Title	部分と全体 : 試料に於ける誤差の本質
Sub Title	
Author	寺尾, 琢磨
Publisher	慶應義塾理財学会
Publication year	1939
Jtitle	三田学会雑誌 (Keio journal of economics). Vol.33, No.2 (1939. 2) ,p.173(27)- 209(63)
JaLC DOI	10.14991/001.19390201-0027
Abstract	
Notes	
Genre	Journal Article
URL	https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=AN00234610-19390201-0027

した最大の政治的勢力は軍部である。しかし、いまや支那事變の全面的發展によつて、軍部の主張は、全國民的支持と主張にまで發展した。この事變の解決の形態が如何であれ、わが國の大陸發展の政策は、現在既に確立され、國民的支持と主張を持つものといつてよいであらう。それは、最早單なる軍中堅部の政治的主張ではない。全國民的主張であり、唯一無二の國策であるといへよう。さういふ意味において、いまや日本は全體主義的構成の下にあるといつてよいであらう。

部分と全體

——試料に於ける誤差の本質——

寺尾琢磨

一　間接調査と部分調査

論理學又は數學の如く純粹の思惟過程を對象とする形式科學に於ては吾人は毫も事實の背景を必要としない。要求されるものは單に命題の合理的發展のみである。然るに何等かの具體的現象を對象とする實質科學に於ては理論の前提もその歸結も常に事實と關聯せしめられねばならぬ。實質科學に於て觀察又は調査の缺く可からざる所以はこゝに在る。蓋し事實は先驗的には把捉されないからである。

併し事實を認識する方法は科學の性質により、またその箇々の場合によつて大なる相違がある。天文學に於ける觀測、物理學及び化學に於ける實驗又は分析、地理學及び考古學に於ける調査、人口學及び經濟學に於ける統計調査の如き、その代表的なるものといへやう。この最後に擧げた統計調査をば、他の認識方法と區別せしめるものは、それが箇々の事實の代りに事實の集團を問題とするといふことである。單位又は部分はそれが構成する全體と必ず

しも性質を同じふしない。一國人口の職業は極めて複雜多岐に亘り、從つてその數、種類乃至割合は全人口について調査する以外に正確に知るを得ない。即ち一國人口を構成する各單位（各人）につき職業なる項目を調査するのであつて、この調査項目を標識（merkmal）といふ。即ち統計調査とは統計集團を構成する各單位につき一定の標識を調査することであつて、單位と全體とが同性質に非ざる場合には、この方法に據る以外には全體を知る途の無い事は言ふ迄もないのである。そして社會科學の對象は言ふ迄もなく社會的事實なる集團的事實であるから、これが把捉に統計調査の要請されるのは當然のことである。

斯く統計調査は統計集團を構成する單位の全部につき所定の標識を直接に調査すべきもので、これを本來の意味に於ける大量觀察といふ。國勢調査はその代表的なものである。然るに斯かる統計調査は次に述べる二つの理由から、常に必ずしも行ひ得ないのである。

（一）第一には統計調査はその本質上大規模に亘らざるを得ないから、勞力及び費用の調達に困難を來すものである。國勢調査では一調査員の受持世帶數は五〇乃至一二〇であるから、全國では少くとも十萬人前後の調査員を必要とする。調査員をば名譽職と定めてゐる我國の如きは調査費用は比較的小額で濟む筈であるが、而もなほ數十萬の經費は避け難いのである。勞力と費用とに大きな制限が附されたのでは、統計調査は到底正確を期し難く、寧ろ斷念するのが得策であらう。

