906.*

13. Bartels u. Fuchs :- Ueber die Bedeutung des Bartelschen Brauchbarkeitsindex. *ibid.*

14. Ranke, K. E. :- Ueber die Bedeutung des Bartelschen Brauchbarkeitsindex. *Zts. f. Morph u. Anthrop., Bd. 8, 1905.*

第四章
種々の分布型

第一項 正常分布型

1. 基本分布型と合成分布型

第一章で三種の基本分布型を記述したが,この外にも種々の分布型がある. 今これを列擧すれば,

基本型
1. 正常分布型,
2. 輕度なる不相稱型,
3. 强度の不相稱型 (又 J-字形分布).

合成型
4. U-字形分布,
5. 二峯及び多峯分布型,
6. 高峯分布型.

等である. 今これらにつき少しく說明して見よう.

2. 正常分布 (Normal distribution; Die normale Verteilung, Die symmetrische Variationsreihe)

最も普通に出現し,その理論的研究も最も容易であり,又研究が行屆いて居る. 正常分布を代表する三つの數値がある(1). 觀測總數と算術的平均と標準偏差とである. この三つが與へられて居る時は變異係數でも,種々の標準誤差(2),若くは確率誤差(2)

(1) 正常ならざるときは代表する數値は4個以上となる.
(2) 第三編參照

でも或は平均偏差,四分の一偏差等皆この三つの數値より計算することが出來る. それのみならず度數分布表までも,亦三つの數値から數學的に計算することが出來るのである.

この分布に類似せる分布を捜して見るに,二項級數の係數がある. この分布も中央高く,兩端漸次低く,遂に0となる. 全く一種の相稱分布である. そこで n を種々の整數とし, $(p+q)^n$ を展開して見れば,

$$
\begin{aligned}
(p+q)^1 &= p+q,\\
(p+q)^2 &= p^2+2pq+q^2,\\
(p+q)^3 &= p^3+3p^2q+3pq^2+q^3,\\
(p+q)^4 &= p^4+4p^3q+6p^2q^2+4pq^3+q^4,\\
&\cdots\cdots\cdots\cdots\cdots\cdots\\
&\cdots\cdots\cdots\cdots\cdots\cdots
\end{aligned}
$$

その係數ばかりを取る時は,次表甲の如き多數の級數となる. これを**二項係數**と云ふ.

第28表
二項係數の表

甲表

$(p+q)^1$	1 1
$(p+q)^2$	1 2 1
$(p+q)^3$	1 3 3 1
$(p+q)^4$	1 4 6 4 1
$(p+q)^5$	1 5 10 10 5 1
$(p+q)^6$	1 6 15 20 15 6 1
$(p+q)^7$	1 7 21 35 35 21 7 1
$(p+q)^8$	1 8 28 56 70 56 28 8 1
$(p+q)^9$	1 9 36 84 126 126 84 36 9 1
$(p+q)^{10}$	1 10 45 120 210 252 210 120 45 10 1

乙表

1 + 1
1 + 2 + 1
1 + 3 + 3 + 1
1 + 4 + 6 + 4 + 1
1 + 5 + 10 + 10 + 5 + 1

甲の係數を書き下すに,第一行より行へば,順次下の行は上の行より計算することが出來る. 即ち,第二行は第一行の二數の和をその數の中間に書き,且つ兩端に1を書き加へ, 1, 2, 1 が出來上る: 第三行は第二行の二數間に其の和を書き,その兩端に1を書き加へ, 1, 3, 3, 1 が出來上る. 同樣に,第四行以下もすべて上の行の二數の間にその二數の和を書き,兩端に1を加へることによつて, n を任意に取りたる時の係數を悉く書き下すことが出來る. 乙はその順序を示したものである. 二項係數の分布は, n が如何なる値を取つても常に對稱であつて,しかも分布の中央最大で,これより兩側には漸次小となり,遂に兩端に到りて0となる. かくの如き點が,正常分布に似て居り,しかも n を大にすれば益類似して來るのである.

n 無限大なる時の二項級數の分布は,

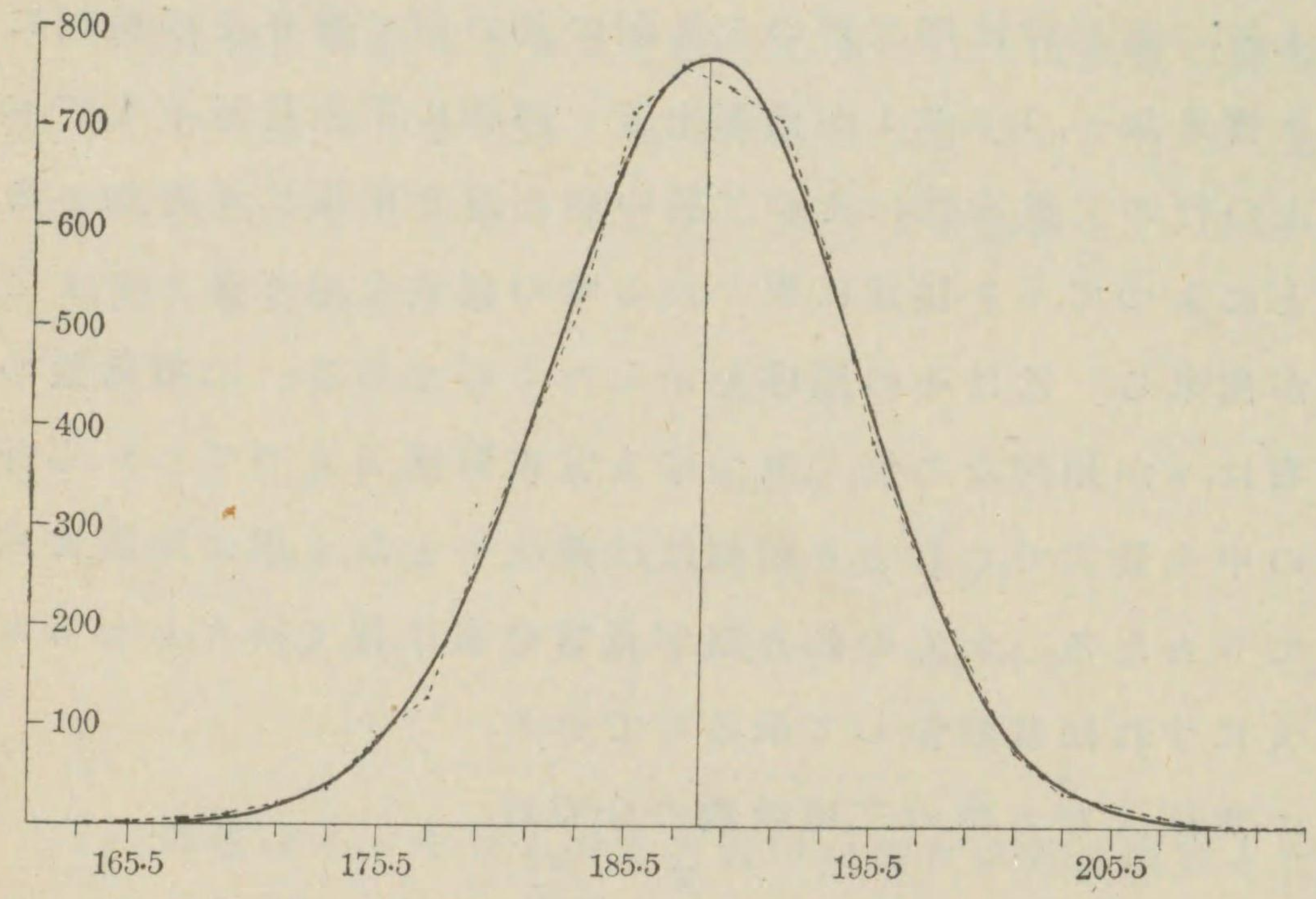
$$y=\frac{N}{\sqrt{2\pi}\,\sigma}e^{\frac{x^2}{2\sigma^2}}$$

なる函數形(正常確率函數)で表示することが出來る. これ正常分布形である.

生物界に於ける度數分布の大部分は,分布の中心に於て度數が最大であつて,之より兩側に向つて對稱的に減少し,遂に度數が0となる. 且つ分布は**正常確率函數**で表はすことが出來る. 之を**ケトレーの法則**(Das Queteletsche Gesetz)と云ふ(第三編參照).

第27圖は日本人男子6,000人の頭最大長徑の度數分布である, 虛線は實測の度數分布で有り,實線は數學的に計算せる正常分布である. この二者が如何によく一致せるかを見よ. 少數の計測から作つた度數分布表ではかくまで一致することはない

が,それは多くは偶然の影響の爲である．正常分布形の計算法等は第三編に詳しく說くこととし,今は,正常分布曲線を目分量で描くことを說明するに止めよう．



第 27 圖　日本人男子 6,000 人の頭最大長徑の度數多角形及びこの分布と N, M, σ を同じくする正常曲線．

3. 正常分布型の目測描畫法

元來正常分布曲線を描畫するには計算によらねばならぬが,目分量でもその概略を描き得るのみならず,目分量の描畫に熟練すれば,計算による正常度數曲線の誤謬を容易に發見し得る樣になるから,これも一通り知つて置きたい．

(a) 先づ度數多角形(又は Histogram)を書き,實測分布が正常分布に近きことを確める．若し實測分布が不相稱型であれば,正常分布を書くことを斷念しなければならぬ(第 28 圖)．

(b) この度數多角形の角點の間を通過し,これらの點に最も

近き正常曲線を書く．曲線は角點を通過する必要はないが,次の事項に注意せねばならぬ．

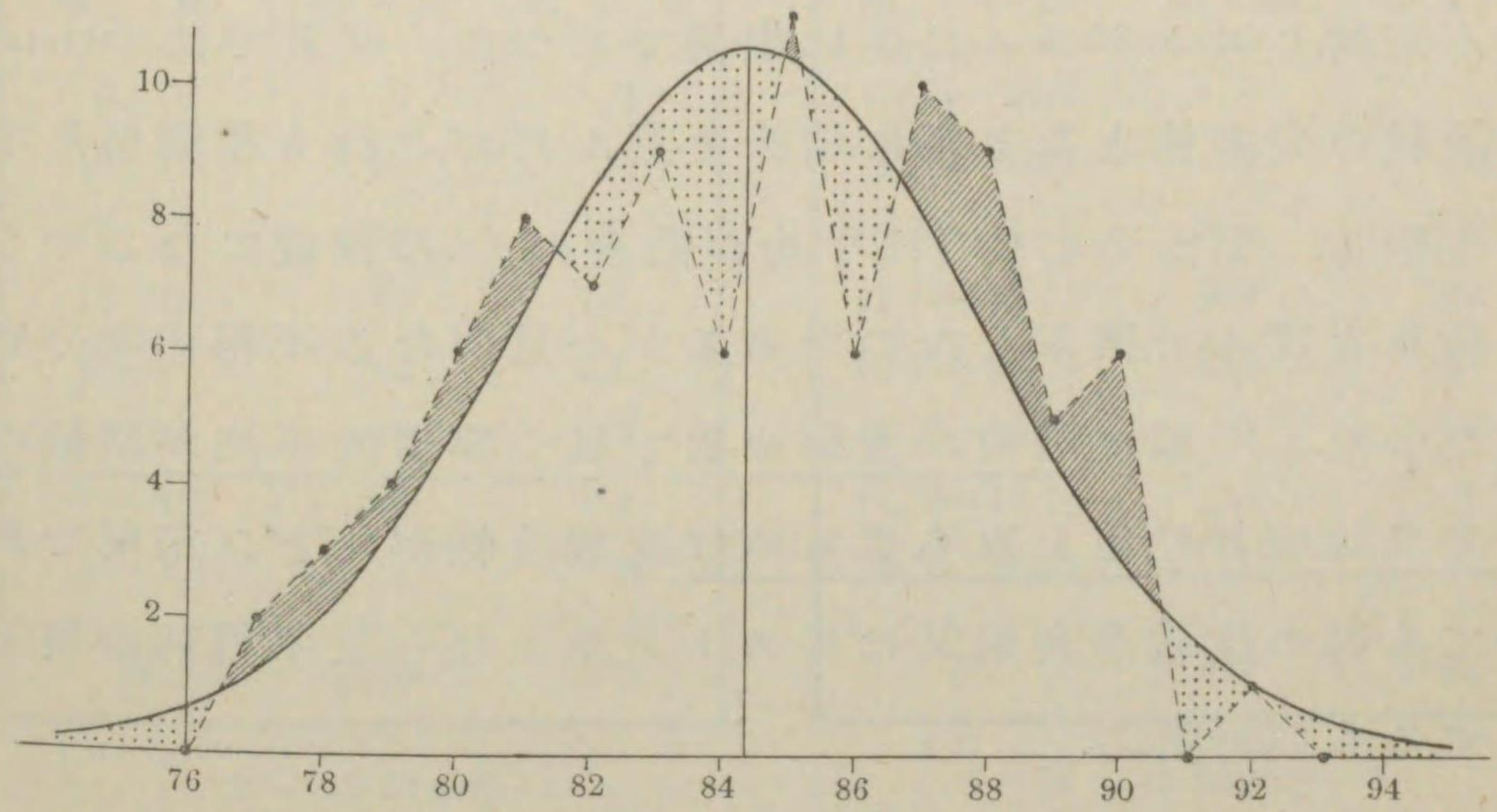
(c) 曲線は平均(M)に最高點があり,この點を中心として兩側に相稱に展開する樣に畫くべきである．しかも曲線は平均の兩側に於ては初め極めて緩徐に,次で急峻に下降し,終に再び緩徐に下降しつつ原線に接近する(第 28 圖其の他)．

(d) 度數多角形の角點の約半數が曲線の一側に,又約半數が他側にある樣に畫かねばならぬ．

(e) 第 28 圖の陰影を施せる部分の總面積と,點を施せる所の總面積とが等しくなければならぬ．これ曲線と橫軸との間にある部分の面積は,度數多角形の面積に等しいが爲である．

(f) 標準偏差を σ とすれば, $M\pm\sigma$ の位置に變曲點がある樣に畫くべきである(第三編參照)．

(g) $M\pm3\sigma$ より兩側で,曲線は橫軸に接近し,極めて緩徐なる傾斜をなすのみである(第 25 圖其の他)．



第 28 圖　度數多角形と度數曲線との比較．

第28圖は93個の朝鮮人男子頭骨の全顏面角の度數分布(虛線)と,計算による正常分布(實線)とを比較したものであるが,以上の諸點に注意して曲線を畫けば,この圖の實線に略一致するであらう.

第二項 不相稱分布型

4. 似て非なる不相稱分布型

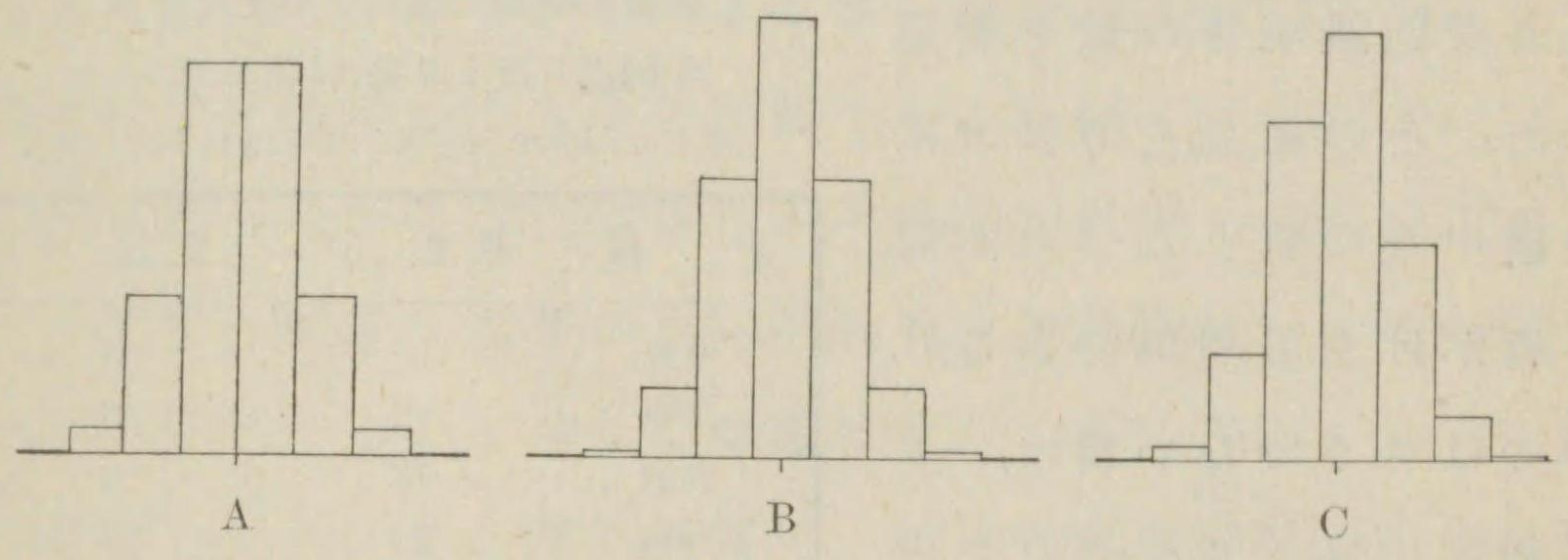
不相稱分布型は種々なる原因で起るが,就中,一見不相稱型で,實は正常分布なる場合がある. これは階級數が少數なる時にのみ起るのであるから,少しく熟練せる者には容易に發見することが出來る.

今,一の正常分布がありとし, $M = 50$ mm., $\bar{\sigma} = 1$ mm. とし,階級幅を 1 mm. に取り,正常分布より計算して Histogram を圖示するに(計算法は第三編參照),三樣に描きうる.

第29圖 A では,平均 50 mm. が階級境界の値となる樣に階級を選んだ所,ヒストグラムは明に相稱であつた. 又 B では, 50 mm. が階級中央の値となる樣に境界を選んだが,これも相稱型を示して居る. しかるに C では平均が階級の $\frac{1}{4}$ の位置になるやうに境界を選んで,圖示したのであるが,一見明なる不相稱型分布となつた. 一般に階級の數が少數で,且つ算術的平均が階級の中央又は境界の値を取らざる時は,度數曲線がたとひ相稱であつても其の度數多角形又はヒストグラムは,一見不相稱の觀を呈するものである.

平均が階級の境界値又は中央の値を取ることは稀であるか

ら,多くの度數分布は常に多少見掛け上の不相稱を呈するのである. 但し階級數多數である時は,この不相稱が一般に目立たない.



第29圖. 正常分布を Histogram にて描寫したる時に生ずる似て非なる不相稱型. A は平均が二階級の間にあり, B は階級の中央にあり, C は階級の $\frac{1}{4}$ の所にある.

第 29 表

階級の取り方によりて似て非なる相稱型のあらはれるを示す.

甲 表

階級	度數	正分布
9—10	7	3
10—11	67	75
11—12	466	488
12—13	761	706
13—14	201	233
14—15	15	17
15—16	5	1
16—17	1	0
計	1523	1523

$M = 12{\cdot}25$ mm,
$\bar{\sigma} = 0{\cdot}767$ mm,

乙 表

階級	度數	正分布
8·75— 9·75	2	1
9·75—10·75	43	38
10·75—11·75	314	353
11·75—12·75	809	738
12·75—13·75	316	353
13·75—14·75	30	38
14·75—15·75	6	1
15·75—16·75	2	0
計	1522	1522

$M = 12{\cdot}25$ mm,
$\bar{\sigma} = 0{\cdot}767$ mm,

例 1.　豆の一種, Schwarze belgische Kruppbohnen の種子 1,522 個を計測して,二様の度數分布表を作つた. 第29表がそれである. 甲と乙とは階級幅を同一に取り,唯階級境界の値を變更した. その結果,乙では M は階級中央の値となつたが爲に,見かけ上も相稱であるが,甲ではさうでない爲に,一見不相稱に見える. これなどは明に,似て非なる不相稱分布である. 甲表及び乙表第三列にこれと同一の M と σ と N とを有する理論的の正常分布を併せ記した. その計算法は第三編にとくこととするが,この正常分布も甲表では一見不相稱であることに注意せよ.

第 30 表

朝鮮人學童の肺活量の度數分布.

年齡滿 6 歲より滿 17 歲に至る.

$M = 1746$ cc. (♂), $M = 1412$ cc. (♀).

階　級	度數 ♂	度數 ♀
600	—	3
700	2	3
800	13	18
900	19	19
1000	25	41
1100	27	33
1200	31	30
1300	38	41
1400	46	41
1500	67	26
1600	46	19
1700	52	14
1800	49	14
1900	33	5
2000	36	11
2100	14	2
2200	14	3
2300	15	2
2400	8	—
2500	8	1
2600	5	1
2700	4	—
2800	7	—
2900	2	—
3000	3	—
3100	1	—
3200	2	—
計	557	327

5. 二種以上の分布の合成のために起る不相稱型

屢發見する不相稱型の中には,多數の正常相稱型の合成せる場合がある.

例 2.　第30表は學童の肺活量の度數分布表であるが,年齡を一定にして居ない. 從て,この分布はことなる多數の**原素分布**の合成であるが,不相稱の原因は恐らく共處にあるだらうとは,何人も氣が付く所である. また第一章例 8(第10表)に於ける牛乳搾取量の度數分布も,牛の年齡や,牛の種類につきて不純なる材料の混合であることが分布の不相稱の原因であらう. 而して不相稱の程度や傾斜の方向,平均や散布度等は,合成せる原素分布の割合によつて,著しく影響を受け,一般に不定であつて,殆ど研究の對象とする價値がない. 純粹が研究材料の持つ重要な性質であることを銘記せよ.

6. 眞の不相稱型

先天的又は後天的條件を同一にして一群の個體を觀測しても,尙ほ,不相稱型の現れることがある.

例 3.　3 年間試驗床に蒔いて選擇を行へる Braune Bohnen の純系を,同一試驗床に蒔き,出來るだけ同一の條件の下に培養し結實せしめ,採取せる5,000個の種子の長徑を計測し,千分比分布表を作つたが,第31表となつた. 實測の結果は著しく不相稱である. 決して見掛け上の不相稱ではなく,又材料の不純の爲に起るのでも無い. 第三列に記入せる理論的正常分布と比較すると,甚だ相違することが明とならう.

正常分布は $(p+q)^n$ なる二項式の展開に於て, $p=q$, $n=\infty$ として導いたのであるが,若し $p \neq q$ である時は, n は有限なる限り,不相稱である.

第 31 表

Braune Bohnen の純系の種子の長徑度數分布.

階級 (mm.)	度數 (實驗的)	度數 (理論的)	階級 (mm.)	度數 (實驗的)	度數 (理論的)
10·0—10·5	1	—	14·5—15·0	246	204
10·5—11·0	3	—	15·0—15·5	175	145
11·0—11·5	6	1	15·5—16·0	77	78
11·5—12·0	8	3	16·0—16·5	16	31
12·0—12·5	17	14	16·5—17·0	2	9
12·5—13·0	30	41	17·0—17·5		2
13·0—13·5	68	95			
13·5—14·0	145	165			
14·0—14·5	206	212			

$M = 14{\cdot}4429\,\mathrm{mm}$,
$\bar{\sigma} = 0{\cdot}911\,\mathrm{mm}$.

例へば, $p=\frac{2}{3}$, $q=\frac{1}{3}$ とすれば,

$$(p+q)^1 = \frac{1}{3}(2+1),$$

$$(p+q)^2 = \frac{1}{3^2}(4+4+1),$$

$$(p+q)^3 = \frac{1}{3^3}(8+12+6+1),$$

$$(p+q)^4 = \frac{1}{3^4}(16+32+24+8+1),$$

$$(p+q)^5 = \frac{1}{3^5}(32+80+80+40+10+1),$$

$p \neq q$ なる時の $(p+q)^n$ 展開の各項はかく不相稱で,或はJ-形分布となり,然らずとも高度の不相稱分布となるのである.その狀態が生物學に出て來る不相稱分布型に似て居るから,不相稱型を二項式の p と q との不等なることより說明を試みた

時代がある (Quetelet). しかし,この考は誤であることが明になつた. 少くとも連續變數の分布における不相稱型には適用することが出來ない. その理由は次の通りである.

正常分布を $\left(\frac{1}{2}+\frac{1}{2}\right)^n$ の展開であらはすことが出來るのは n が無限大なる極限に於てなしうることである. 不相稱分布を $(p+q)^n$ から導かん爲に, n を無限大にする時は, p と q とが等しからずとも相稱分布となつてしまふのである. 前頁に示した二項係數が, n 漸次大となるに順ひ,不相稱の程度を減少したのも其の一端を示すものであらう. 今 $1000\left(\frac{2}{3}+\frac{1}{3}\right)^{40}$ を展開し,各項の小數點以下を四捨五入する時は次表の如くなる.

第 32 表

$1000\left(\frac{2}{3}+\frac{1}{3}\right)^{40}$ の展開の各項と正常分布との比較.

項の番號	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
$1000\left(\frac{2}{3}+\frac{1}{3}\right)^{40}$ の開展の各項	1	2	6	13	27	48	75	102	123	133	128	111	87	61	39	23	12	6	2	1
正常分布	1	3	7	14	27	47	72	98	121	133	130	114	90	63	39	22	11	5	2	1

第一項より第四項まで,及び第25項以下は數値小であるから,省略した(Kapteyn). これを見ても第二行の$(p+q)^n$の展開は n が大なる時,正常分布に近づくことが明になる. n が益大になる時は益接近するから,生物統計學に於ける不相稱を p と q との不等なることより說明が出來ない.

Kapteyn は生物の發育には自ら限度があつて, Goethe も云へる通り「樹木は天まで生長するものにあらず」人體の榮養の如き,又次表に示す大麥の缺殼(Schartigkeit)の如き,生物の數量的現象

で取り得る値は一定の制限あるを免れないことに歸して居る.この制限に近づく時は發育は自ら抑制せられるが爲に,不相稱を呈するのであらうと云ふ.

第 33 表

大麥の或る純系の缺殼百分比の度數分布表.

Schartigkeit (%)	0—5	5—10	10—15	15—20	20—25	25—30	30—35	35—40	40—45	45—50	50—55	55—60	60—65	65—70	70—75	75—80
1902	181	441	292	97	11	7	—	2								
1903	4	30	77	99	107	76	31	13	2	2						
1904	4	9	16	40	80	87	111	112	75	86	40	25	16	9	6	1
1905	91	160	69	16	3	—	1	1								

Schartigkeit が 0 になることは困難である故に,又 0 より小となり得ないが故に,成績の良い年では (1902 年 1905 年) 分布が一方の極端 0 に接近して居るため,甚しき不相稱型を示した. しかし分布曲線が右の方に偏して居る年では,不相稱が輕度であつて,正常に近いのである.

7. 不相稱の程度の計算

散布の狀態を簡單に表示する爲に,σ が計算せられたが不相稱型の程度を簡單に表示する數値が無いであらうか. 一つの値で不相稱の方向と程度とを表示することが出來れば便利であり,又それが必要なることである.

(i) 不相稱分布なる時には,算術的平均と中央値と形態とが一致せず,M_0 は曲線の最高所にあるが,M は緩傾斜の脚に偏し,Med. は M_0 と M との中間にあるものである. 而して M と M_0 の差が大なるほど不相稱が甚しいから,標準偏差を單位として $M-M_0$ を計つたものを**歪度** (skewness ; Die Schiefheitziffer) と云ふ.

卽ち歪度を S' にてあらはせば

$$S' = \frac{M-M_0}{\sigma} = \frac{3(M-\text{Med.})}{\sigma}, \quad \cdots\cdots\cdots\cdots \text{(IV. 1)}$$

S' には正と負とあり,急傾斜の脚が左方にある時は S' は正である(第 20 圖). 若し,反對に急傾斜脚右にある時は負である.

(ii) 又不相稱分布なる時は,中央値を中心として兩側の分布の密度が異ふから,Q_2 と Q_1 とは同一でない. この事を應用して,Q_2-Q_1 を,Q を單位として計るのである. これを歪度 S'' とすれば,

$$S'' = \frac{Q_2-Q_1}{Q} = \frac{q_3-\text{Med.}-(\text{Med.}-q_1)}{Q}, \quad \cdots\cdots$$

卽ち,
$$S'' = \frac{q_1+q_3-2\,\text{Med.}}{Q}, \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \text{(IV. 2)}$$

S'' は正又は負であつて,急傾斜脚が左方にある時は S'' は正,然らざる時は負である. 又 $|S''|$ は最小 0 最大 2 であるが,普通は

$$-1 < S'' < +1$$

である. 但し,S'' は S' に比して,不相稱に對して稍鈍感であると云ふ缺點がある.

(iii) 最も合理的な歪度は偏差の三乘から計算するものである.

今,算術的平均 M を原點としたる各觀測値の偏差を a とすれば

$$\frac{1}{N}\sum_{i=1}^{N} a_i = 0,$$

又
$$\frac{1}{N}\sum_{i=1}^{N} a_i^2 = \sigma^2$$

であつた(第二章及び第三章). 然らば, $\sum_{i=1}^{N} a_i^3$ は如何なる値を取るであらうか. 相稱なる分布型では $\Sigma a^3=0$ となり,不相稱なる分布では, Σa^3 は正又は負の値を取る. 急傾斜脚が左方にある時は Σa^3 は正となり,反對の時は負となる.

$$\mu_3 = \frac{1}{N}\sum_{i=1}^{N} a_i^3$$

又,度數分布表よりは,

$$\mu_3 = \frac{1}{N}\sum_{i=1}^{K} f_i x_i^3$$

として, μ_3 を計算することが出來るが,不相稱の程度は μ_3 に關係すること勿論である. しかし μ_3 はそのまま歪度と出來ない. μ_3 を σ^3 を單位として計りたるものを歪度 S と云ふ.

即ち,

$$S = \frac{\sum_{i=1}^{N} a_i^3}{N\sigma^3} = \frac{\sum_{i=1}^{K} f_i x_i^3}{N\sigma^3} = \frac{\mu_3}{\sigma^3} \qquad \cdots\cdots\cdots\cdots\cdots(\text{IV. } 3)$$

歪度 S の計算は次の通り考へてもいい. 絶對の偏差 a は一の分布における偏差の程度を明確に表はすものでないから,偏差を,標準偏差を單位として計る. これを關係偏差と云ふ. 關係偏差を R とすれば,

$$R = \frac{a}{\sigma},$$

そこで R^3 の總和を N で除して,

$$S = \frac{1}{N}\sum_{i=1}^{N} R_i^3 = \frac{1}{N\sigma^3}\sum_{i=1}^{N} a_i^3 = \frac{\mu_3}{\sigma^3}$$

とする.

μ_3 の計算にも簡便なる方法がある.

a を算術的平均を原點とせる時の觀測値の偏差, α を任意の

原點 A よりの觀測値の偏差とし, $A-M=d$ とすれば,

$$V-M=a, \qquad V-A=\alpha,$$

$$a-\alpha=d,$$

$$\sum_{i=1}^{N}\alpha_i^3 = \sum_{i=1}^{N}(a_i-d)^3,$$

$$= \sum_{i=1}^{N} a_i^3 - 3d\sum_{i=1}^{N} a_i^2 + 3d^2\sum_{i=1}^{N} a_i - Nd^3,$$

而して, $\sum_{i=1}^{N} a_i = 0,\ \frac{1}{N}\sum_{i=1}^{N} a_i^2 = \frac{1}{N}\sum_{i=1}^{N}\alpha_i^2 - d^2$ なるを以て,

$$\frac{1}{N}\sum_{i=1}^{N}\alpha_i^3 = \frac{1}{N}\sum_{i=1}^{N} a_i^3 - \frac{3d}{N}\sum_{i=1}^{N}\alpha_i^2 + 2d^3,$$

$$\frac{1}{N}\sum_{i=1}^{N} a_i^3 = \frac{1}{N}\sum_{i=1}^{N}\alpha_i^3 + 3\frac{d}{N}\sum_{i=1}^{N}\alpha^2 - 2d^3,$$

故に,

$$S = \frac{\mu_3}{\sigma^3} = \frac{\nu_3}{\sigma^3} + 3d\frac{\nu_2}{\sigma^3} - 2\frac{d^3}{\sigma^3} \qquad \cdots\cdots\cdots(\text{IV. } 4)$$

となる[1].

但し,

$$\nu_3 = \frac{1}{N}\sum_{i=1}^{N}\alpha^3,$$

$$\nu_2 = \frac{1}{N}\sum_{i=1}^{N}\alpha^2,$$

又觀測値が階級に分類せられてあり,且つ階級幅大ならざる時は,

$$\nu_3 = \frac{1}{N}\sum_{i=1}^{K} f_i x_i^3,$$

[1] この式では, $A-M=d$ と定義したるが故に, (IV. 4)式となつた. 若し, $M-A=d$ と定める時は, これと異なる式となる. 常に d の附號に注意すべし.

$$\nu_2 = \frac{1}{N}\sum_{i=1}^{K} f_i x_i^2,$$

で計算する.

8. J-型分布

不相稱が甚しくなる時は, **J-型分布** (J-shaped distribution; Die einseitige Variationsreihe) となる. 第一章にあげた Ranunculus の花瓣の數の分布などその例であるが,一般に花瓣の數にこの種の分布が多い.

例 5. *Weigelia amabilis* の花瓣の數の度數分布 (de Vries) なども此の例であらう(第34表).

第 34 表

花瓣數	3	4	5
度數	61	196	888

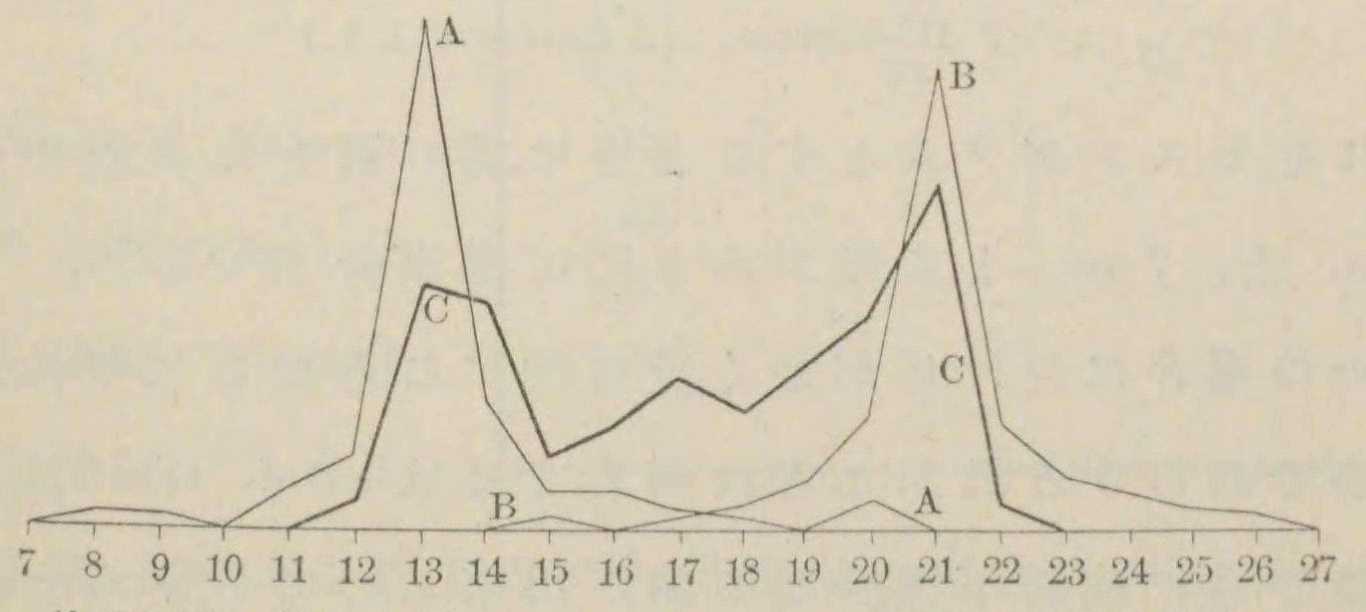
第三項 雙峯及び多峯分布型

9. 雙峯分布型と多峯分布型

度數分布は1個の峯を有するのが普通である. 稀に峯(分布曲線の極大)が2個又は3個以上あることがある. これを雙峯及多峯分布 (Die dimodale u. polymodale Variationsreihe) と云ふ. この型は多くは基本型の合成に分解することが出來るのである.

例 6. de Vries は諸地方の植物園から, *Chrysanthemum segetum* なる野生菊の種子を集めた. これを蒔いて生じたる莖の頂に咲いた花の舌狀花の數を數へて,度數多角形を作つたところ,第30圖のC曲線となつたのである. この分布には, 13 と 21 とに峯があり,明に雙峯分布である. この二つの峯が何故に出來たか

は, de Vries の數年の培養の結果明になつた. 統計は莖の頂に咲いた花のみにつきて舌狀花數を計へることとし,舌狀花數が12及び13なる花を殘し,他は悉く開花と同時に拔きすてて,交配を豫防したのである. かくて舌狀花12及び13なる花より,採種して蒔きたるものよりは,分布多角形Aを得た. 分布の頂は13にある單峯分布となつたのである. 又その翌年には別に舌狀花數21なるものを殘して,他を拔き棄て21なるものより採種して播き,これより舌狀花數の統計を取つた所度數多角形Bとなつた. これは分布の頂が21にある單峯分布である. 而して,AとBとの合成は略C多角形となることが明である. 即ち,最初彼が諸地方より集めたる種子は,二變種の混合であつて舌狀花數13を中心として分布するを特性とする一變種と,舌狀花數21を中心としてその兩側に分布する一變種との混合であることが明になつた.



第 30 圖 *Chrysanthemum segetum* の二峯分布とその單位分布.
C は合成分布, A 及び B は單位分布 (de Vries 原圖).

生物界には,かくの如く遺傳的に異なる二變種(若は品種),又は多數變種の混合せる場合に,雙峯分布又は多峯分布を示し,これが淘汰の結果單一なる變種に分離しうる場合が多くあらうと

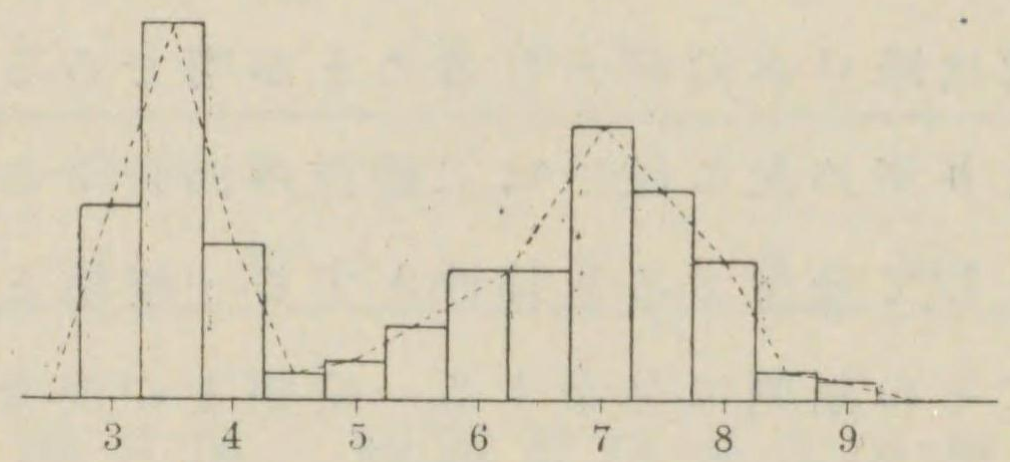
信じられる.

例 7. Bateson は Northumberland 附近の Farne island に住むハサミ蟲 (*Forficula*) を捕へ,その雄 584 個の鋏の長さを度數分布表としたのが第35表である.

第 35 表

Forficula の雄 584 個の鋏の長さの度數分布 (mm.).

鋏の長さ	3	3·5	4	4·5	5	5·5	6	6·5	7	7·5	8	8·5	9
度數	64	125	52	7	12	24	42	42	90	68	44	8	6



第 31 圖 *Forficula* 雄の鋏の長さを表はす Histogram. (Johannsen より.)

これをヒストグラムとするときは,第31圖となり,鋏の長さ 3·5 mm. と, 7 mm. とに峯があり,明に雙峯分布である. これは de Vries の場合に於けるが如く,遺傳的にことなる二變種が存在するのであらうと云ふことは極めて疑はしい. Giard は一種の寄生蟲の爲に成長不充分なる鋏が出來るのであらうと說明して居る.