（二）第二の困難は技術的方面に求められる。國勢調査は一國の人口を或る時點で捉へ以てその大さと構成を知る事を目的とする。然るに人口は一瞬と雖も靜止狀態に在るものではなく、出生・死亡及び移住來住によつて不斷に增減しつゝあるものである。斯かる動態をもし直接の調査によつて確定しやうとすれば、全く一瞬の休みもなく絕えず調査を續けて居らねばならぬ。これは言ふは易く、行ふは困難といはんよりは不可能である。變動が極めて規則的に行はれる事が確實な場合には、二つの異る時期における靜態調査に補間法を施す事によつて、中間の各時期の狀態を正確に求める事が出來る。假りに人口が一定率で增加するものとすれば、この增加は複利の原則に從ふから、例へば一九三〇年と一九三五年の二回に國勢調査を行ひ、その人口をそれぞれA及びBとすれば

$$B=A(1+r)^5$$

となり、この式からr即ち增加率を求める事が出來るから、これを利用して一九三一年、三二年等の人口を算出する事が出來るのである。然るに變動率が全く一定なるが如き現象は殆どあり得ないから、補間法による計算は多かれ少かれ近似的な値を示すに止まる。故に人口について正確な報道を得んとすれば每年國勢調査を繰返へさねばならず、否每月又は每日繰返さねばならぬ理である。その不可能な事は改めて言ふ迄もないところであらう。

斯くて本來の統計調査即ち直接且つ悉皆的統計調査に代りうるものが必要となるのであつて、その一は間接調査であり、その二は部分調査である。前者は直接に統計の蒐集を行ふ代りに、既存の資料を單に編纂する事によつて統計を作製する事であり、後者は統計集團を構成する全單位の代りにその一部分を調査する事によつて所期の目的を果さんとするものである。言ふ迄もなくその何れに於ても、これによつて能く當該統計集團の本質が把捉さるべ

きためには一定の要件の具はる事を必要とするのであつて、不用意な濫用は統計の正確性を破壞する以外の何物でもない。

一、**間接調査**　計畫的持續的な事業遂行、例へば國家行政・企業・家計乃至は科學的研究の遂行に際しては、關係事項について自ら豐富な數字的資料が蒐集されるものである。人の出生・死亡又は婚姻はその都度屆出づ可き事が法律的に規定されてゐる。これは全く行政上の目的に出づるもので、決して統計の作製を目的とするものではない。併し吾人はこれを利用する事によつて出生統計・死亡統計又は婚姻統計を作製しうるのであつて、この意味に於ては、恰も最初から斯かる統計の作製を目的として統計調査を行ふのと毫も異らない。斯くの如く、元來非統計的目的によつて蒐集された資料をば統計に轉化したものを一般に二次統計(Secondary statistics)といふ。企業經營に於ては物品及び金錢の收受はその都度帳簿に記載されるが、これを編成する事によつて經營統計が作製される。同樣にして毎日の家庭生活の收支を記載した家計簿は家計調査なる統計調査の殆ど唯一の資料となつてゐる。即ちこれらは本來の目的は決して統計作製にあるのではないが、而も結果からいへば毫も異るものではないのである。

併し二次統計は次の點に於て一次統計と根本的な相違を持つてゐる。即ち二次統計は元々統計の作製を目的とする特別の調査から生れたものではないから、單位や標識の概念決定に於て必ずしも統一があるとは限らない。前例の出生や死亡の二次統計については、事柄が簡單なだけに殆どこの憂ひは無いが、例へば賃銀統計の如きに於ては、抑も何を以て賃銀と見做すかは會社や工場によつて必ずしも一樣ではない。貨幣賃銀を賃銀として記載するところもあらうし、或ひは更に實物給與をこれに附加して記載するところもあらう。故に賃銀支拂簿から賃銀統計を作製することは——當該工場又は會社のみを問題とする場合を除き——甚だ危險だと言はねばならぬ。賃銀其他の勞働條件については當然各工場又は會社に正確な帳簿がある可き筈であり乍ら、而も政府が勞働統計實地調査の如き大規模の直接調査を實施する主たる理由は、畢竟間接調査を以てしては調査の統一を期し難いことに在るのである。そして社會事象に關する概念は著しい程度に習慣や便宜によつて決定されるから、統計の正確を期する限りは、間接調査の行へる範圍は極めて狹少たらざるを得ない。そこで代用調査の他の一つたる部分調査が重要視されるのである。