例 8. も一つ例をあげる, 見波は家蠶の繭の重量を計測して分布表を作つた所,第 36 表第二列となつた. 繭の重さ 900 の階級內に形態のあること及び1,150の階級內に第二の形態があ

るから雙峯分布である. 然るに出て來る蛾の性に由つて分けて統計すれば第三列及び第四列となる. この二つは共に輕度の不相稱ではあるが單峯分布であつて,一は 900 の階級內に頂を有し,他は 1,150 の階級內に頂を有す. 故に,初めの雙峯分布は合成の分布で明に單峯なる二分布に分離することが出來たのである (Der Geschlecht-dimorphismus).

第 36 表

繭の重量の度數分布.

階級	度數 (♂+♀)	度數 (♂)	度數 (♀)
600	4	3	1
650	3	2	1
700	13	11	2
750	25	21	4
800	57	46	11
850	173	163	10
900	303	290	13
950	272	252	20
1000	142	97	45
1050	80	16	64
1100	139	3	136
1150	162	2	160
1200	151	—	151
1250	61	—	61
1300	21	—	21
1350	5	—	5
1400	1	—	1
1450	3	—	3

10. 度數分布の合成

de Vries 又 Bateson, 見波等の例の如く淘汰または他の方法によつて,合成の分布をその原素分布に分離することが出來る場合はあるが,しかし度數分布を見ただけでは分離し得ない. 正確な實驗と觀測に由つて初めて可能なのである. また,多くは分離に多大の勞を要すべく,中には分離不可能なものも多くあらう. また合成分布は必ずしも雙峯或は多峯分布とならず,時には輕度の不相稱型,

又は正常分布型なることもある．人體の計測の場合がそれであつて,一般には條件をことにせる多數の個體群の計測を行ふのであるから,出來上つた分布は決して單純な分布ではなく,合成の分布である．しかもこの合成が不相稱分布なることも稀で,最も普通には合成分布も亦正常分布である．

例へば第一章例3(第6圖)に掲げた頭最大長徑の度數分布は日本人全體としては正常分布であるが,これを地方的に分つ時は各地方それぞれ異つた分布をするもので,第37表はその著しく異なる二地方の分布のみを取出したものであるが,その合成は雙峯分布に近い．しかも,この各地方の分布(小分布)も決して眞の原素分布でなくして,先天的並に後天

第37表

日本人男學生の頭最大長徑度數分布.

伊勢及び隱岐の部.

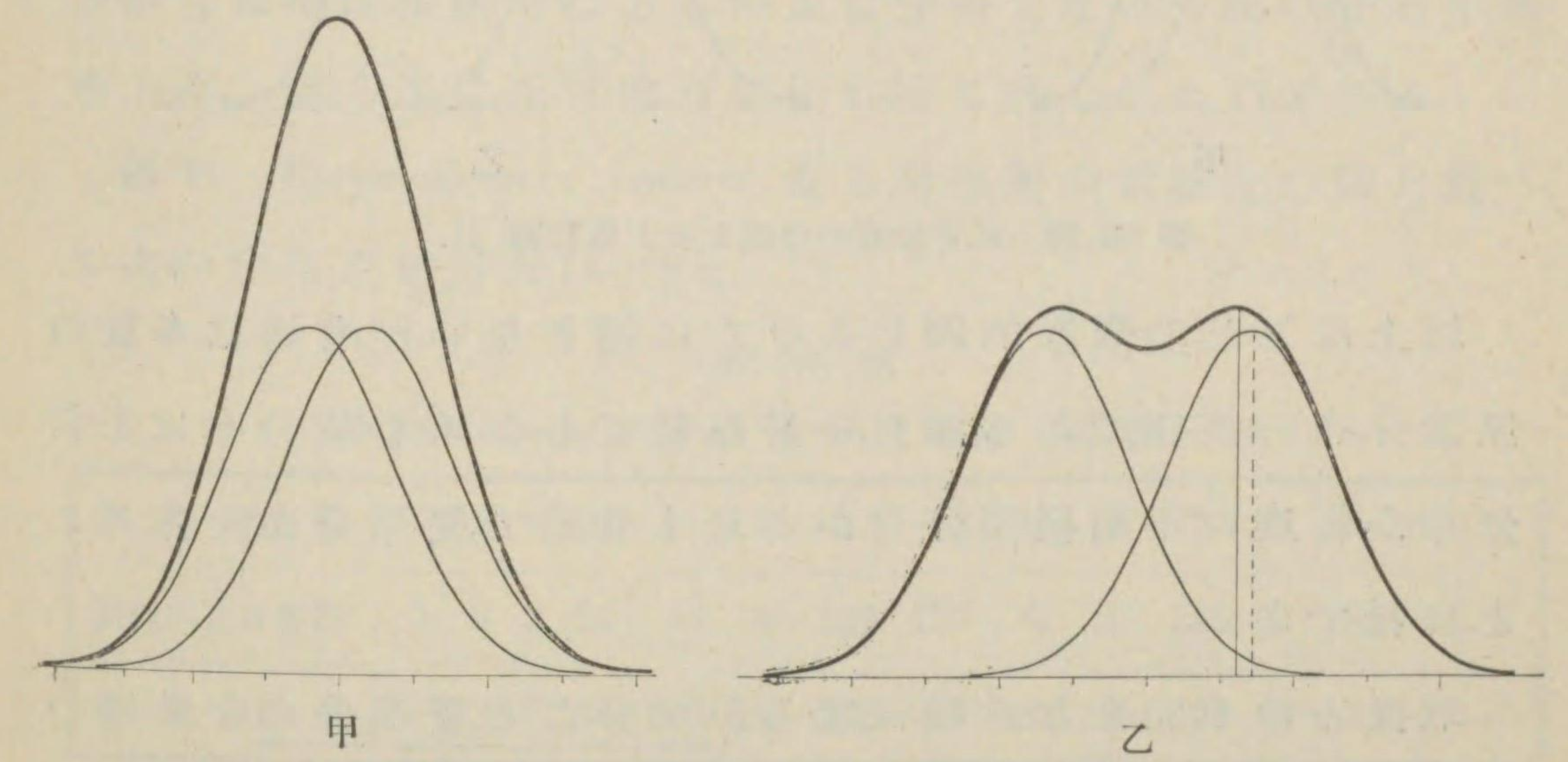
階級	伊勢	隱岐	計
× —170·5	2	—	2
170·5—172·5	—	—	—
172·5—174·5	—	—	—
174·5—176·5	1	—	1
176·5—178·5	8	—	8
178·5—180·5	19	2	21
180·5—182·5	17	2	19
182·5—184·5	15	3	18
184·5—186·5	23	4	27
186·5—188·5	20	7	27
188·5—190·5	18	8	26
190·5—192·5	15	12	27
192·5—194·5	7	11	18
194·5—196·5	4	14	18
196·5—198·5	3	7	10
198·5—200·5	1	9	10
200·5—202·5	—	—	—
202·5—204·5	—	2	2
204·5—206·5	—	2	2
206·5—208·5	—	—	—
208·5— ×	—	1	1
計	153	84	237

的條件を異にせる多數分布の合成から成るのであらう．多數の度數分布の合成で如何なる度數分布が出來るかは,興味の深い問題である．今それに深入りすることは出來ないが,最も簡單な場合を一二圖示して讀者の參考に供する．

(I) A, B 2個の原素分布が共に正常で,これを構成する個體が略同數なる時の合成分布．

(a) 2個の正常分布の中心が,相接近して居る際は,合成分布も亦正常分布である(第32圖甲).

(b) 2個の正常分布の中心が遠く相距つて居る時は,合成分布は相稱なる二峯分布となる．而して合成分布曲線の2個の峯(形態)は原素分布曲線の峯に一致しない,前者は分布の中央値に接近せんとする傾向を示す(第32圖乙).



第32圖 正常分布の合成を示す假想圖 I (水島原圖).

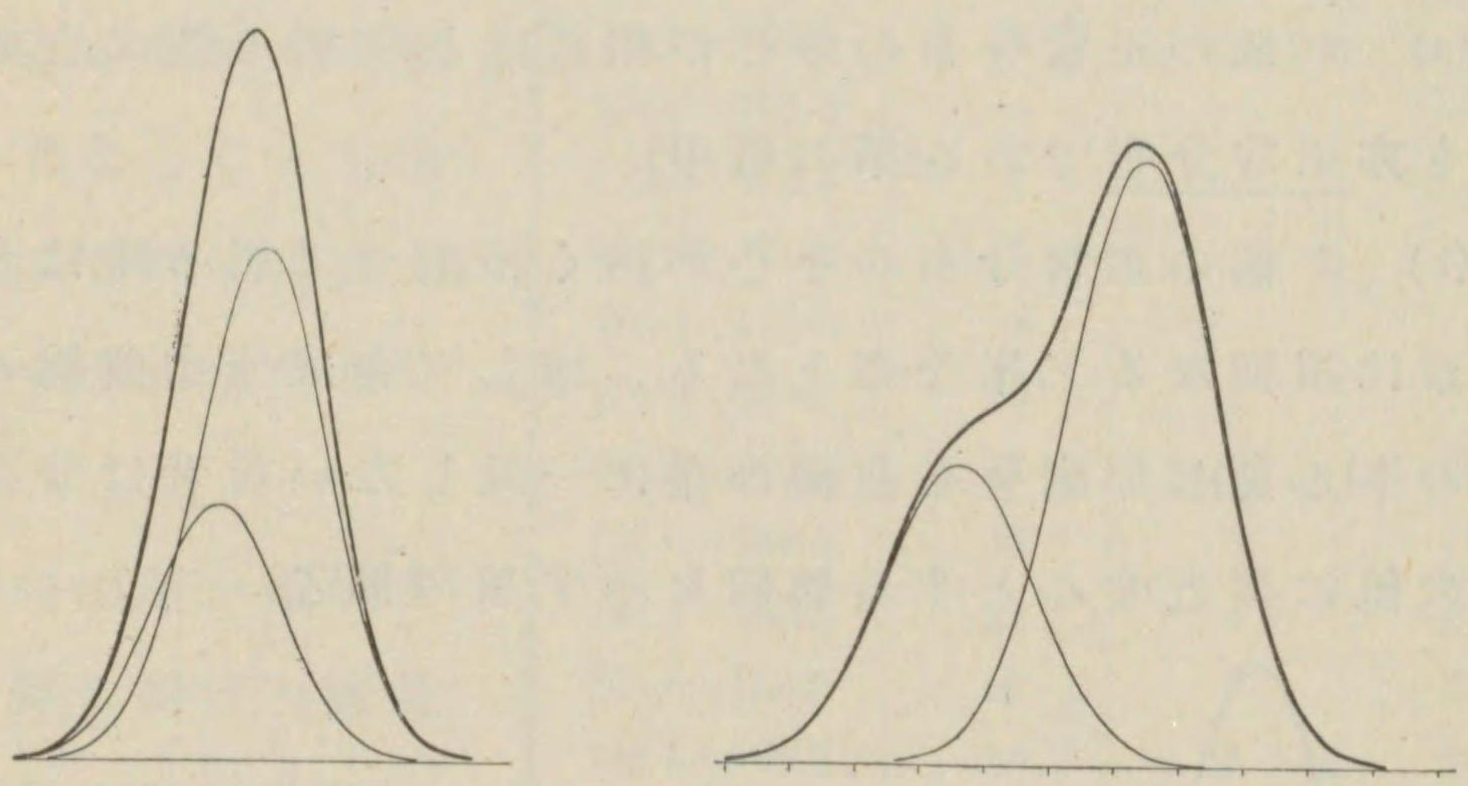
(II) A, B 2個の原素分布が共に正常で,これを構成する個體數が著しく相違して居る場合．

(c) 若し,二分布の中心が接近して居る時は極めて輕度の

不相稱型となる(第33圖甲).

(d)　若し二分布の中心が互に遠く距つて居る時は,不相稱型分布となる(第33圖乙).

(e)　若し,相距ること更に甚しき時は,相稱ならざる雙峯分布となる. この場合も合成分布の形態は移動することを免れない(圖省略).



第33圖　正常分布の合成を示す假想圖 II.

以上は二三の假設的例をあげたに過ぎないが,普通は多數の原素分布の集合であるから,一層複雜で,しかも合成分布は正常分布か輕度の不相稱型分布かが最も普通で,雙峯分布を作ることは稀である.

取扱方の不完全なる爲に起る似て非なる雙峯分布を,普通の雙峯分布と混同してはならぬ. 第一章第7節(iv)に說明せる如く,計測せる値を他の單位に換算せる場合に,屢この人爲的多峯分布が起る. 又人爲的ではないが觀測數が少いが爲に分布中央附近で,偶然の影響を受けて不定の振動をすることが極めて

多い. 階級幅を小に取る時ことに甚しい. 是も似て非なるもので眞の雙峯又は多峯分布と混同してはならぬ(第28圖). 故に,雙峯分布は輕率に決定することが出來ないのであつて,それには正常分布及び,不相稱なる分布や確率論についての深い知識を必要とする.

第四項　高峯分布型

11. 正常に近い多數の原素分布の合成は,正常分布に類することは屢說明したることである. しかし,合成分布にて,一原素分布を構成する個體數が特に多い場合には,第33圖乙に示した樣に不相稱の分布となることがある. 又,多數の原素分布の中中央の一分布が特に多數の個體を有し,周邊の分布の個體も亦少からぬ時は相稱ではあるが正常分布でない一種の分布が出來上る. 卽ち峯は正常分布に比し特に高くなるのである.

例 9. *Chrysanthemum segetum* なる野生菊の舌狀花の數を數へて次の分布表を得た (Ludwig).

第38表

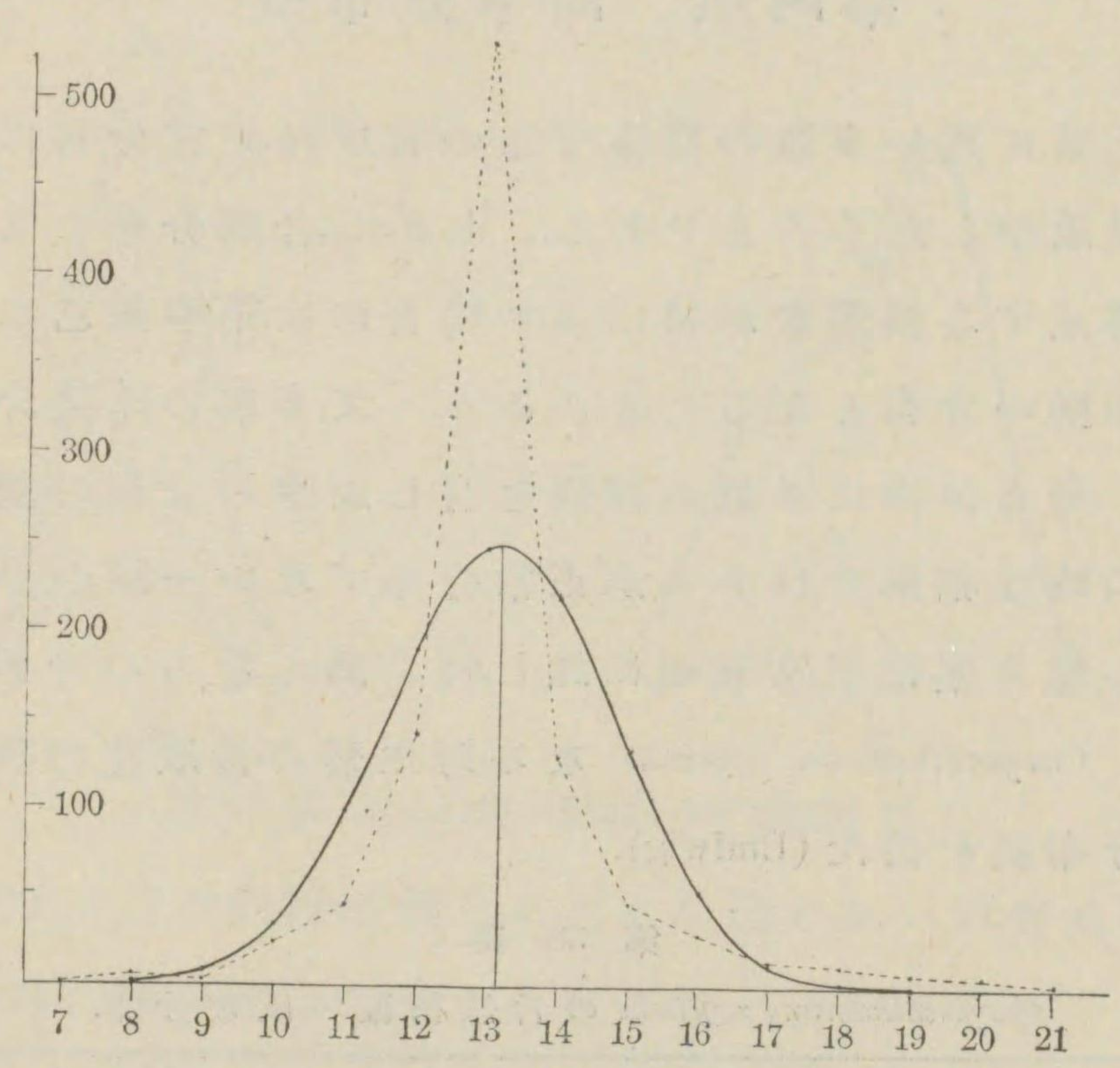
Chrysanthemum segetum の舌狀花數の度數分布.

舌狀花數	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
觀測による度數	1	6	3	25	46	141	529	129	47	30	15	12	8	6	.2
正常分布	—	2	9	37	100	188	243	215	132	55	16	3	—	—	—

$M = 13{\cdot}183,\quad \sigma = 1{\cdot}609.$

實際觀測による度數分布(第二行)を見るに,分布の最高所は舌狀花數13にある. 算術的平均も亦これに近い. $M = 13{\cdot}183$, $\sigma = 1{\cdot}609$.

しかるに,この平均とこの標準偏差とを有する正常分布は,第三行に示す通りであつて,實際の度數と大分異つて居る．即ち實際の度數分布は正常分布に比して,分布の中央が著しく高く,その兩側に於ては實驗上の分布は反對に低くなり,更に分布の兩端に至つて,再び實驗上の分布の度數が著しく高くなつて居る[1] (第34圖).



第34圖 *Chrysanthemum segetum* の高峯型分布 (Ludwig).

斯く同一の平均と同一の標準偏差を有しながら,正常分布と著しく相違し,就中分布の中央と兩端に於て特に高き度數を有する分布を高峯分布型(Die hochgipflige oder exzessive Variationsreihe)と云ふ.

(1) Ludwig のこの *Chrys. segetum* は de Vries の實驗に用ひたものの中Aなる品種即ち舌狀花數が少き方の品種で有つたのであらう.

高峯分布型は觀測個體數の多い場合と混同してはならぬ. Nの大なるが爲に分布曲線が高くなることは,第23圖に示した多數の曲線で理解し得るが,なほ觀測數の爲に理解の誤らるることを避けるには,百分比分布,又は千分比分布等に改めればいい. 次に高型分布はσの小なる分布と混同してはならぬ.

高型分布であることを決定するには,必ず同一のNと,同一のMと,同一のσとを有する正常分布を計算し,これを比較する. (その計算法は第三編を見よ.) 次に高峯分布型は如何なる場合に出來るかとの疑問が起る. 恐らく性型的(先天的),若くは現象型的(後天的)に單一ならざる分布に於て,その中央に位する一分布が他に比し,非常に多數の個體を含んで居る場合に起るものであらう.

例 10. 或る豆 (Bohnen) の純系につきて, 533 個の重量を計測して次の度數分布表を得た (Johannsen).

第 39 表

533 個の豆の純系の重量の度數分布.

階級	20—25	25—30	30—35	35—40	40—45	45—50	50—55	55—60	60—65	65—70	70—75	75—80 以上
實驗的度數分布	3	9	28	51	111	174	101	44	6	0	1	5
正常分布	1	7	31	67	119	135	101	50	17	4	1	—

$$M = 46{\cdot}51, \quad \sigma = 7{\cdot}675.$$

變數が連續なる場合の高峯分布型の例である.

12. 高峯分布型と反對に,低峯型分布がある. これは又負の高峯型分布と考へることが出來る (Die negative Exzessivität, Die tiefgipflige Form). これも多數の正常分布の合成であつて,就中分

布の中央を構成する原素分布の個體が比較的少數であり,中間の分布が多數を占むる樣な場合に起るのである. 若し中間の二分布のみが特に多數を占むる時は,第32圖乙の如く,雙峯分布となることがある. 例6, 例7, 例8の如き雙峯分布は皆一種の低峯分布である.

尙ほ U-型分布は2個の J-型分布の合成と見るべきであらう.