二、**部分調査**　二次統計と並んで廣く行はれる代用調査は所謂部分調査である。もし全體の構成が均一であるとすれば、吾人はその一部分の構成を調査することによつて全體の構成を推しうることは勿論である。鑛泉や井水の構成を試驗管に汲取つた小量を分析することによつて求めるが如きこの適例であらう。併し斯かる場合には統計の概念は全く介入しない。蓋し統計調査の要請される所以は、全體と部分との間に異る性質の現はれるためであつて、もし部分が正確に全體の縮圖であるならば、何を苦んで尨大な大量觀察を行ふ必要があらう。換言すれば統計調査の要請されることとそれ自體が、部分の觀察によつて直ちに全體を推し得ないといふ事實を物語つてゐるのである。

乍併全體は部分の集團であるから、一定の前提の下に於ては統計調査に於ても部分から全體に遡ることも許される筈である。この前提とは第一には全體と部分の性質の間に特に大なる相違の認められない場合であり、第二には

部分が全體に對して比較的に大なる場合である。この第一の前提は多くの場合には寧ろ經驗を俟つて斷定せねばならぬから、事前にこの前提を置くことは概して不可能である。斯くて少くとも統計調査に關する限りは原則としては寧ろ第二の前提に據る外はない。百箇の球の入つてゐる箱から二箇又は三箇を取出したところ、それが何れも白球であつたからといつて全部が白球とは毫も考へられないが、五十箇を取出してそれが何れも白球であつたなら、殘りの五十箇も亦白球ではあるまいかといふ推測は可成り合理的となり、もし九十箇を取出してそれが何れも白球であるとすれば殘りの十箇も亦白球であらうといふ推測は一層合理的となるであらう。即ち吾人は理論的には部分を大ならしめる以外には一部から全體を推す方法は無いと言へるのである。故に部分調査とはいへ、その部分は一般に可成り廣汎に亘るを原則とするのであつて、この事は國勢調査速報に於ける抽出調査や米穀收穫高推定に於ける坪刈調査等の實状によつて容易に肯けやう。

併しこの場合にもなほ第一の前提が出來うる限り保持されることが望ましい。換言すれば抽出された部分をして出來うる限り全體の性質を具備せしめるといふことである。前述の如く吾人は豫め全體の性質を知り得ないから、積極的に部分を全體に等しからしめることは不可能であるが、而も消極的にこれに接近することは必ずしも不可能ではない。即ち部分を選擇するに當つて出來うる限り公平を期することこれである。百箇の球が白球と黑球から成るとし、もし黑球は白球より重いとすれば、特に適當に混和されぬ限り、最初には白球のみが取出される傾きがある。十歳兒童の體重を單に都會の小學校のみについて求めても、到底全國の狀態は推し得ないであらう。即ち單に部分を大ならしめる努力の外に、更に部分の選擇そのものに特殊の努力の必要とされる所以である。

部分から全體を推す場合、統計學の用語では部分をば標本又は試料(Sample)といひ、これに對して全體を全域又は母域(Population or Universe)といふ。母域はこれを有限母域と無限母域に分つ事が出來る。前者は一國の人口や生產額の如き一定の數を有するもので、これに對して後者は大氣中の各點に於ける氣壓の如き不定の數を持つものである。吾人の扱ふ數量は大部分は有限であるが、併しそれが極めて大なる場合には、理論的にはこれを無限母域と認めることが出來、これによつて却つて合理的に取扱へるのである。次に試料は人によつて幾多の見地から分類されてゐるが、要はその選擇が無差別に行はれるか又は意識的に行はれるかによつて分類してよからう。代表的部分が最初から明瞭な場合には、斯かる部分のみを選擇することが最も望ましいのは勿論である。例へば家計調査の如きこれである。家計の狀態は家族の構成や收入によつて著しく相違するから、正しい家計調査は最も正常的な世帶について行ふ可きであり、而も如何なる世帶が正常であるかは人口統計や所得統計から容易に割り出せる。故に全國各地に亘つて選出された數千の斯かる代表的世帶の家計狀態の平均を以て全國の家計狀態の反映と見做すことは最も妥當な方法といへやう。

併し家計調査に於けるが如く豫め代表的な單位乃至は部分を決定しうる場合は寧ろ例外に屬し、一般には斯かる代表的單位乃至部分は逆に母域から導かれるのである。男女の出生比とか人口の年齡別又は性別構成の如きは母域から斷定する外には、全くこれを知るの途なく、延いて代表的單位乃至部分を云々する途はないのである。

意識的選擇が不可能乃至無効の場合には、選擇に際して何等の先入觀をも介せしめない方法を採る外はない。千人から成る母域の學業成績を知らんとするに當つて、假りにその內の二百人を全く無差別的に、即ち手當り次第に抽出したとすれば、この二百人より成る試料は少くとも意識的に母域と性質を異にせしめられたものではない。そしてこの試料が母域に對して比較的に大なる限りに於ては、この試料を母域の縮圖と解しても大なる誤りはない筈である。旣に述べた通り、大部分の統計調査に於ては意識的に試料を選擇することが許されないから、如上の無差別的選擇こそ部分調査の原則的形態と認められるのである。この方法をば任意抽出法といひ、この方法によつて抽出された試料をば任意試料(Random Sample)といふ。こゝに謂ふ任意とは前述の如く、選擇に際して何等の先入觀をも介入せしめない事で、より正確に言へば、母域を構成する各單位が抽出されるに當つて其の抽出される可能性即ち確率が各單位に於て全く等しいといふことである。千人の學生のうち二百人が優等生なりとし、いま二百人を抽出するに當つて故意にこの優等生のみを選擇したとすれば、この二百人の優等生の各々の抽出される確率は、二百分の二百即ち一――これを確實といふ――なるに對し、殘りの八百人の各々の抽出される確率は八百分の零即ち零――これを不可能といふ――であつて、從つて千人なる母域を構成する各單位の抽出される確率は等しくない。この場合には試料は母域の縮圖ではなく、從つて試料から母域を推すことは出來ないのである。右の例に於ては、二百人の優等生を母域とし、これに對して悉皆調査を行つたことを意味するに留り、部分調査の概念は全く介入しないのである。

反之もし全く無差別に二百人が抽出されゝば、千人の各々が抽出される確率は千分の二百即ち〇・二であつて、從つてこの二百人は任意試料となる。何故に任意抽出法が優れた部分調査法と認められるかといへば、この方法が數學的確率論の原則によつて支持されるからである。部分は全體の一部に過ぎないから、試料と母域とは必ず多かれ少かれその性質を異にするが、而も任意資料に於てはこの相違は確率論的誤差法則によつて嚴格に規定され得るのである。殊に注意すべきは、旣に一言した通り、吾人は殆ど如何なる場合にも有限母域をも理論的には無限母域乃至はその一部と認めうるといふことである。換言すれば殆ど一切の統計調査は部分調査と見てよく、從つて統計資料は多かれ少かれ試料と見てよいのである。然らば統計資料に基く結論は、自ら誤差法則によつて律せられねばならぬと言はねばならぬ。私は曾て本誌上に「法則に於ける必然性と蓋然性」と題して統計學の確率論的基礎を論じたことがある(第三十二卷第三號)。この意味に於て、誤差法則を主題とする本稿は右の續稿と認めらる可きものである。

二　誤差法則

一、誤謬と誤差　誤差は英語の Error 又は獨逸語の Fehler の譯字で、動もすれば誤謬とか過誤とかの意味と混同される惧れがあるが、この兩者は嚴に區別されねばならぬ。誤謬又は過誤は人の意識的又は無意識的な不用意に基くもので、何等かの意味に於ける失策に外ならない。故に誤謬は吾人の注意によつて避けうる性質のもので、從つて數字上の誤謬は調査や計算を入念にすれば起り得ないのである。反之、科學上の用語としての誤差は斯かる