演習問題

1. 第一章例2(第4表)兵士の身長分布の度數多角形を描き,目分量にてその正常分布曲線を描き添へよ.

2. $p=\frac{3}{4}$, $q=\frac{1}{4}$ とし, $(p+q)^2$, $(p+q)^4$, $(p+q)^6$, $(p+q)^8$, $(p+q)^{10}$ $(p+q)^{11}$ の展開の各項を計算し,これを度數多角形として描畫せよ.

3. 第一章,例6(第8表)につきて, S', S'', S を計算せよ.

4. 第36表,伊勢及び隱岐に於ける男子頭最大長徑の分布を,度數多角形にて描畫し,この二つの合成分布をも描き添へて比較せよ.

参考文獻

本章に關するもの Yule, Johannsen の外,

1. Pearson, Karl :- Skew Variation in Homogeneous Material. *Phil. Trans. Roy. Soc., Serie A, Vol. 186, 1895.*

2. Bateson, W. :- Materials for the Study of Variation. London, 1894.

3. Kapteyn, J. C. :- Skew Frequency-curves in Biology and Statistics. Groningen, 1903.

4. MacLeod, J. :- The quantitative Method in Biology. Manchester, 1926.

5. de Vries, Hugo :- Die Mutationstheorie I. Leipzig, 1901.

6. de Vries, Hugo :- Einige zweigipflige Variationskurve. *Arch. f. Entwick.-Mech., Bd. 2, 1896.*

7. Vogler, P. :- Neue variationsstatistische Untersuchungen am Kompositen. *Jahrb. 1910 der St. Gallischen Naturwiss. Gesell. 1911.*

8. 水島治夫 :- 東京醫事新誌, 2726, 1931/V.

9. 　„　 :-　„　, 2734, 1931/VII.

第二編

相　　　關

第五章

相　關　表

1. 相關表

これまでは1個の變數についての統計法を記述したのであるが,本章以下では2個の變數を取扱つた場合の研究にうつる. 二つの變數を取扱ふことは,今までの統計法でも全く無いことはなかつた. 指數,例へば頭長幅指數では,頭最大長徑 Y と,頭最大幅徑 X との比を取り,その比の分布を研究することが出來た. しかし,これはやはり $100\frac{X}{Y}=Z$ として, Z を取扱つたのであるから,實は一變數である. 本章では,比 Z でなく, X と Y とを考へ,一の變數 X が他の變數 Y に如何に關係して居るかを檢討するのである.

相關の研究には,第一に**相關表**を作らねばならぬ.

例 1. 山東苦力1,096人につき,身長と腸骨上前棘高とを變數として,この二變數間の相關表を作つたのである(高牟禮). 階級の幅は,身長でも腸骨上前棘高でも,共に 2 cm. とした. 身長の大なるものでは棘高も亦大であらうとは,何人も考へる所であるが,後者は必ずしも前者に正比例して大とはならない. 然ら

第 40 表

山東苦力の身長と腸骨上前棘高との相關.

X は身長(階級幅 2 cm.), Y は棘高(階級幅 2 cm.)

Y \ X	146·45—148·45	148·45—150·45	150·45—152·45	152·45—154·45	154·45—156·45	156·45—158·45	158·45—160·45	160·45—162·45	162·45—164·45	164·45—166·45	166·45—168·45	168·45—170·45	170·45—172·45	172·45—174·45	174·45—176·45	176·45—178·45	178·45—180·45	180·45—182·45	182·45—184·45	184·45—186·45	計
102·45—104·45																		1	3		4
100·45—102·45															2		3	1		1	7
98·45—100·45													1	1	6	7	9	2	1		27
96·45—98·45												4	4	12	15	9	4	2			50
94·45—96·45										2	6	18	28	29	18	6	1				108
92·45—94·45									3	10	28	40	43	24	16	2					166
90·45—92·45								6	22	40	52	44	29	11	4	1					209
88·45—90·45					1	1	6	16	43	60	29	24	9	1	1						191
86·45—88·45				2		4	19	47	40	56	13	3	1		1						186
84·45—86·45				2	1	9	29	22	14	12	2										91
82·45—84·45				2	5	9	14	7	1	2											40
80·45—82·45			1	4	2	3		4													14
78·45—80·45			1		1																2
76·45—78·45																					
74·45—76·45	1																				1
計	1		2	10	10	26	68	102	123	182	130	133	115	78	63	25	17	6	4	1	1096

$M_x = 167{\cdot}208,\quad M_y = 90{\cdot}744,$

$\sigma_x = \ 5{\cdot}590,\quad \sigma_y = \ 4{\cdot}073.$

$r = \ 0{\cdot}847,$

ば,この二つの値は如何に關係して居るであらうか？ 第40表を見ればその概略を知ることが出來る. 身長の階級を一定にすれば,その階級內では棘高は一定の分布を示し,しかも正常分布で有つて,分布の中心は,身長の大となるに從ひ亦漸次大となるのである. 又上前腸骨棘高の一階級を取るも,各階級の身長は正常分布型で,棘高を大にすれば,その階級に於ける身長分布の中心も亦大となるのである.

かく,二變數 X, Y 間の度數分布の關係を示す表を**相關表**(Correlationtable; Der Korrelationstabelle), 又は**二重度數分布表**と云ふ.

ここに一つ注意すべきことは階級境界の値である. 例 1 では計測は 1 mm. を單位としたのであるから, 1465mm.—1484 mm. を一階級とした. この 1465 も 1484 も四捨五入した値だから,實はこの階級は 1464·5 mm.—1484·5 mm. としなければならぬ. 階級境界の値の端數はそれに由來するのであるが,この端數は後の計算に少しも障りとはならない.

X を第一變數, Y を第二變數と名づくる. 但し第一と第二變數とは,變更しても差支はないが,便宜上,上欄に記入するを第一とし,他を第二と定める. 又,縱の區分を**列**(Columns; Die Colonne) と稱し,表の左端の列には變數 Y の階級を記入する. 横の區分を**行**(Rows; Die Zeile) と稱し,上端の行には變數 X の階級を記入する. X が k 個の階級に, Y が h 個の階級に區分せられて居る時は, kh 個の小階級に分れ,その度數は相當せる區劃內に記入せられる. この區劃を**欄**と云ふ. 最終行には各列の度數の總計,最終列には各行の度數の總計を記入す. 卽ち,最終行は X

變數のみを取りたる時の度數分布表,最終列は Y 變數のみを取りたる時の度數分布表となる.

例 2. 第41表は鷄卵の重量とその孵化時に於ける雛體重との相關表である. X は雛體重, Y は卵重量(谷山). この表で X も Y も 2 gr. を階級幅としたものである. 例 1 と異つて居る點は X, Y ともに階級中央の値が與へられて居ることである. 從て境界の値は隣合つた二値の和を二分したものである.

第 41 表

鷄卵重量と孵化時に於ける雛體重との相關表.

X は雛體重, Y は卵重量(階級幅は X, Y 共に 2 gr.)

$Y \backslash X$	32	34	36	38	40	42	44	46	48	計	M_{ax}
67								1	3	4	—
65								2	3	5	47·2
63							3	3	2	8	45·75
61						6	20	4		30	43·87
59				1	9	24	16			50	42·20
57				1	24	15	2			42	40·86
55				18	19	5				42	39·38
53			7	14	4					25	37·76
51		1	12	4						17	36·35
49		8	10	1						19	35·26
47	4	9								13	33·39
45	5									5	32·00
計	9	18	29	39	56	50	41	10	8	260	—
M_{ay}	45·89	48·11	50·79	53·87	56·36	58·24	60·17	63·0	65·25	—	—

$M_x = 40{\cdot}053,$

$M_y = 55{\cdot}920,$

$\sigma_x = 3{\cdot}708, \quad \bar{\sigma}_x = 3{\cdot}663,$

$\sigma_y = 4{\cdot}723, \quad \bar{\sigma}_y = 4{\cdot}689,$

$r = +0{\cdot}931,$

第 42 表

父の身長(*X*)とその男兒の身長(*Y*)との相關.

(階級幅は共に 1 吋)

男兒＼父	58·5—59·5	59·5—60·5	60·5—61·5	61·5—62·5	62·5—63·5	63·5—64·5	64·5—65·5	65·5—66·5	66·5—67·5	67·5—68·5	68·5—69·5	69·5—70·5	70·5—71·5	71·5—72·5	72·5—73·5	73·5—74·5	74·5—75·5	
78·5—79·5														·25	·25			·5
77·5—78·5											1	1		·25	·75			3
76·5—77·5										1·25	·25	1			1·5			4
75·5—76·5										1·25	·25		·5	1	1			4
74·5—75·5										1	2		2·5	·75	1·75	·5		8·5
73·5—74·5					1		1·5	1·5		5·25	2·25	2·5	6·5	3·25	3·25		2	29
72·5—73·5							·75	·75	2·5	7·5	6·5	6	7·5	6·25	3·25	·5	·5	42
71·5—72·5						·25	3	1·25	7	7·75	10·75	11·25	10	8·5	2·75	·5		63
70·5—71·5						3·25	5	8·75	10·75	19	14·75	20·75	10·75	8	5	1	1	108
69·5—70·5					1	2·5	5·75	18·75	11·75	19·5	22·5	19·5	14·5	6·25	3·5	1·5	1	128
68·5—69·5			1		5·25	5	12·75	18·25	16	24	29	21·5	10	3·5	2·25		1	149·5
67·5—68·5			1·5	2	7·5	10	10·25	24·25	31·5	23·5	29·5	13·25	8·5	9·5	2·25			173·5
66·5—67·5		1·5	2	4·75	3·5	13·75	19·75	26·5	25·75	19·5	12·5	13·75	3·25	·5	1			148
65·5—66·5		·5	1	2·25	5·25	9·5	10	16·75	17·5	16	5·25	2	2·5	1				89·5
64·5—65·5	2	1	·5	2	3·25	9·5	13·5	10·75	7·5	5·5	3·5	2·5						61·5
63·5—64·5	1		1·5	3·75	3	4·25	8	9·25	3	1·25	1·5	·75	1·25					38·5
62·5—63·5		·25	·25	2·25	2·25	2	4	5	2·75	1·25		·25	·25					20·5
61·5—62·5		·25	·25		·5	1	·25	·25	·5	·5								3·5
60·5—61·5					·5				1									1·5
59·5—60·5					·5	·5	1											2
	3	3·5	8	17	33·5	61·5	95·5	142	137·5	154	141·5	116	78	49	28·5	4	5·5	1078

$r = +0·15.$

X の値の大になると共に, *Y* の値も亦大となり, *Y* の値の大となるに從ひ, *X* も亦大となる. しかも,その相關の程度は例 1 より一層高度である.

例 3.　父の身長と,子の身長を取りて相關表としたのが第42表である (Pearson & Lee).

例 1 は一個體の或る長さと他の長さとの相關であり,例 2 も亦これに近いが,例 3 は父と子とを一對とし,父の身長を *X*, 子の身長を *Y* としたのである. かくの如く一個體に屬せざる對の値の相關を研究することも甚だ多い. 階級幅は 1 吋,階級中央の値は整數であるから,階級境界の値は 0·5 の値を取る. 又計測値が二欄の丁度中間の値を取る事多く,その時は 0·5 を各の欄に加へ,又一對が四欄の中間の値を取つた時は, 0·25 を四個の欄に加へたのである. かく處置すれば,度數に小數が出來るので計算の不便甚しい. この不便をさける爲には,例 1 の如く階級を定めるがいい. また例 2 では境界の値が端數で無いに拘らず,度數に小數がないが,これは計測原表が精密な數字を與へて居たが爲に,卵重量や雛體重に實際整數値が無かつたのであらう.

例 3 に於ても明なる相關がある. 父の身長大となると共に,子の身長も概して大となり,子の身長大となれば,その父の身長も亦大となる. 但し,相關の程度は例 1, 例 2 の如く高度ではない.

例 4.　第43表は英國人男子の頭長徑と,視覺に對する反應時間との相關表で, Garton の計測したものを後に Harmon が整理

第 43 表

英國人男子の頭長徑 X(階級幅 0·1 吋)と視覺に對する反應時間 $Y\left(\text{階級幅 } \frac{2}{100} \text{ 秒}\right)$との相關.

Y \ X	6·805	6·905	7·005	7·105	7·205	7·305	7·405	7·505	7·605	7·705	7·805	7·905	8·005	8·105	8·205	8·305	8·405	8·505	8·605	計	M_{ax}	σ_{ax}
43·955										1										1	—	—
41·955																					—	—
39·955						1														1	—	—
37·955																					—	—
35·955											1									1	—	—
33·955							1													1	—	—
31·955								1			1									2	—	—
29·955					1				1	4	2	1	2	1						12	7·772	·225
27·955			1		1		1	5	5	2	4	2	2	2						25	7·673	·259
25·955				3	3	5	13	13	13	12	14	5	5	6	2					94	7·643	·259
23·955			1		7	10	24	22	58	45	42	24	19	11	5	1		1		270	7·695	·232
21·955		1	1	7	11	25	41	53	76	56	78	49	28	25	8	5		1		465	7·688	·256
19·955	2	1	5	13	37	58	122	148	176	196	215	138	81	63	31	9	3		2	1300	7·693	·256
17·955		5	4	11	22	43	76	143	191	201	177	146	83	53	22	11	1	2		1190	7·705	·244
15·955			4	2	14	37	67	115	131	138	145	121	71	39	17	6	1	2		910	7·714	·237
13·955			2	4	6	10	23	29	30	49	39	27	19	9	7	1				255	7·690	·248
11·955				1	1	1	5	9	10	15	15	5	8	6	4	2	1			83	7·770	·258
9·995			1	2	1	5	5	10	10	12	12	7	6	4	1					76	7·673	·255
7·995						1		1	1											3	7·47	—
5·995								1												1	—	—
計	2	7	19	43	104	196	378	550	702	731	745	525	324	218	97	35	6	6	2	4690	—	—
M_{ay}	—	18·81	18·27	18·79	19·30	18·81	18·98	18·51	18·86	18·53	18·72	18·44	18·44	18·83	18·43	18·41	17·62	18·96	—	—	—	—
σ_{ay}	—	—	3·85	3·61	3·21	3·51	3·23	3·22	3·16	3·29	3·23	2·86	3·26	3·31	3·09	2·71	—	—	—	—	—	—

$$M_x = 7{\cdot}699, \qquad \sigma_x = 0{\cdot}248, \qquad r =$$
$$M_y = 18{\cdot}669, \qquad \sigma_y = 3{\cdot}210,$$

したのである. 階級中央の値が與へられて居るが,この値に端數が出て居り,例 1 と類似して居る. 頭長徑では單位は吋,階級幅は 0·1 吋,視覺に對する反應時間の單位は $\frac{1}{100}$ 秒,階級幅は 2 單位である.

この表では頭長徑が大となつても,視覺に對する反應時間が殆ど大とならない. 又小ともならない樣である. 同樣に視覺に對する反應時間が大となつても,頭長徑は必ずしも大とならず,又小ともならない. 卽ち,相關が無いらしいのである. (相關の有無とその程度につきては次章を見よ.) 表に於て,

M_{ax} は各行分布の算術的平均, σ_{ax} は其の標準偏差;

M_{ay} は各列分布の算術的平均, σ_{ay} は其の標準偏差

である.

例 5. 燕麥の穀粒の重量とその脂肪含有量百分比との相關表 (Krarup).

この例では, Y は穀粒の重量であるが, X は穀粒の含有する脂肪百分比であつて化學的量を取扱つた一例である.

化學的操作が形態學的計測より困難な爲に,觀測數の少きは止むを得ぬことで,階級數も亦少數である. 穀粒の重量, Y の單位は mg. であるが,脂肪含有量は絕對量で表はされて居ない. 表を一覽すれば,穀粒重量を增大するに從ひ,脂肪含有のプロセントが反對に減少することが明になる. この點で例 1 乃至例 3 とは反對であり,又例 4 とも異つて居る[1].

[1] この相關表は例 2 と同じく,階級境界の値に端數が出ても居らず,度數に小數も出て居ない. 計測の際に注意して,變數が階級境界の値(例へば 50 mg.)を取る時は,精密な數字を記入して置く. 例へば 50·01 とか 49·99 であるとかを明記して置くのである. さうすれば 50 mg. を境界とした時,何れに編入すべきか迷ふ樣なことはない.

第 44 表

燕麥の穀粒の重量 Y (階級幅 5 mg.) とその
含有脂肪百分比 X (階級幅 0·5 %) との相關.

Y ＼ X	4·5—5·0	5·0—5·5	5·5—6·0	6·0—6·5	6·5—7·0	7·0—7·5	7·5—8·0	8·0—8·5	計	平均脂肪量
55—60			1						1	5·75
50—55		2	1	1					4	5·63
45—50		1	12	11	2				26	6·02
40—45	1	2	10	48	38	8	1		107	6·43
35—40		1	6	22	33	10	2	1	75	6·62
30—35					8	2	1		11	6·93
計	1	6	30	82	80	20	4	1	224	
平均重量	42·5	45·8	44·3	41·9	40·1	39·0	37·5	37·5	—	—

$M_x = 6{\cdot}453\,\%, \quad \sigma_x = 0{\cdot}516\,\%,$

$M_y = 41{\cdot}16\,\text{mg.}, \quad \sigma_y = 4{\cdot}145\,\text{mg.}$

$\bar{r} = 0{\cdot}488,$

例 6. 英國の貴族について,母の生んだ子供の數と,その娘の一人が生んだ子供の數との相關を表とした(第45表).

この例では, X も Y も共に子供の數であるから,不連續變數の例であるが,又統計する材料に注意を拂はなければならぬ一例である. 子供の數は,夫婦の結婚年齢や結婚繼續期間に影響せらるること著しいから,これらの影響を除く爲に,結婚繼續期間15年以下のものは表より除いた. 又,夫の結婚年齢30歳以上,妻の年齢25歳以上なるものも表から除くべきであるが,實際上制限を嚴格にすることは困難であつて,この表では妻の結婚年齢にも夫の年齢にも制限を附することは出來なかつた. 又, 1人の母に多數の娘が有つた場合,何れの娘を取るべきやを決定

することも必要である. 即ち,前記の條件に叶つたものは全部取るべきか,又は多數の娘中1人を取るべきか, 1人を取るとすればどの娘を取るべきか. 例へば1人の母に5人の子あり,その中3人の娘が前記の條件に叶つて居るとし, 3人中Aは子供3人, Bは0人, Cは4人を持ちたりとし,三組の値(5, 3), (5, 0)

第 45 表

母の生める子供數とその娘の生める子供數との相關表.
X は娘の生める子供の數, Y は母の生める子供の數.

母の子の數 ＼ 娘の子の數	0	1	2	3	4	5	6	7	8	9	10	11	12	13	計	M_{ax}
16			1												1	—
15													1		1	—
14									1		1				2	9·00
13					1	2	1	1	2				1	2	10	8·10
12	3	2		3	5	2	3	1	1	1	1				22	4·23
11	2	2	2	2	1	3	3	2	4	2			1	1	25	5·52
10	3	3	4	8	2	8	5	10	3	2	2	2			52	5·13
9	6	5	2	4	14	14	7	5	12	3	3	1			76	5·11
8	9	8	3	10	10	13	14	8	5	5	6		1		92	4·93
7	8	9	9	13	18	12	15	9	8	7	4		1		113	4·69
6	15	13	15	14	15	9	12	13	9	4	3	1		1	124	4·15
5	21	10	18	9	21	23	15	4	9	3	1	1	3	2	140	4·09
4	18	15	15	11	17	17	11	8	12	3	2	1	2		132	4·02
3	11	14	10	16	19	7	8	3	4	4	1	2	1		100	3·74
2	9	5	9	10	5	6	5	4	2				2		57	3·53
1	5	12	9	5	5	7	4	5	1						53	3·15
計	110	98	97	105	133	123	103	73	73	34	24	8	13	6	1000	
M_{ay}	5·39	5·10	4·95	5·63	5·80	6·13	6·22	6·45	6·77	6·85	7·63	6·25	6·46	8·83		

$M_x = 4{\cdot}335, \quad \sigma_x = 2{\cdot}984,$

$M_y = 5{\cdot}898, \quad \sigma_y = 2{\cdot}830.$

(5,4)の全部を表に載すべきかと云ふに,嚴密にはその中の一組を載すべきであらう. ここでは唯一組,就中最初に結婚した一組を選ぶこととしたのである. 若し,この選擇を誤る時は,誤れる相關を呼ぶことが有る. 或は故意に子供の多き娘を取るとか,故意に子供の少き娘を取るとか,或は故意に親の子供數に近きものを取るとか,又はその反對を取るとか,表の作製者の意志を加へ,その意志が相關に反映する如きことは極力避けねばならぬ. 統計は實に用意周到なる人のみが能く爲し得るものである.