觀察上の心理的技術的原因に基くものではなく、實に事象發顯の樣態そのものに基くものを指すのである。事象の發顯は殆ど常に多かれ少かれ偶然的要素を含み、從つて觀察値と本質値とは必ずしも一致するものではない。日本人成年男子の身長の本質値が假りに五尺三寸なりとするに、觀察される各人の身長は或ひは五尺八寸とか或ひは五尺一寸とかの區々たる値を示してゐる。本質値と觀察値との斯かる差を誤差といふのであつて、この意味に於て誤差を特に偶然誤差と呼ぶのは極めて至當である。

誤謬と誤差との間には理論上次の如き重大な相違がある。即ち誤謬は上記の如く觀察者の心理又は觀察用具の不完全さに基くが、この不完全さは一種の性癖を有し、從つて觀察値は常に同じ方向に誤り易い。縮んだ物指で長さを計れば常に過大の方向に誤りを犯す事になり、從つて觀察を重ねれば重ねるほど、誤謬は大きくなる一方である。然るに誤差は事象發顯の偶然性に起因するから、その生起は偶然法則即ち確率論の原理に從ふべく、斯くて觀察を重ねれば重ねるほど正（プラス）の誤差は負（マイナス）の誤差によつて相殺される傾きを生じ、結局に於ては本質値の所在を示すに至るのである。

第一圖

甲	○	○	●	●
乙	○	●	○	●
	1	2		1

二、誤差法則　誤差生起の法則を最も簡單な例から研究しやう。甲乙二枚の鑄貨を投じて現はれる表裏の組合せは（○は表、●は裏とする）上の第一圖の示すとほりの四通りである。二枚を投げたのであるから、一方が表で一方が裏の場合が最も合理的な結果即ち本質形であることは言ふ迄もないが、右によれば四回のうち二回はこの本質形と一致し、他の二回は甲乙とも表か又は裏である。二枚とも表の場合、即ち本質形に乙が合致しない場合を負の誤差とし、二枚とも裏の場合、即ち本質形に甲が合致しない場合を正の誤差とすれば、本質形が二回現はれる間に正負の誤差が一回づゝ現はれる事になる。

同樣にして甲乙丙三枚の鑄貨を投げて見る。この時の本質形は一枚が表で二枚が裏か、又は二枚が表で一枚が裏の場合である。然るに全部が表又は裏の場合も起り得るのであつて、これらを例記すれば次の如くである。

第二圖

甲	○	○	●	●	●	○	○	●
乙	○	●	○	●	○	●	○	●
丙	○	●	●	○	○	○	●	●
		3			3			
	1	6						1

即ち本質形が六回現はれる間に正負の誤差が一回づゝ現はれる事になる。次に更に一枚を增して四枚を投げて見れば本質形は二枚が表で二枚が裏の場合であるが、この際全部が表なるとき、三枚が表で一枚が裏なるとき、三枚が裏で一枚が表なるとき、及び全部が裏なるときの四つの非本質形的な形即ち誤差が起りうる。その分布は次の如くである。

第三圖

甲	○	○	○	○	●	○	○	○	●	●	●	●	●	●	○	●
乙	○	○	○	●	○	○	●	●	●	○	○	●	●	○	●	●
丙	○	○	●	○	○	●	●	○	○	○	●	●	○	●	●	●
丁	○	●	○	○	○	●	○	●	○	●	○	○	●	●	●	●
	1	4				6						4				1