又,結婚期間の長いものを取る代りに,長女,二番目等を取るも差支はない. この表に就いて見るに相關はあるらしい, 例1,及び例2程高度ではないが,類似の相關があるらしい. 又,この二重度數分布表で特色あることは,分布の中心が著しく一方に偏し,分布が不相稱なることである.

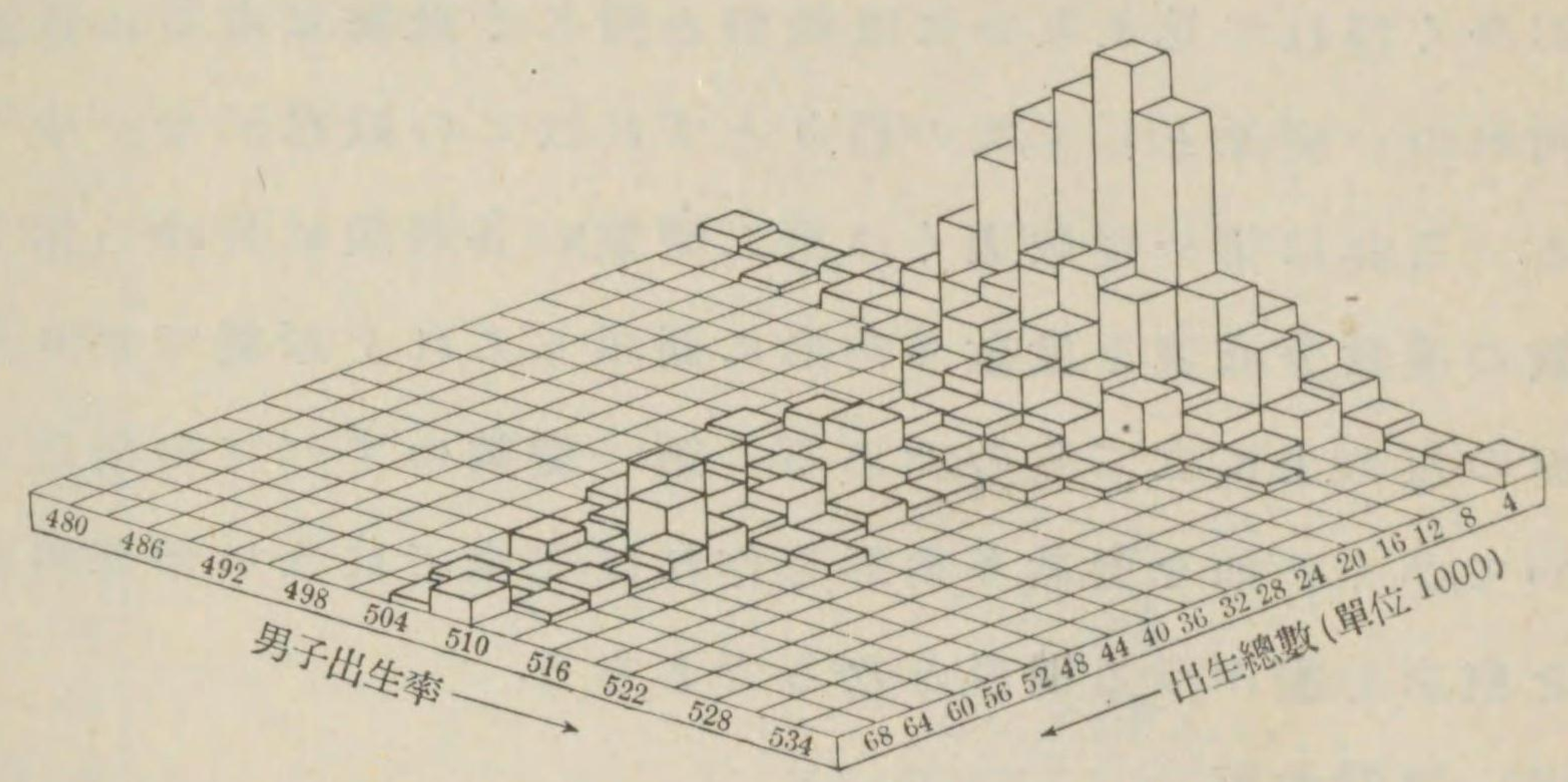
2. 相關の圖示法

單一の度數分布と同じく,二重分布はこれを圖示して視覺に訴へることは,極めて有效な手段で,理解を助けること大である. 單一度數分布は平面上に曲線又は屈折線として表はされたのであるが,二重分布表は當然立體となり,二變數は平面上に直交せる X 及び Y を軸とし,度數 Z は X と Y とに對して直立せる一定の高さの柱列等を以て表はさる. 從て,圖示は一般に困難である.

(i) ステレオグラム

これは Histogram に相當するもので,一平面上の直交軸 X, Y を

作り,これを階級數に等分し,碁盤目を作り,その相當欄に度數だ

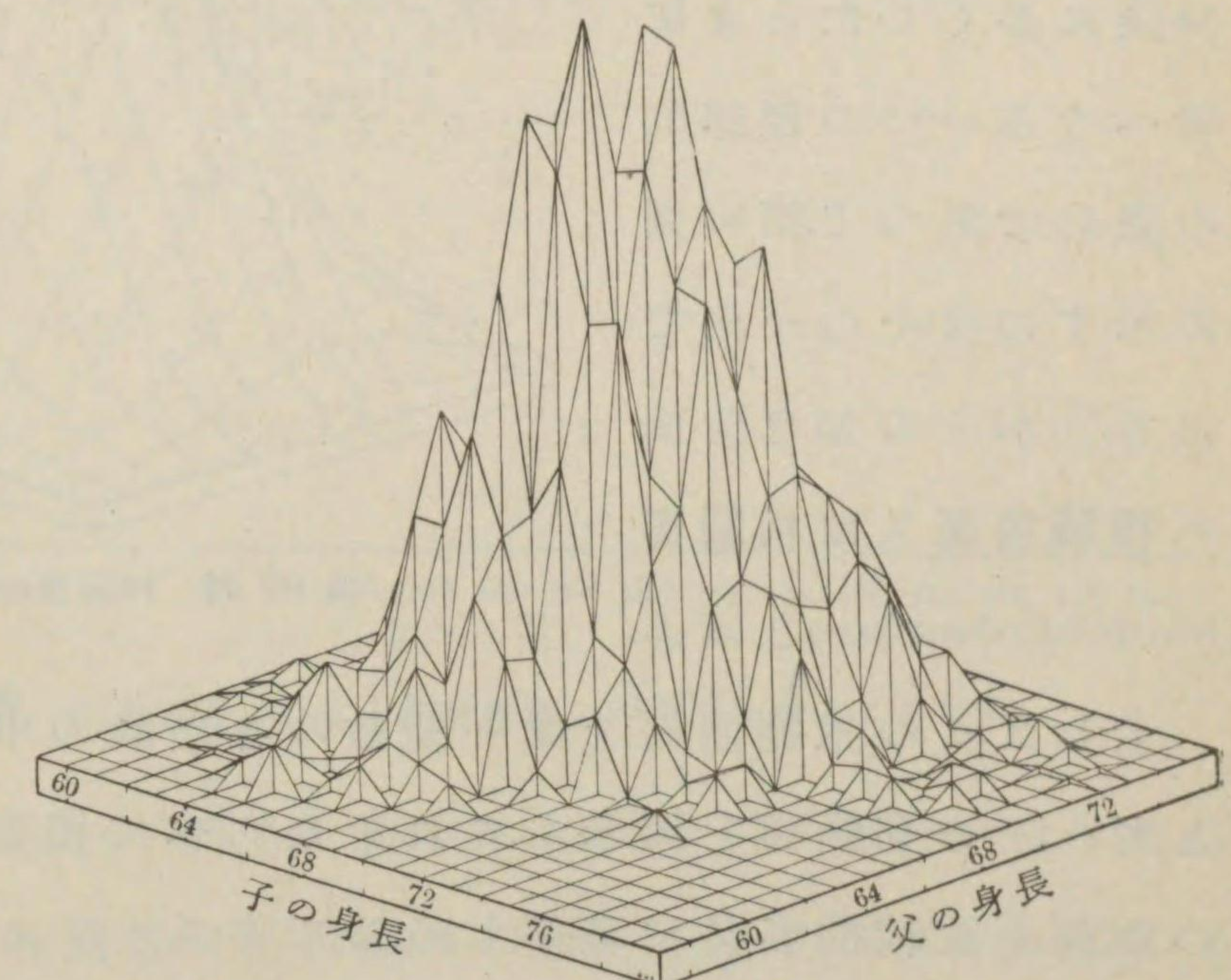


第35圖 ステレオグラム (例8,第47表)

けの高さの柱を立てた如く圖示する. これを Stereogram と云ひ,第35圖はその一例であるが,多くの場合分布の半ばが見えないから,これを概觀することが困難なる缺點がある.

(ii) 相關多角形

單一變數分布の度數多角形に類するもので,厚紙でピラミッド形の模型を作り,これを碁盤面の適當欄に立てるのである.

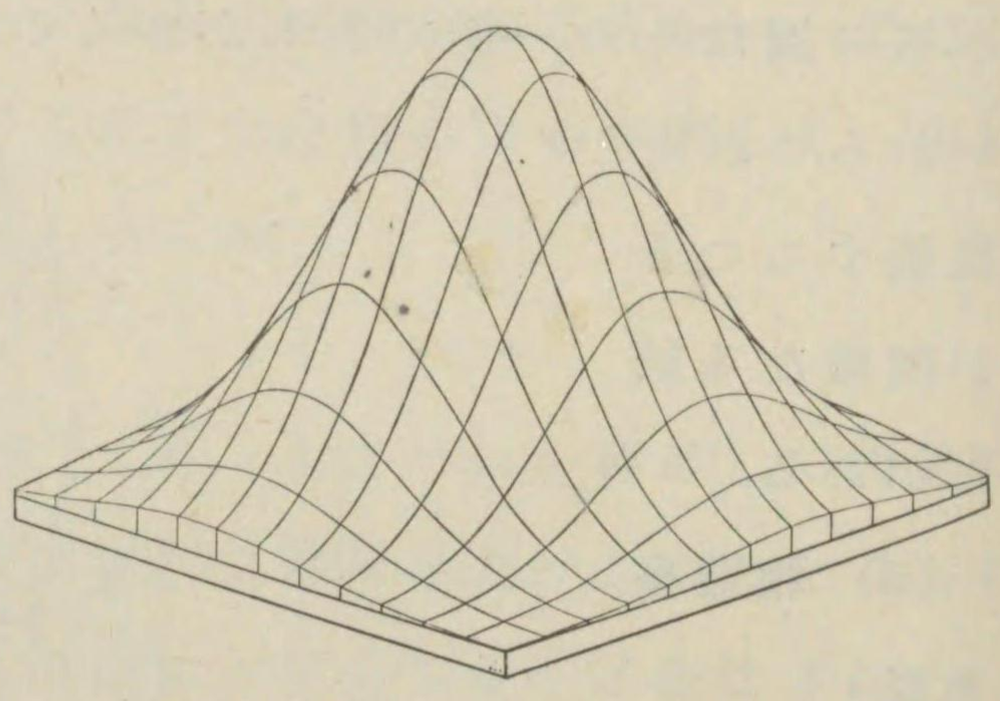


第36圖 父と子の身長の相關を表はす相關多角形 (Yule)

第36圖は第42表を立體的に圖示したものである. 度數の山の裏が全く隱れて見えないことは(i)と同じく缺點である. 相關多角形の一變法として,次の通りとすればこの缺點がやや少くなる. それは第一變數(若くは第二變數)の各階級に於ける第二變數の度數分布表を,度數多角形に圖示し,これを厚紙で切り碁盤面の板の上に立てるのである. 第一變數のすべての階級について,かくの如き度數多角形を作れば,完全ではないが眞相に近き概念を握みうる圖形となるであらう.

(iii) 相關曲面

單一變數の場合に於ける度數曲線に一致するものである. 第35圖に於ける Stereogram に於て,若し X, Y の階級幅を無限に小とし,同時に觀測數を無限に多くしたときに考へうる一つの理想的の圖形であつて,第37圖に示すのはその一例である. かくの如き曲面を**相關曲面**又は**相關面**(Correlationsurface)と云ふ.



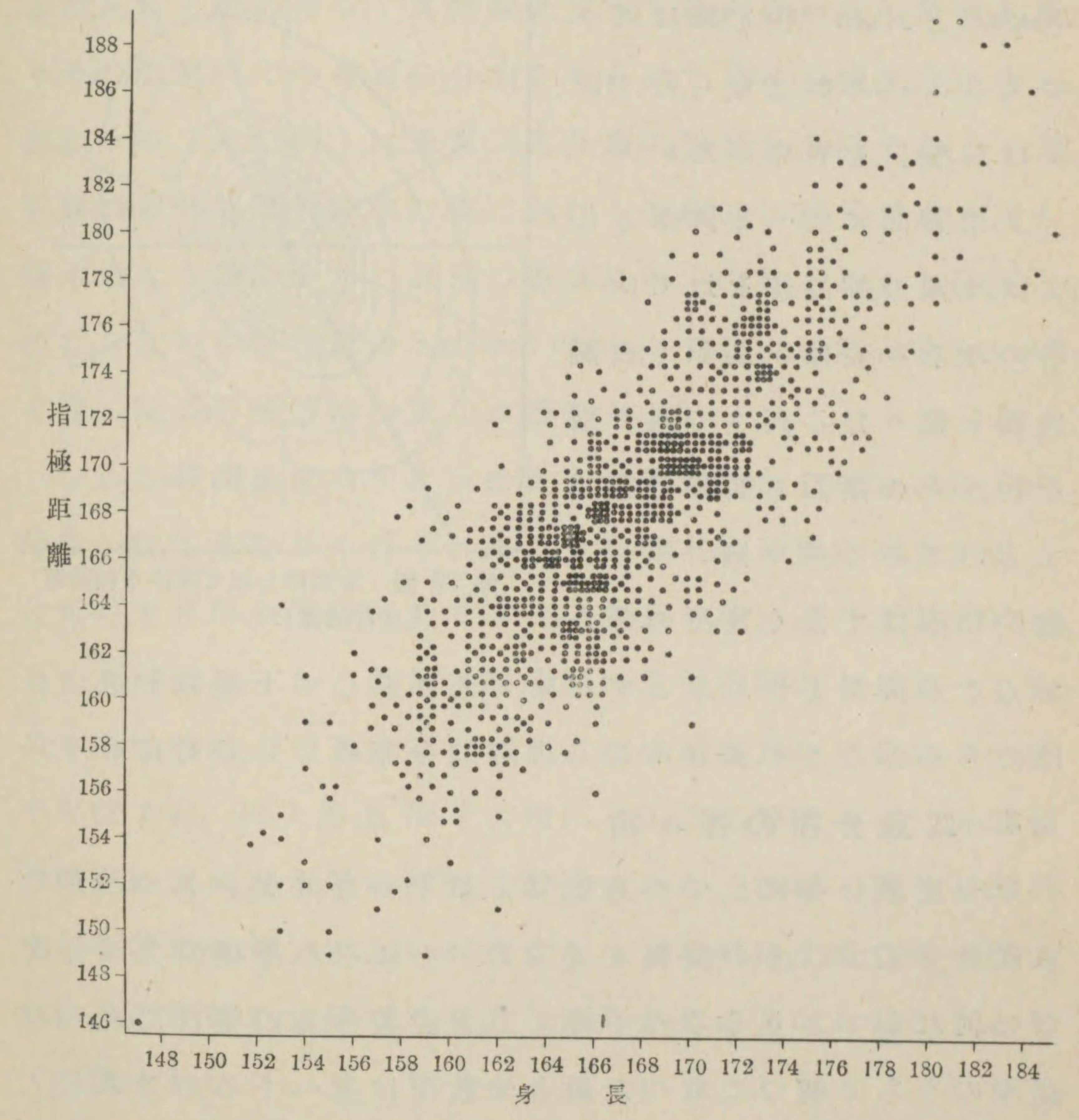
第37圖 相關曲面 (Yule)

若し又,この相關曲面を第一變數の各階級の中央で切斷したと考へ,その切斷面を圖示し,これをならべて模型とする時は,この曲面を比較的容易に理解しうるであらう.

(iv) 相關の平面圖示法

上記 (i) 及び (ii) の如く,相關表を立體に表現するには手數が

掛ること,立體の裏の見えないこと等,種々の不便がある. 若し,平面的に簡單に圖示し得れば便利であらう.

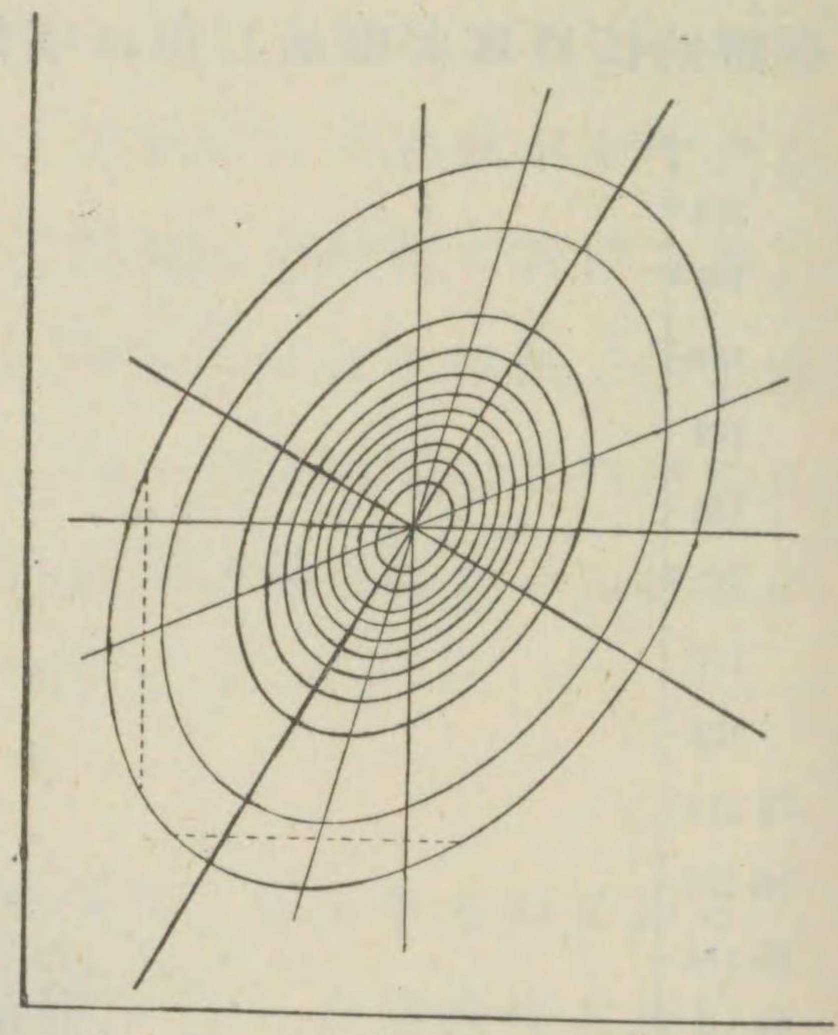


第38圖 山東苦力の身長と指極距離との相關を表はすグラフ.

第38圖の如く一平面上に直交する X 軸と Y 軸とを取り,この X 及び Y 上に各變數の階級境界の値を取り,この境界を通じて並行線を引く. 卽ち,欄を盤目に描き,各欄にそれぞれの度數だけの點を打つのである. 點の密度の大なる欄は度數の多い階

級である．度數を高さで表示せず,點の密度で表示したのがこの法である．これを Scattered diagram と云ふ．但し點は同一の大さに打たねばならぬ．(圖には碁盤目を省略した.)

又,相關面に代へる圖形としては,地圖で等高線を以て山や谷の地形を表はすが如く,相關曲面を種々なる高さで水平斷を行ひ,その斷面を曲線で圖示するのである(第39圖)．但し,理論的相關面でなく,實測の分布からこの圖形を作らんとすれば,可なり甚しい不規則を忍ばねばならぬから,この圖法は專ら理論的分布について行はれる。



第 39 圖　等高線を以て描ける相關面(正常相關).

3. 二重分布の種々相

單一度數分布に三つの基礎型と,若干の合成型があつたが,二重度數分布では,その變數 X 及び Y がそれぞれ基礎型又は合成型の何れかでありうるので,從て二重度數分布の型はこれらの諸型のうち 2 個づつ取つた組合せ數だけ考へうるのである．しかし,その中で殆ど出現しない型もあるから,今は重要なる二三の型について說明する．

(i) 正常二重分布(正常相關)

何れの變數の何れの一階級を取るも,その階級內で他の變數の度數分布が正常分布なる如き二重度數分布を云ふのである．

例 1 に於て,身長の何れの一階級を取るも,その階級內での腸骨前上棘高は正常分布である．偶然の爲に多少の不規則さはあるが,正常と認め得る．又變數を更へて,棘高の何れの階級を取り,その階級內での身長の分布をしらべて見ても,皆正常分布と認めうる．故に,例 1 は正常二重分布の適當なる例である．第 37 圖の相關曲面,及び第 39 圖は,何れも理論的に正常二重分布を圖示したものである．實測の値ではかくの如く正しき圖形とはならないが,第 38 圖の Scattered diagram で點の密度と等しい所を結ぶ時は,略楕圓形の重つた圖形となる．

しかし,觀測總數の少き爲に,實測の相關表からは,各行各列の分布が皆正常なりや否やを決定すること不可能な場合が多く,分布の中央部の階級に於ても,計測數少い時は甚だ不規則になることがある．かくの如き時簡單に正常相關なりや,否やを決めるには,普通 X の全分布及び Y の全分布が共に正常なりや否やを檢する．例 1 乃至例 4 は何れも正常相關である．正常相關圖形は常に或る軸に對して相稱である．相稱二重分布は必ずしも常に正常ではないが,正常ならざる相稱二重分布の例は生物界には極めて稀であるから,相稱分布は常に正常なりと考へて差支ない．

(ii) 輕度の不相稱なる二重分布

一變數の何れの階級を取りても,その階級內での他の變數の分布が輕度に不相稱である場合は,一方に偏したる相關面が出來る．かくの如き分布を輕度に不相稱なる二重分布と云ふ．例 6 及び例 7 では, X に於ける各階級, Y に於ける各階級,共に

不相稱な場合である．觀測數少い時は,各行又は各列の分布を見極めることは困難であるから，X 若くは Y の全分布について,その輕度に不相稱なりや,否やを決定すれば簡單である．例5も亦輕度の不相稱型である．しかし此の場合は X の全分布は正常であるが, Y の全分布は不相稱なるが爲である．

第 46 表

乳牛の年齡(X)とこれより1週間に搾取せる酪脂量(Y)との相關.

Y ＼ X	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	
13·5—14·0				1					1								2
13·0—13·5							1										1
12·5—				1													1
12·0—				1		1			1								3
11·5—					2												2
11·0—			1	3	1	1	3		1		1						11
10·5—		1	6	4	3	3	3		2								22
10·0—		6	15	11	6	11	4	3	4	5				1			64
9·5—	2	11	15	23	25	14	6	10	4	1							111
9·0—	1	27	40	23	23	28	11	14	5	4		2					178
8·5—	4	44	61	73	51	37	25	14	5	5	2	1					322
8·0—	7	80	81	80	22	44	25	27	18	5	5	1	1		1		437
7·5—	8	102	119	110	88	60	33	31	10	11	3	1	2			1	579
7·0—	12	152	149	122	92	74	32	37	21	10	5	4	2				713
6·5—	22	186	188	117	102	58	49	32	16	11	8	1			1		791
6·0—	13	176	166	96	74	36	39	27	18	10	2	2	1				660
5·5—	22	162	94	70	61	31	24	16	5	7	4		1		1		498
5·0—	12	118	67	45	25	11	16	7	5	3	1	2			1		313
4·5—	3	41	30	21	15	8	4	3	4	1	1			1			132
4·0—	5	18	13	10	4	2	1	2	1	1		1					58
3·5—4·0	1	4	1	1	2					1							10
3·0—3·5		2	1						1								4
	112	1129	1047	812	636	419	276	223	122	75	32	15	7	2	4	1	

例 7.　乳牛より1週間に搾取する乳の酪脂(Butterfat)の量と,牛の年齡との相關關係を檢討する目的で,相關表を作つた．第46表がそれである．年齡の各階級を取れば,脂肪量は殆ど正常

分布を呈して居るが,(極めて輕微なる不相稱)脂肪量の何れの階級を取るも,その階級內での年齡の分布は著しく不相稱であり,從て全體として分布の中心は著しく左に偏して居る．例6と共にこれも亦輕度の不相稱の一例である．相稱,不相稱と云ふことは相關の程度とは關係が無いのであつて,例6及び例7は不相稱であり,且つ相關の程度の輕微な例である．

(iii) 高度の不相稱なる二重分布

X の(若くは Y の)各階級に於ける Y の(若くは X の)分布が,甚しき不相稱なる度數分布を呈する時は, Y につきて(若くは X につきて)**高度に不相稱なる二重分布**と稱す．

例 8.　第47表は England と Wales とに於ける，1881 年より 90 年迄の延べ數 632 地區につきて,出生全數(Y)と男子出生率(X)との相關表である．

出生全數 4,000 以下の地方 149 地方あり,この階級內で男子出生率の分布は略正常である．Y の他の階級を取つても, X は略同樣に正常である．然るに,男子出生率(X)の一階級を取り,その階級內での出生全數の分布を見ればこれは甚しく不相稱であるから,この二重分布は高度の不相稱型である(第35圖)．

(iv) 不規則なる二重分布

單純なる不相稱とのみ云へない樣な不規則な合成二重分布が出て來ることがある．しかし,かくの如き場合は數個の二重分布に分離すれば,稍規則正しき分布にあらためることが出來る．

第 47 表

England と Wales に於ける 1881 年より 1890 年迄に於ける延べ數 632 地區につきて出生全數(Y)と男子出生率(X)との相關表.

(男子出生率(X)は千分比を以て表はし,出生全數(Y)は 1000 人を單位とす).

Y ＼ X	464·5—467·5	467·5—470·5	470·5—	473·5—	476·5—	479·5—	482·5—	485·5—	488·5—	491·5—	494·5—	497·5—	500·5—	503·5—	506·5—	509·5—	512·5—	515·5—	518·5—	521·5—	524·5—	527·5—	530·5—	533·5—	536·5—	539·5—	542·5—	
108—112															1													1
104—108															1													1
100—104																												—
96—100																2												2
92—96																												—
88—92															1	1												2
84—88															2													2
80—84															1													1
76—80																												—
72—76														1	3													4
68—72														2		1												3
64—68															1	2												3
60—64														3	1	1												5
56—60														1		2												3
52—56														2	5													7
48—52														2	6	3												11
44—48														2	5	2	1	1										11
40—44															3	4	1											8
36—40															4		2											6
32—36														3	3		1											7
28—32												1		2	6	2	1											12
24—28													2	3	5	3	1	1										15
20—24														1	5	4	3	1	1									15
16—20														5	6	9	4	1	1									26
12—16													1	10	8	12	8	6	1	1	1							48
8—12											2	1	5	15	17	18	16	7	4	1								86
4—8							1		2	2	5	7	20	27	29	42	36	18	10	4				1				204
0—4	1		1			2	1	2	2	4	8	9	18	19	12	21	14	12	9	5	3	1	1	2		1	1	149
	1	—	1	—	—	2	2	2	4	6	15	18	46	98	125	129	88	47	26	11	4	1	1	3	—	1	1	632

例 9. 第48表は英國人につきて,肺活量と年齢との相關表である. 一見極めて不規則であるが,これは種々なる年齢を取りたるが爲に起るのであつて[1],年齢 23 歳までは高度の不相稱型で,且つ年齢と共に肺活量が增大するのであつて,例 6, 例 7 に類する. 又, 24 歳より 50 歳まではこれも不相稱で,且つ相關の程度の低い,卽ち殆ど相關が無い分布である. 又, 51 歳以上でも輕度の不相稱型であるが,年齢の增大するに從ひ肺活量は却て減少するし, 23 歳までとは反對の相關が成立するのである. かく三時期により,相關關係は同一でないから,互に相殺して全體としては相關が殆ど無いかの觀を與へるが,實は相當に濃厚な相關が存在するものである.

4. 相關表作製に當りて注意すべき事項

度數分布表作製に當り注意すべき事項は,そのまま,相關表の作製に當りても適用すべきであるから,今これを再說しない. その他に注意すべき二三の事項をあげる.

(i) 二つの變數の各階級の幅が小にすぎる時は,表が甚しく厖大となり,取扱は困難となる. 故に,階級幅はなるべく大とし,從て階級數を少くし,以て計算や取扱を簡便にしなければならぬ.

X, Y 何れも階級數を 12—20 位に限つても,修正を施すならば精密さを害することはない(第七章參照). 各の階級を 30 とすれば,欄の數 900 となり,中には度數 0 なる欄も多いけれども,計算

[1] 一般に X 若くは Y が年齢である時は不規則又は不相稱の二重分布を呈することが多い.

第 48 表

英國人に於ける肺活量と年齢との相關表.

(肺活量 X は階級幅 15, 年齢 Y は階級幅 3 年)

年齢 \ 肺活量	42·5—57·5	57·5—72·5	72·5—	87·5—	102·5—	117·5—	132·5—	147·5—	162·5—	177·5—	192·5—	207·5—	222·5—	237·5—	252·5—	267·5—	282·5—	297·5—	312·5—	327·5—	342·5—	357·5—	372·5—	387·5—402·5	
78—80					1					1															2
75—77		1					1		1																3
72—74						1	1		1																3
69—71							1	3		1															5
66—68					1	1		1		1	1														5
63—65						4	2	1	5	4	1	2													19
60—62							1	1	1	2	4	1	1	1	1										13
57—59					1	2	1	4	2	3	6	5	2												26
54—56						2	1	2		3	1	7	2		2	1									21
51—53					2	1	2	2	2	1	4	4	4	2		1									25
48—50							3	4	4	6	7	6	7	2	3		2								44
45—47				1		2	2	4	7	10	5	2	9	11	2	3	1		1		1				61
42—44						1	1	2	2	14	4	12	7	8	2	3		2							58
39—41		1				2	1	3	1	12	4	12	15	12	4	5				1					73
36—38						3		5	2	12	8	17	13	7	6	5	1	3		2	1				85
33—35		1	1				1	2	2	11	11	16	16	22	12	12	2	2		1					112
30—32				2		1	5	2	2	12	10	29	13	26	10	8	5	5	2						132
27—29						1	2	2	4	18	19	36	30	42	14	22	9	6	3	4	1				213
24—26						2	1	8	9	33	19	53	41	51	27	28	10	8	2	1					293
21—23					1	2	2	22	12	34	39	89	73	82	37	35	16	16	3	3	1				467
18—20				3	1	2	10	10	29	70	84	151	138	112	76	54	25	22	5	6				1	799
15—17				3	6	13	27	50	51	84	48	70	60	39	20	14	4	3	2	1					495
12—14	1	1	2	29	51	59	51	47	18	13	9	4	3	1		1									290
9—11		6	14	24	26	29	8	2	1	1			1												121
6—8	3	5	5	5	3																				21
	4	15	22	67	93	128	124	177	156	346	284	516	435	418	216	192	75	67	18	19	4			1	3377

の煩瑣なるを察することが出來る.

(ii) 相關の研究には,單一變數の研究の場合よりも,多數の觀測を必要とする. 觀測總數が少い時は誤つた相關の觀念を呼起し易いから,かくの如き時は必ず相關係數と共に,その標準誤差若くは確率誤差を計算し,愼重の態度を取らねばならぬ.

演習問題

1. 本章,例 2, 鷄卵重量とその孵化時に於ける雛體重との相關表(第 41 表)を Scattered diagram に描寫せよ.

2. 本章,例 5, 燕麥穀粒の重量とその含有脂肪百分比との相關を Stereogram に描寫せよ.

3. 本章,例 8, England と Wales とに於ける生出全數 Y と男子出生率 X との相關を Scattered diagram に描寫せよ.

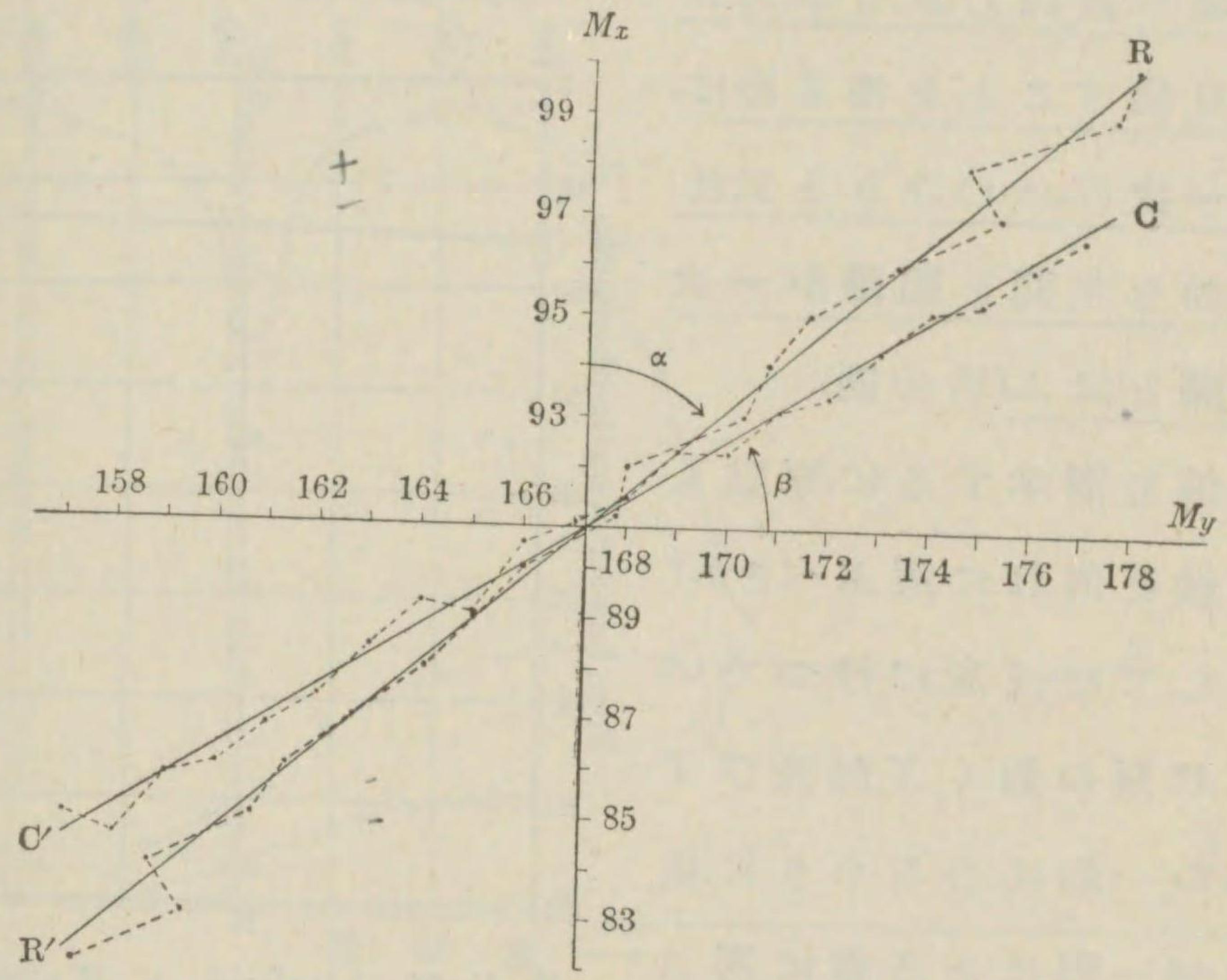
第六章
相關の諸計算

1. 回歸線

相關表を作れば,相關の大體を概觀することが出來るが,表が厖大で捕捉に困難である. 單一變數の場合平均や散布度で簡單に度數分布を代表せしめた如く,相關の場合でもその相關の程度を簡單な値で代表せしめたい. この爲に相關係數,回歸係數等を計算するのであるが,今これらの係數の意義を順を追うて說明することとする.

相關表を見て第一に注意すべきことは,相關表における各行及び各列の算術的平均である. 例へば,例1(第40表)に於て第一變數 X の各階級內で,第二變數 Y の算術的平均を(卽ち各列の算術的平均を)計算し,第40圖の如く平面上に點を打てば,算術的平均は略一直線 CC′ 上にならぶ. 又,第二變數 Y の各階級內で,第一變數 X の算術的平均(卽ち各行の平均)を圖上に取れば,これも同じく一直線 RR′ 上,またはその附近に點在する. 吾々が最も普通に遭遇し,且つ規則正しき相關表では,かく各行及び各列の平均が直線上にならぶ場合であるが,實測の二重分布では,偶然の影響を受くるを以て,完全に直線上にあること殆どなく,點は直線の兩側にならび,これらの點を順次結ぶ線は鋸齒狀を呈する. 觀測數が少く,階級數が多い時ことに甚しい. しかし確率

論で豫期出來る範圍內の不規則さならば,これを直線と見なす



第40圖　山東苦力の身長(X)と棘高(Y)との相關表より計算せる回歸線 RR′ 及び CC′. (階級幅 1 cm. の表より圖示せり.)

ことが出來る. 第40圖,第43圖乃至第45圖は,これらの諸點が直線上にありと見做して差支ない場合である. 但し,分布の周邊の階級では度數が少く,偶然の影響が大きいから,度數の少い二三の階級は圖から省いた方が安全である.

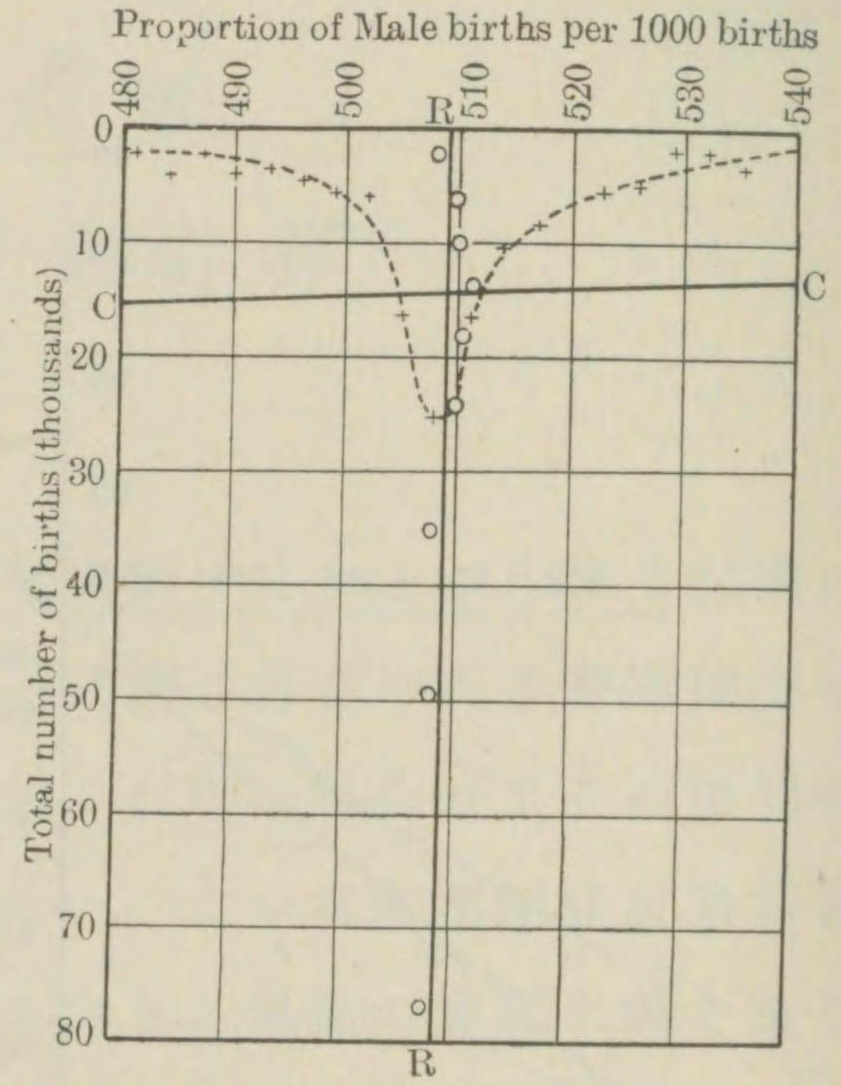
然るに,各行及び各列の平均が直線上にありと見なし得ないことが稀でない.