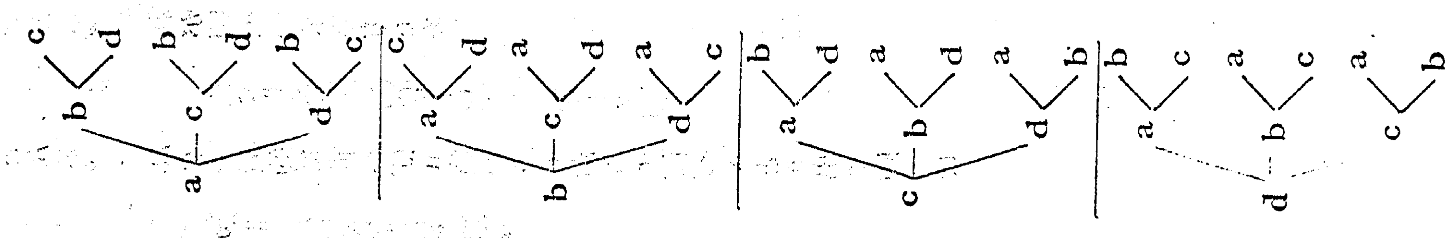
即ち本質形が六回現はれる間に、極端な正負の誤差（全部が表又は裏）が一回づゝ、中程度の正負の誤差（四枚のうち三枚が表又は裏）が四回づゝ現はれる事になるのである。

上記の例から直ちに推しうる事は、擲錢に於ける表裏出現の狀態は、二項式 $(a+b)^n$ の展開によつて求められる各項の係數に一致するといふ事である。二枚の場合に1、2、1と現はれるのは $(a+b)^2$ 即ち $a^2+2ab+b^2$ の三つの項の係數と一致し（a^2 の係數は1、ab のそれは2、b^2 のそれは1である）、三枚の場合に1、3、3、1と現はれるのは $(a+b)^3$ 即ち $a^3+3a^2b+3ab^2+b^3$ の四つの項の係數に外ならず、更に四枚の場合の1、4、6、4、1は $(a+b)^4$ 即ち $a^4+4a^3b+6a^2b^2+4ab^3+b^4$ の五つの項の係數と一致する。斯くて例へば十枚の鑄貨を投げたときの出現狀態は $(a+b)^{10}$ を、五十枚を投げたときのそれは $(a+b)^{50}$ を展開する事によつて容易に知りうるであらう。蓋しこの二項式の展開は「組合せ」(Combination) の應用による所謂二項定理(Binomiald Theorem)に據ればよいからである。二項定理とは

$$(a+b)^n=a^n+na^{(n-1)}b+\frac{n(n-1)}{1\cdot2}a^{(n-2)}b^2+\frac{n(n-1)(n-2)}{1\cdot2\cdot3}a^{(n-3)}b^3+\cdots\cdots$$

のことで、この各項の係數は組合せの原理に從つて決定される。上記の四枚の擲錢に於ける係數は1、4、6、4、1であつたが、これを一考して見やう。四枚を投げれば全部が表なる場合がありうるが、これは四箇から四箇とり出す組合せに外ならぬ。組合は順列(Permutation)に於ける排列の順序を無視したもので、次の如く理解されやう。

a、b、c の三箇から二箇づゝとり出す列べ方は ab, ba, ac, ca, bc, cb の六通りである。これを三箇から二箇づゝとり出す順列といひ ${}_3P_2$ と記し、$3\times2=6$ と計算する。何故然るかといふに、最初に a が來れば次には b か c かであるから、列べ方は二通りであり、b が最初に來れば次には a か c かであるから同じく二通り、最初に c が來たときも同樣にして二通りである。最初に來る文字は a、b、c 即ち與へられた文字全部であるから、この場合は3である。その各々が二通りの列べ方を持つから從つて全部では六通りとなるのである。これと同じ理由から四箇のうちから二箇づゝとり出す順列は ${}_4P_2=4\times3=12$ 通りである。いま四箇のうちから三箇づゝとり出す順列を見れば、次の如くである。