第41圖は各行の平均が略直線上にあるが,各列の平均は一の曲線上にある場合である. 又,第42圖は各列平均は直線上にありと見做せぬことも無いが,各行の平均は明に曲線上にある.

各列の平均若くは各行の平均を結ぶ直線又は曲線を**回歸線**

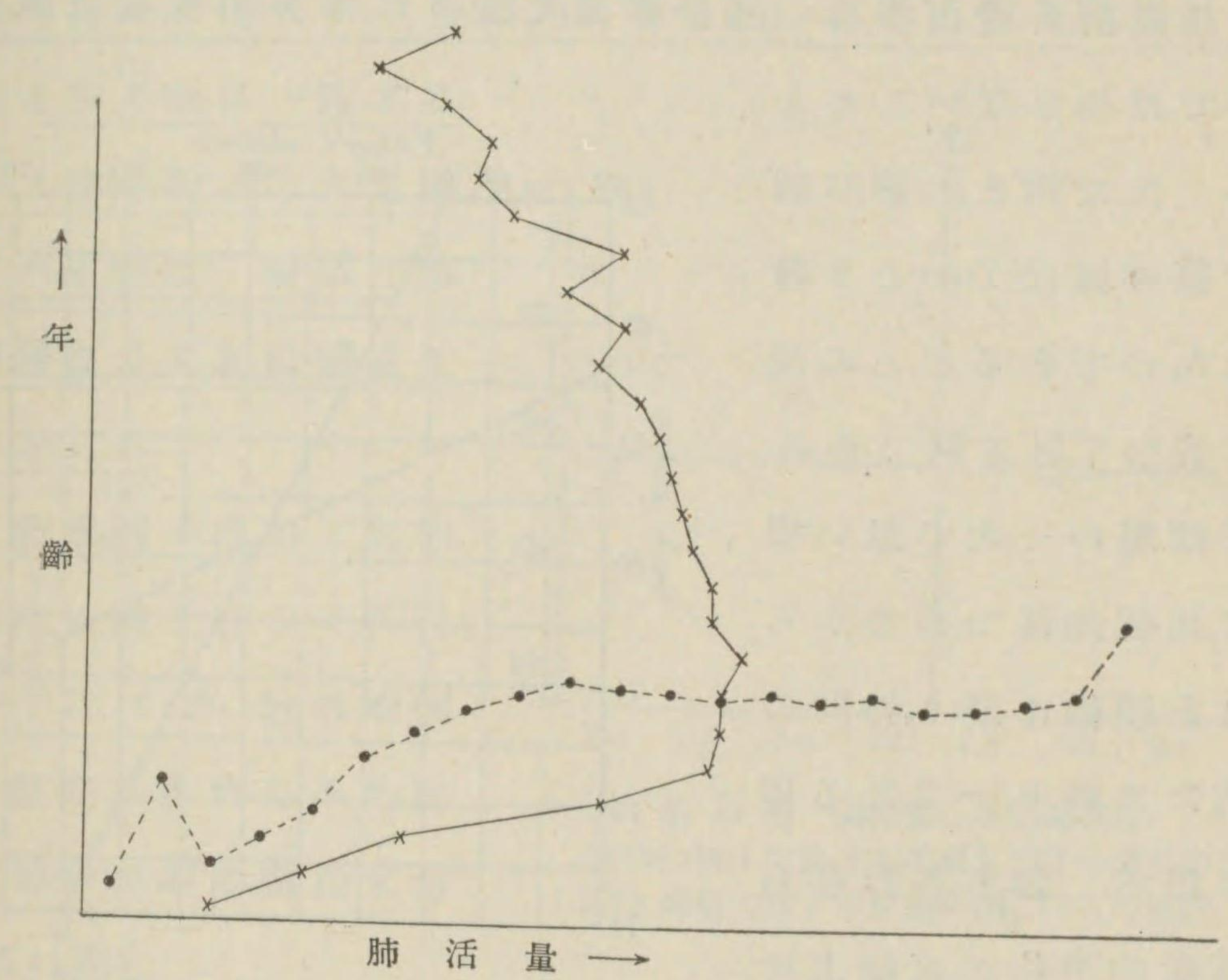
(Lines of regression; Regressionslinien) と云ふ.

回歸線が直線であるか,又は直線と見做すことを得る時は,**回歸が一次**(linear)なりと云ひ.かくの如き相關を**回歸が一次なる相關**と云ふ(第40圖).



第41圖 England と Wales に於ける延べ數632地方に就いて出生總數と男子出生率との相關. (Yule)
各行の平均は一直線の兩側にあるが各列の平均は二曲線上にあるを示す. 又 Y の正の方向は他の圖形と反對に取つてあること他の圖形と相違してゐる(第47表).
$r=-0{\cdot}014.$

回歸線を圖示するに,原點及び座標軸を何れに選ぶべきか? 原則としては任意に行つていい. 第42圖の如く X 軸及び Y 軸を圖の一端にあるやうに,從て原點が一隅にある樣に,選ぶのも一つの方法である. しかし,第40圖等の如く,變數 X 及び Y の算術的平均 M_y 及び M_x を座標軸とし,その交點を原點に選ぶのが便利である. かくすると分布の全區域が座標軸で4個の象限に分たれ,相關の研究に都合がいい. 特別の記述なき限り,今後はこの規約に從ふこととする. 從て,各觀測對の偏差 x, y は各點の座標 x, y で表はされる. 即ち $x=X-M$ であるから變數 X に代ふるに x を以てすることとなる. 又 x, y の正負は大體平面解析幾何學の規約に從ふ. 即ち, x の正の方向は x を X 軸(即ち M_y 軸)の右の方向に取りたる場合,負の方向は左方に取りたる場合と定め, y の正の方向は y 軸(即ち M_x 軸)を原點より

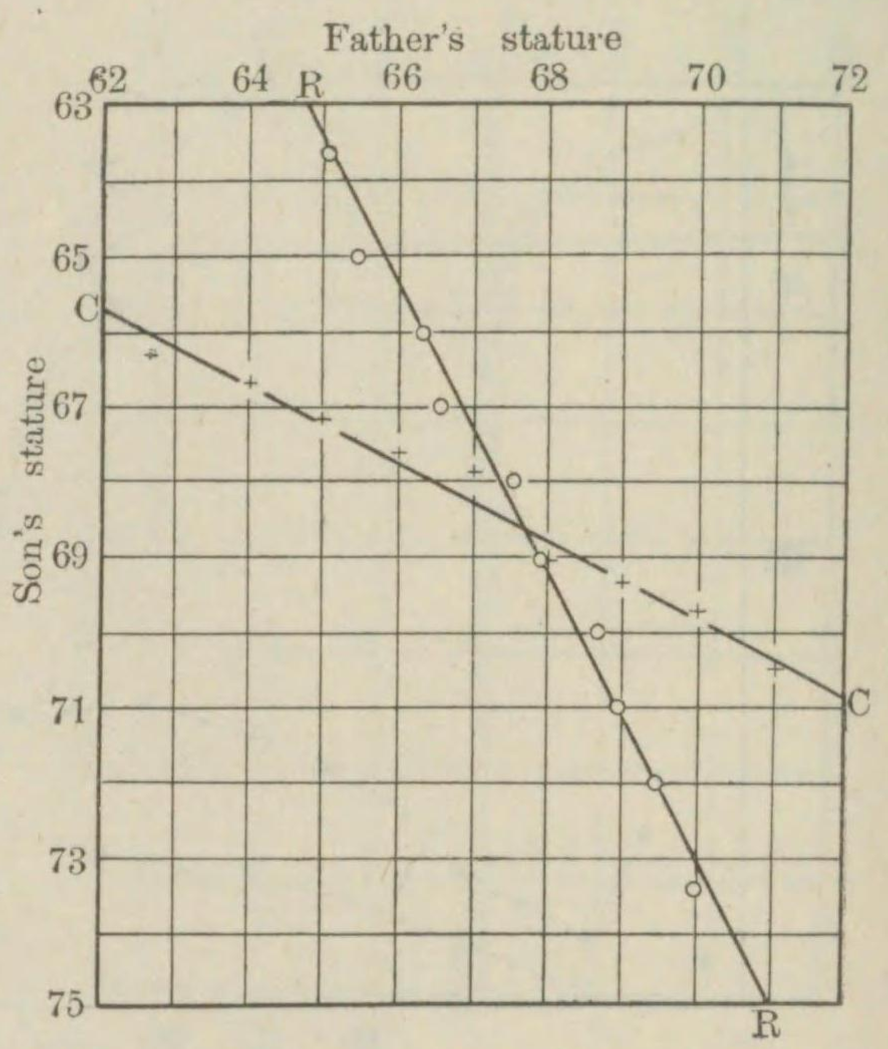


第42圖 英國人肺活量と年齡との相關(♂).
各列の平均は殆ど一直線上にあれども各行の平均は一曲線上にあることを示す(第48表).

上方に,負の方向は下方に取りたる場合と定める. 又,角度は X 軸を原線とする場合は,時計の針の回轉と同一方向に取りたる時を負とし,反對の方向に取りたる時を正と定める. Y 軸を原線としこれより角度を計る時は,時計の針と同方向に回轉したる時は正,然らざる時は負と定める. 第40圖の α, β, は共に正,第46圖では共に負である.

多數の著書中には別の規約に從つて圖示する人もある. 例へば x の正負の方向は上の通りであるが, y は原點より下方に取りたる時を正とし,上方に取りたる場合を負と定める人もある. さうすれば角の正負の方向も變へなければならなくなり,

且つ平面解析幾何學との連絡が悪く,餘りに推奨出來ないから,本書では用ひないこととした. ただ例とし第41圖及び第43圖は Yule より轉載したのであるが,この規約に從つて居る例である. さて回歸の一次で無い場合は,比較的稀であるからこれを續編に讓り,特別の記述なき限り一次なる場合に限る. 即ち,各行各列の算術的平均が直線上にあるか,又は直線上に在りと見做して差支ない場合に限ることとする.



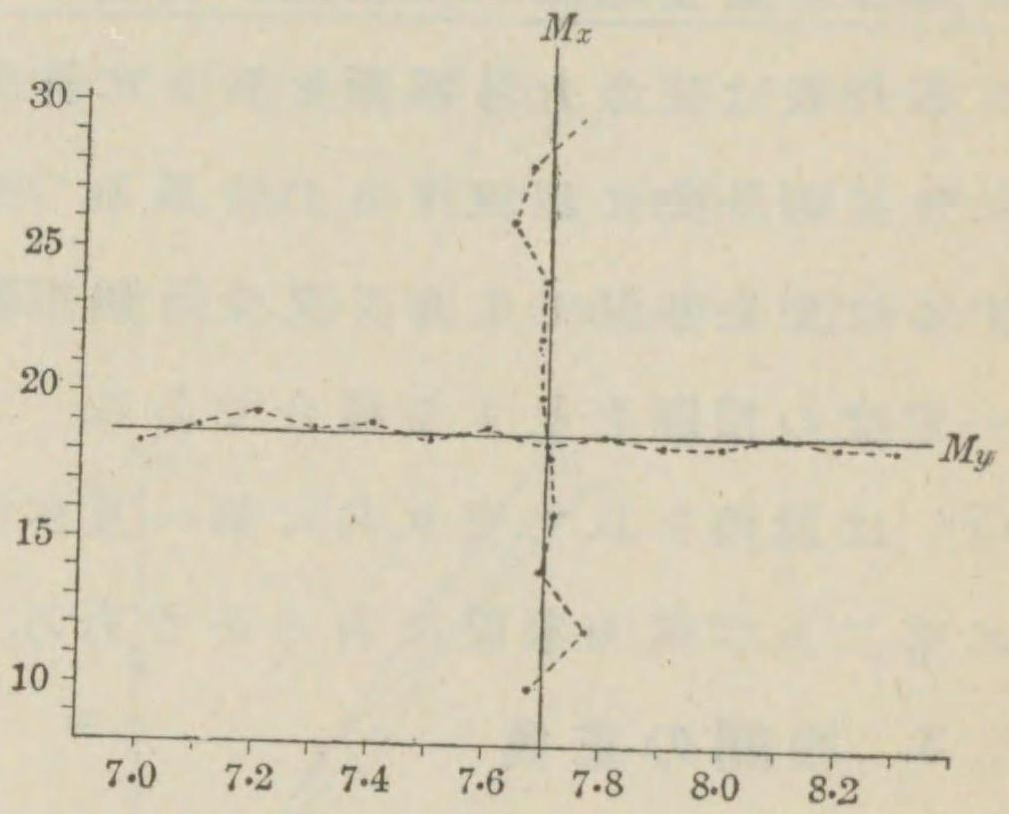
第 43 圖 父と子との身長の相關 (Yule) ($r = +0{\cdot}51$).
Yの正の方向の取り方の異なれる一例. (第 42 表)

2. 無相關と完全相關

回歸が一次で有る時,2 個の回歸線 RR′, CC′ が直角に交ることも銳角で交ることも,又角が更に小となり遂に二直線が全然合して一となることもある. 先づ,互に直角に交る場合を攻究して見る. 各列の平均を結ぶ直線 CC′ は X 軸に一致し,各行の平均を結ぶ線 RR′ は Y 軸と一致する. (X 軸及び Y 軸を M_y, M_x 以外に選ぶ時は RR′, CC′ は座標軸に並行する.) 換言すれば,各列の算術的平均は悉く全觀測に於けるYの平均M_yに等しく(又は近似的に等しく),各行の算術的平均は悉く全觀測に於けるX

の平均M_xに等しい(又近似的に等しい). かくの如く,回歸線がX軸及びY軸に一致するか,又は一致すると見做しうる時は,二變數は**無相關なり,又互に獨立なり**と云ふ.

前章,例 4 は殆ど無相關なる例であつて,各行の平均は M_y 軸の兩側に散在し,各列の平均は M_x軸の兩側に點在する (第44圖).



第 44 圖 無相關なる回歸直線.
前章の例 4 に於ける英國人男子の頭長徑 (X) と視覺に對する反應時間 (Y) との相關.

第 49 表

作爲せる完全相關の一例 (x, y 共に正常分布である場合).

y \ x	−8	−7	−6	−5	−4	−3	−2	−1	0	+1	+2	+3	+4	+5	+6	+7	+8	計
+8																	1	1
+7																3		3
+6															9			9
+5														22				22
+4													45					45
+3												78						78
+2											116							116
+1										147								147
0									158									158
−1								147										147
−2							116											116
−3						78												78
−4					45													45
−5				22														22
−6			9															9
−7		3																3
−8	1																	1
計	1	3	9	22	45	78	116	147	158	147	116	78	45	22	9	3	1	1000

次に又,2個の回歸線が互に一致することがある. かくの如き場合を**完全相關**又は**相關が完全**なりと云ふ.

第49表は完全なる相關を假りに作爲したので有つて,かくの如きは例外的に出現するのである. 生物學的統計で常に遭遇するは,完全相關でもなく,又全く無相關でもなく,完全ならざる一定度の相關を呈する場合である. この時は回歸線 RR′ 及び CC′ は鋭角を以て交り,共に第一及び第三象限にあるか,又は共に第二及び第四象限に有るかである.

3. 相關の正負

無相關の場合を除いて,相關が有りさへすれば,その完全なると不完全なるとを問はず,正負二つの場合に區別しうる. 即ち,

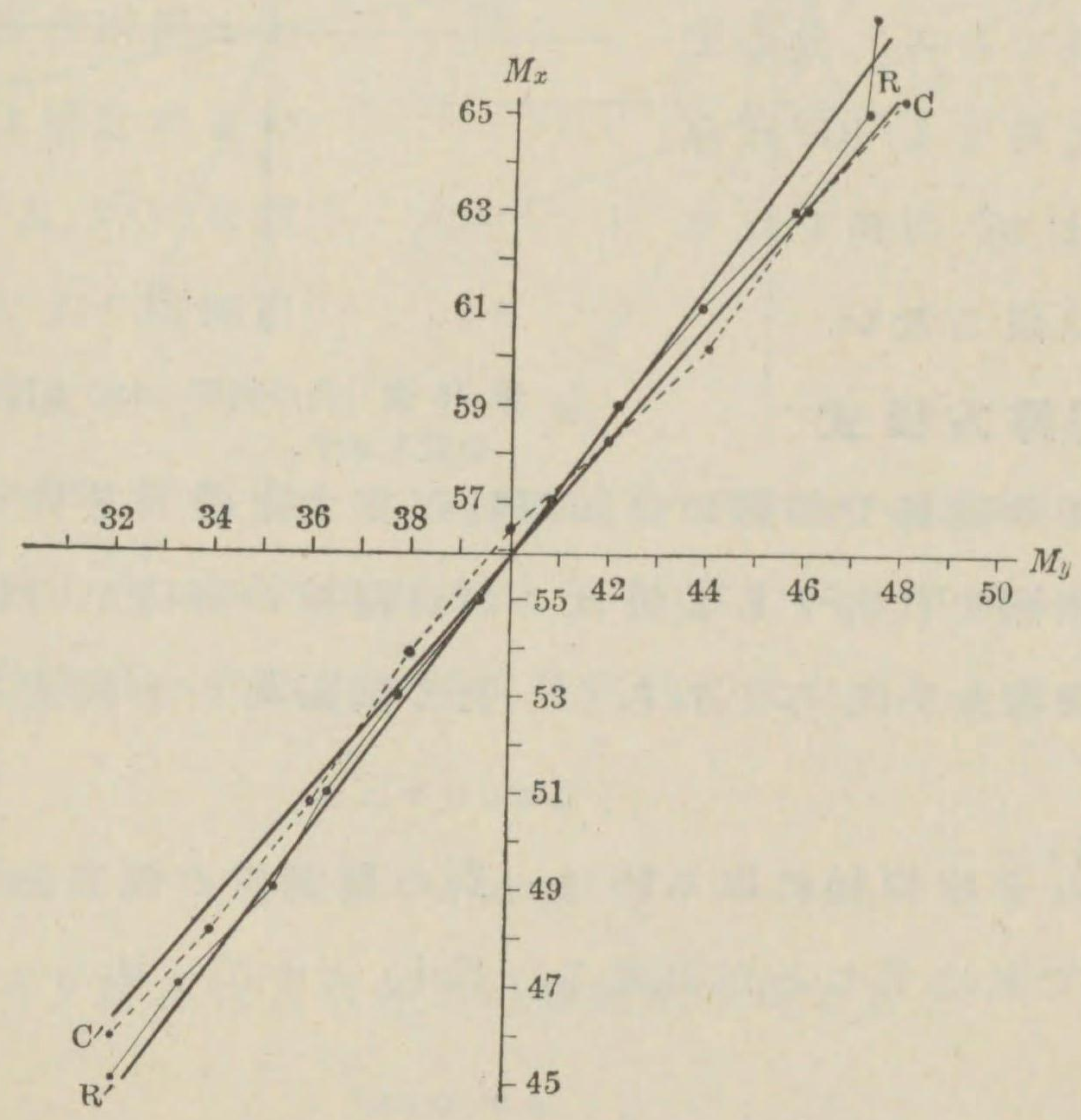
(i) Xが増大するに從つて,Yの平均も増大し; 同樣にYの増大するに從つて,Xの平均も亦増大する. 前章,例1—例3等がそれである. この時は回歸線 RR′ 及び CC′ は共に原點を通り,第一及び第三象限內にある(第40圖及び第45圖).

この場合を**相關が正なり**又**正相關**と云ふ.

(ii) 第二にはXの増大するに從ひて,Yの算術的平均は反對に減少し; 又Yの増大するに從ひて,Xの算術的平均も亦減少する場合であつて; RR′ 及び CC′ は共に原點を通り,第二及び第四象限內にあり. この場合を,**相關が負なり**又**二變數は負の相關關係にあり**と云ふ. 前章,例5がそれである(第44表及び第46圖).

回歸線 RR′ 及び CC′ が初めX軸とY軸とに一致する場合(無相關)より,同時に第一及び第三象限に動き出せば,正の相關とな

るのであつて,しかも初めは相關が低いが,二直線が漸次接近し,RR′ と CC′ との挾む角漸次小となれば,正の相關が漸次高度となり,遂に二直線が合する時は正の完全相關となる. RR′ が M_x 軸との挾む角を α とし,CC′ が M_y 軸との間に挾む角を β とすれば,RR′, CC′ が第一及び第三象限に在る限り,α も β も正である

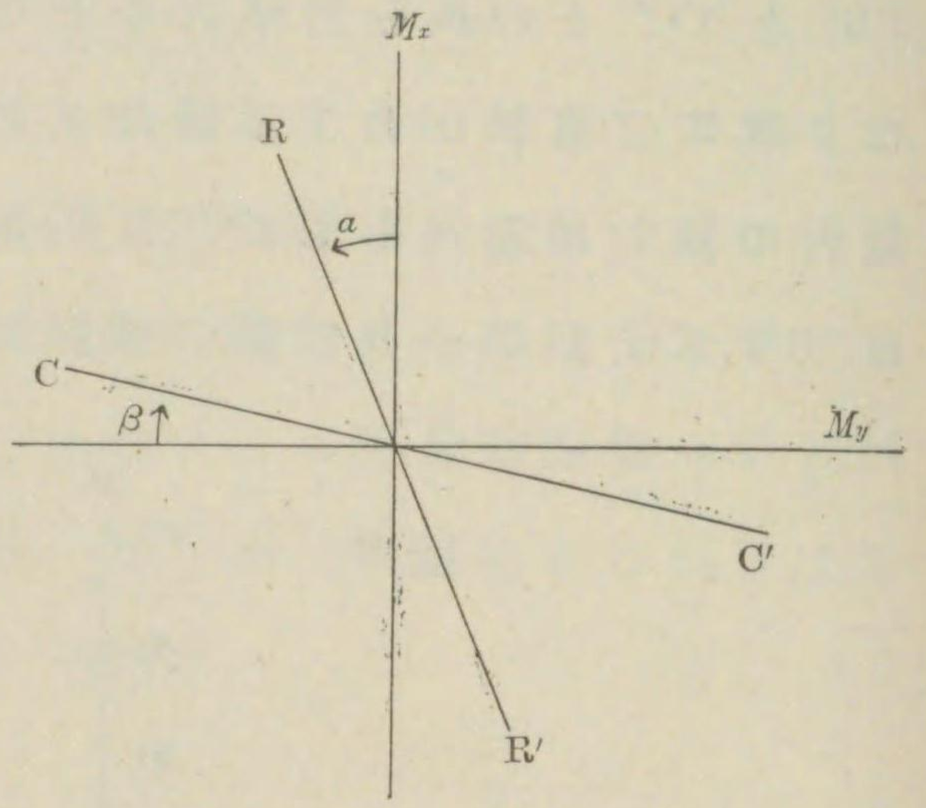


第45圖 鷄卵重量と孵化時における雛體重との相關. RR′ と CC′ とを示す(表41表).

(第40圖). しかし α と β とは必ずしも等しくない. 又 RR′ と CC′ とが合一せる場合(正の完全相關)に於ても,この直線はX軸と45°の角で交るとは限らない.

次に RR′, CC′ が座標軸に一致せる位置より,第二及び第四象限に向つて同時に回轉し始むれば,負の相關となるのであつて;

直線漸次接近すれば,負の相關漸次高度となり(第46圖),遂に二直線合すれば,負の完全相關となるのである.



第46圖 負の相關における回歸線の位置を示す.

負の相關でも角αとβとは必ずしも等しくないが,共に負である. 負の完全相關に於ても, RR′は座標軸とは 45° の角を以て交るとは限らない.

4. 回歸方程式

今までの記述で,相關には回歸線が重大な役目を持つものであつて,相關を代表する數値はこの回歸線の研究から得られる事が略理解せらるるであらう. 先づ回歸線の方程式を求めて見よう.

M_y, M_x を座標軸に取る時は,一對の觀測値の偏差 (x, y) は座標(x, y)で表はすことが出來る. 但し $x = X - M_x$, $y = Y - M_y$ である.

變數 x の階級中央の値を, $x_1, x_2, x_3, \cdots, x_k$

變數 y の階級中央の値を, $y_1, y_2, y_3, \cdots, y_h$

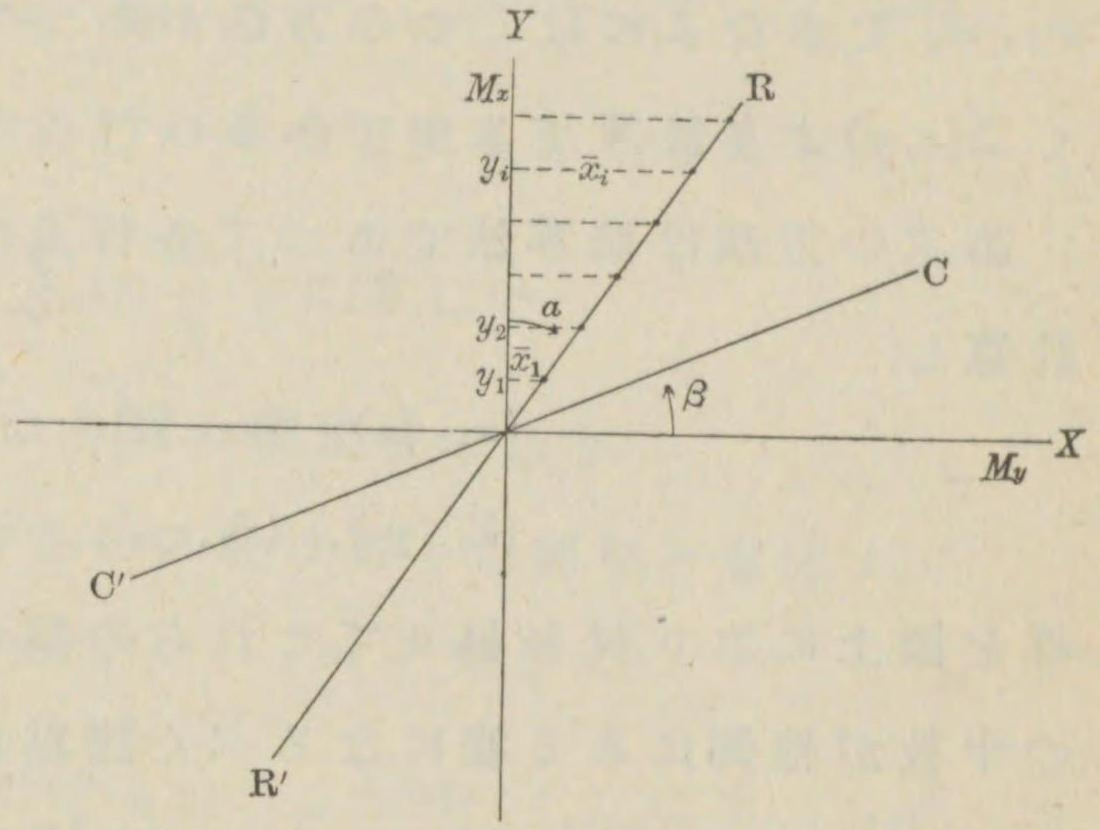
とす. 今 y の各階級內で, x の算術的平均(各行の平均)を,

$$\bar{x}_1,\ \bar{x}_2,\ \cdots,\ \bar{x}_k$$

とす.

若し回歸が正しく一次なれば, $(\bar{x}_1, y_1)$, $(\bar{x}_2, y_2)$, $(\bar{x}_3, y_3)$, $\cdots$, $(\bar{x}_h, y_h)$ は何れも一直線 RR′ 上にある(第47圖).

又, x の各階級內での y の算術的平均を, $\bar{y}_1, \bar{y}_2, \cdots, \bar{y}_k$ とすれば, $(x_1, \bar{y}_1)$, $(x_2, \bar{y}_2)$, $\cdots$, (x_h, y_h) も亦一直線 CC′ 上に無ければならぬ.



第47圖 正の相關における回歸線の位置.

若し,これらの點が完全に直線上になくして,直線の兩側にならぶ時は適當にこの直線RR′及びCC′を畫くことにより回歸直線を得られる[1].

そこでRR′がM_x軸(Y軸)と挾む角α, CC′がM_y軸と(X軸)挾む角βを知ることが出來れば,回歸線の方程式は解析幾何的に次の通りとなる.

$$\left.\begin{array}{ll} \text{RR}'\text{は} & \bar{x} = y\tan\alpha \cdots\cdots\cdots\cdots\cdots \\ \text{又 CC}'\text{は} & \bar{y} = x\tan\beta \cdots\cdots\cdots\cdots\cdots \end{array}\right\}\text{(VI. 2)}$$

但し, $\bar{x}$, $\bar{y}$ は各行及び各列の算術的平均である.

今,
$$\left.\begin{array}{l} \tan\alpha = m_1 \\ \tan\beta = m_2 \end{array}\right\}\cdots\cdots\cdots\cdots\cdots\text{(VI. 2)}$$

と置けば,

$$\left.\begin{array}{lll} \text{RR}'\text{は} & \bar{x} = m_1 y \cdots\cdots\cdots & \text{(a)} \\ \text{CC}'\text{は} & \bar{y} = m_2 x \cdots\cdots\cdots & \text{(b)} \end{array}\right\}\cdots\cdots\text{(VI. 1)}$$

となる. これを回歸方程式と云ふ. その中で(a)式は y を知り

[1] 適當に直線を畫くには最小自乗法を用ふる(5條參照).

て其の行に於ける x の平均を計算する式であり,(b)は x の値を知りて y の平均を計算する式である. 而して, α 及び β 或は m_1, m_2 を求むるには二つの方法があつて,一は次條で述べる如く $\Sigma(x,y)$ より計算するのである.

第二の方法は概算法であつて,各行及び各列の算術的平均を計算し,

$$(x_1, \bar{y}_1),\ (x_2, \bar{y}_2),\ \cdots\cdot$$
$$(\bar{x}_1, y_1),\ (\bar{x}_2, y_2),\ \cdots\cdot$$

等を圖上に取り,絲を張りてこれらの點の半數が絲の一側に他の半數が他側にある樣に,なるべく諸點に近く且つ原點を通る直線 RR′ 及び CC′ の位置を定めて,これを描くのである. 次で, RR′ の y 軸と挾む角 α,及び CC′ が X 軸と挾む角 β を分度器で計る. α, β が明になれば m_1, m_2 はこれより三角函數表を用ひて求めることが出來る. 或は Y 軸の任意の一點 y_i よりこれに垂線を引き, RR′ と交らしめ,その點を $\bar{x}_i$ とし, $\bar{x}_i$, y_i の長さを計測して,

$$m_1 = \frac{\bar{x}_i}{y_i},$$

より m_1 を計算する. m_2 も同樣に CC′ から計測,計算することが出來る.

式(VI. 1)を**回歸方程式**と云ひ, m_1 及び m_2 を**回歸係數**と云ふ.

5. 回歸係數

最初に各行及び各列の平均が完全に直線 RR′ 及び CC′ 上にある場合につき,回歸係數の計算式を誘導し,次に直線と見做しうる場合に擴張せん.

$(\bar{x}_1, y_1), (\bar{x}_2, y_2), \cdots\cdot$ は RR′ 上にあるから(第47圖),

$$m_1 = \frac{\bar{x}_1}{y_1} = \frac{\bar{x}_2}{y_2} = \cdots\cdot = \frac{\bar{x}_h}{y_h}, \quad \cdots\cdots\cdots\cdots\cdots (A)$$

而して, $\bar{x}_1, \bar{x}_2, \cdots\cdot$ は各行の平均であるから,それぞれ

$$\frac{1}{n_1}\sum x_1,\ \ \frac{1}{n_2}\sum x_2,\ \cdots\cdot \text{ に等しい.}$$

但し, $n_1, n_2, \cdots\cdot, n_h$ は各行の觀測數であり,

$\Sigma x_1, \Sigma x_2, \cdots\cdot$ は各行に於ける x の總和を意味す.

今,(A)式の分母子に, $y_1, y_2, \cdots\cdot, y_h$ を乘じて,

$$m_1 = \frac{\Sigma x_1 y_1}{n_1 y_1^2} = \frac{\Sigma x_2 y_2}{n_2 y_2^2} = \cdots\cdot = \frac{\Sigma x_h y_h}{n_h y_h^2}, \quad \cdots\cdots\cdots\cdots (B)$$

とすることが出來る. この (B) は比例式であるから,分母の總和と分子の總和を取るも m_1 に等しい[1]. 即ち,

$$m_1 = \frac{\Sigma x_1 y_1 + \Sigma x_2 y_2 \cdots\cdot + \Sigma x_h y_h}{n_1 y_1^2 + n_2 y_2^2 + \cdots\cdot n_h y_h^2} \quad \cdots\cdots\cdots\cdots (C)$$

即ち,

$$m_1 = \frac{\Sigma(xy)}{N\sigma_y^2}$$

$\Sigma(xy)$ は總ての觀測對について(xy)なる積の總和, N は觀測總數である.

今, $p = \frac{1}{N}\Sigma(xy)$ とすれば,

[1] (A)式が成立する時は, m_1 を計算する爲に(C)式を誘導せずともいいわけであるが, $(\bar{x}_1, y_1), (\bar{x}_2, y_2), \cdots\cdot$ 等が完全に一直線上に無いが一直線の兩側に分布する場合へ擴張する準備として,迂遠ながら(C)式を誘導したのである.

同様に,
$$\left.\begin{array}{l} m_1 = \dfrac{p}{\sigma_y^2}, \quad \cdots\cdots\cdots\cdots\cdots \\ m_2 = \dfrac{p}{\sigma_x^2} \quad \cdots\cdots\cdots\cdots\cdots \end{array}\right\} \text{(VI. 3)}$$

なる式を誘導することが出來る.

以上は各行及び各列の平均が一直線上に在りと假定して,回歸係數 m_1, m_2 を計算したのであるが,若し,完全に直線上にないが,一直線と認めて差支なき時も,この計算式を用ふることが出來る. 次に,最小自乘法の原理を應用して式を誘導する.

今, RR′ の方程式を $\bar{x} = m_1 y$,

CC′ の方程式を $\bar{y} = m_2 x$

とす. 但し m_1, m_2 の値は未定である.

N 對の觀測値の各を上式に代入するに,上式は必ずしも滿足せられない. 算術的平均に代ふるに觀測値を以てするから,

$$x - m_1 y = \delta,$$
$$y - m_2 x = \delta'$$

に於ける δ, δ' は 0 とならず.

最小自乘法によれば,誤差 δ, δ' の平方の和を最小ならしむる様に m_1, m_2 を定むるのが最も正しいとせられて居る.

即ち, $(x_1 - m_1 y_1)^2 + (x_2 - m_1 y_2)^2 + \cdots\cdots + (x_N - m_1 y_N)^2 = \varDelta_1^2,$

$(y_1 - m_2 x_1)^2 + (y_2 - m_2 x_2)^2 + \cdots\cdots + (y_N - m_2 x_N)^2 = \varDelta_2^2$

とし, $\varDelta_1^2$, $\varDelta_2^2$ を最小ならしむる様に, m_1 及び m_2 を定むれば,それが最も正しい回歸係數である.

上式より

$$\varDelta_1^2 = x_1^2 + x_2^2 + \cdots\cdots + x^2_N$$

$$+ 2m_1(x_1 y_1 + x_2 y_2 + \cdots\cdots + x_N y_N)$$
$$+ m_1^2(y_1^2 + y_2^2 + \cdots\cdots + y^2_N)$$

即ち, $\varDelta_1^2 = N\sigma_x^2 - 2m_1 Np + Nm_1^2\sigma_y^2,$

但し $Np = \Sigma(xy)$ とす.

又, $\dfrac{\varDelta_1^2}{N\sigma_y^2} = m_1^2 - 2m_1 \dfrac{p}{\sigma_y^2} + \dfrac{\sigma_x^2}{\sigma_y^2};$

即ち,右邊を最小ならしむる m_1 の値は $\varDelta_1^2$ を最小ならしむ.

しかるに, $\dfrac{\varDelta_1^2}{N\sigma_y^2} = \left(m_1 - \dfrac{p}{\sigma_y^2}\right)^2 + \left(\dfrac{\sigma_x^2}{\sigma_y^2} - \dfrac{p^2}{\sigma_y^4}\right), \cdots\cdots\cdots \text{(D)}$

(D) 式の右邊の第二項は常數であるから,第一項を最小ならしむる値,即ち $\left(m_1 - \dfrac{p}{\sigma_y^2}\right)^2$ を最小ならしむる m_1 の値が $\varDelta_1^2$ を最小ならしむ.

(D)式の第一項は平方であるから,最小値は

$$\left(m_1 - \frac{p}{\sigma_y^2}\right)^2 = 0,$$

即ち, $m_1 = \dfrac{p}{\sigma_y^2}$

が $\varDelta_1^2$ を最小ならしむる値である.

即ち, m_1 を $m_1 = \dfrac{p}{\sigma_y^2}$

に選びたる時の方程式

$$\bar{x} = m_1 y$$

が最も確らしき回歸直線である. m_2 につきても同様である. 即ち

定義. 各行及び各列の平均が一直線上にあるか,又はその兩側にありて事實上一直線上にありと見做しうる時は,回歸直線 RR′ 及び CC′ の方程式は,

$$\left.\begin{array}{ll}\mathrm{RR}': & \bar{x} = m_1 y, \\ \mathrm{CC}': & \bar{y} = m_2 x\end{array}\right\} \cdots\cdots\cdots\cdots (\mathrm{VI.\ 1.\ a})$$

である. 但し $m_1 = \dfrac{p}{\sigma_x{}^2}$ $m_2 = \dfrac{p}{\sigma_x{}^2}$ とす. これ**回歸方程式**(Equations of regression; Regressionsgleichungen) である.

x, y に代ふるに,變數 X, Y を以て回歸方程式をあらはすことがある.

$$\left.\begin{array}{ll}\mathrm{RR}': & X - M_x = m_1(Y - M_y), \\ \mathrm{CC}': & Y - M_y = m_2(X - M_x),\end{array}\right\} \cdots\cdots\cdots (\mathrm{VI.\ 1.\ b})$$

回歸とは Galton が身長の遺傳の研究に際し初めて使用せる名で,子の身長は必ずしも親の身長に等しからず常に多少親全體の平均に近づく,卽ち退行する又は回歸すると稱したのである. 今日ではこの說は正しいとは考へられて居ないから,回歸線,回歸方程式等と稱するのは不都合であるとの理由で, Yule は**特性線** (Characteristic lines), **特性方程式** (Characteristic equations)等と名づけるが宜いと云つて居る.

尙ほ, X と Y とが無相關である時は,

$$m_1 = m_2 = 0,$$

又完全相關の時は

$$m_1 m_2 = +1$$

である.

6. 相關係數

回歸係數, m_1, m_2 は 2 個で相關の程度をあらはすが,今

$$r = \pm\sqrt{m_1 m_2}$$

と置く時,この r を**相關係數** (Coefficient of correlation; Der Korrelationskoeffizient) と云ふ. 言葉で云へば,

定義. 回歸係數 m_1 及び m_2 の幾何的平均を相關係數と云ふ.

卽ち,
$$r = \pm\sqrt{m_1 m_2} \cdots\cdots\cdots\cdots (\mathrm{VI.\ 5})$$

或は又,
$$r = \frac{p}{\sigma_x \sigma_y}, \cdots\cdots\cdots\cdots (\mathrm{VI.\ 6})$$

又, $p = m_1\sigma_y{}^2 = m_2\sigma_x{}^2$ であるから,

$$r = \frac{m_1\sigma_y}{\sigma_x} = \frac{m_2\sigma_x}{\sigma_y}, \quad \cdots\cdots\cdots (\mathrm{VI.\ 7})$$

$$\left.\begin{array}{l} m_1 = r\dfrac{\sigma_x}{\sigma_y}, \\ m_2 = r\dfrac{\sigma_y}{\sigma_x}, \end{array}\right\} \cdots\cdots\cdots\cdots (\mathrm{VI.\ 4})$$

r, m_1, 及び m_2 の附號は常に同一であつて p の附號を取らしむ.

(VI.4)式により, m_1, m_2 は常に r 及び σ_x, σ_y より計算することが出來るから,普通には r を計算し m_1, m_2 は計算しないでいい. 相關係數は 1846 年 Bravais により創められ, Kapteyn により新しく數學的に意義づけられ, Pearson, Yule 等により採用せられたものである.

7. 相關係數及び回歸係數の性情

(i) 相關係數 r は相關に最も必要な數値であつて,單なる無名數である. 又, r を計算する式の分母と分子とは x 及び y につきて同次であるから,その大さは x, y の量られたる單位の大さには無關係である. 又 m_1, m_2 は無名數ではあるが, x につきて又 y につきて分母子は同次でないから, m_1, m_2 は X, Y を量つた單位の大さによつて變化する. 例へば, X, Y 共に mm. で計測した場合と, X は mm. で他は cm. で計つた時とは, m_1, m_2 に大