即ち最初に來る文字は a、b、c、d の何れでもよいから全部で四通りあり、その各々がその次に殘りの三文字の一つ

を持ち、またこの二番目の各々の文字がその次に殘りの二文字の一つを持つ事が出來るから、

$$_4P_3=4\times3\times2=24\ \text{通り}$$

となる。これを一般的に言へばn箇のうちからr箇づゝとり出す順列は

$$nPr=n(n-1)(n-2)\cdots(n-\overline{r+1})$$

である。これが順列の求め方である。

順列では排列の順序を考慮に入れ、例へばab. と ba とは別の物として取扱ふが、問題によつては内容さへ等しければ順序は問はぬ場合が尠くない。四枚の錢を投げて四枚とも表なるとき、もし順序を云々すれば $_4P_4=4\times3\times2\times1$ 即ち三十二通りあるが、順序を全く無視して單に全部が表なることだけを問題とすれば、言ふ迄もなく一通りしかない。上記のa、b、c、dの三文字のうち三文字づゝとり出す順列は $_4P_3=4\times3\times2$ 即ち二十四通りであるが、このうち先づ abc, acb, bac, bca, cab, cba の六通りは相等しく、同樣にして abd, adb, dab, dba, bad, bda の六通り、acd, adc, dac, dca, cad, cdaの 六通り、及び bcd, bdc, dbc, dcb, bcd, bdc の六通りはそれぞれ相等しい。故に全部で二十四通りの順列も、順序を無視すればその六分の一即ち僅か四通りの列べ方があるだけである。斯かる列べ方を組合せといひ、n箇のうちからr箇づゝとり出す組合せを

$$nCr$$

と記す。そしてその計算は

$$nCr=\frac{nPr}{r!}$$

なる公式に從つて行へる。蓋しn箇のうちからr箇づゝとり出す組合せの各々に於て、其中のr箇を種々の順序に列べればr!づゝの順列を得られるが、組合せの各々について順序を變更するとすれば、全部ではn箇のうちからr箇づゝとり出す順列になつて了ふ。（r!とは $1\times2\times3\times\cdots\cdot r$ のことで、これをrの階乘といふ。即ち

$$r!nCr=nPr$$

であり、これを變形すれば前記の公式が得られるのである。例へば上例の四文字から三文字づゝとり出す組合せは

$$_4C_3=\frac{_4P_3}{3!}=\frac{4\cdot3\cdot2}{1\cdot2\cdot3}=4\ \text{通り}$$

である。

組合せに關する如上の知識から、吾人は二項式の展開を、延いて誤差出現の法則を明かならしめうるのである。四枚の錢貨を投げた場合の表裏の現はれ方は既に說明した通り $(a+b)^4$ の展開に於ける各項の係數に合致するが、この係數は $a^4+4a^3b+6a^2b^2+4ab^3+b^4$ 即ち1、4、6、4、1である。これは二項定理

$$(a+b)^n=a^n+na^{(n-1)}b+\frac{4\cdot3}{1\cdot2}a^{(n-2)}b^2+\cdots\cdot$$

に於けるnを4とすればよい。

先づ第一の項（即ち全部が表なる場合）の係數が1なること、換言すればその出現度數の一回なることは、n箇か

らn箇をとつた組合せ${}_nC_n$（この場合は${}_4C_4$）は常に一なることから明かである。第二項（三枚表、一枚表）の係數はn箇から一箇をとり出す組合せと考へられるから${}_nC_1$であり、これは常にnとなる。本例ではnが4であるから從つて係數も4となるのである。第三項（二枚表、二枚裏）の係數は四箇のうちから二箇づゝとり出す組合せであるから、${}_4C_2$即ち$\frac{4\cdot3}{1\cdot2}=6$であり、第四項（表一枚、裏三枚）の係數は${}_4C_3=\frac{4\cdot3\cdot2}{1\cdot2\cdot3}=4$で第二項と同じになり（註）、最後の項（全部裏）は第一項と全く同じ理由から1となるのである。

（註　組合せの原理に從へば${}_4C_3={}_4C_1$即ち一般に${}_nC_r={}_nC_{n-r}$である。何となれば

$$ {}_nC_r=\frac{n(n-1)(n-2)\cdots(n-r+1)}{r!}=\frac{n!}{r!(n-r)!}=\frac{n!}{(n-r)!\,r!}={}_nC_{n-r} $$

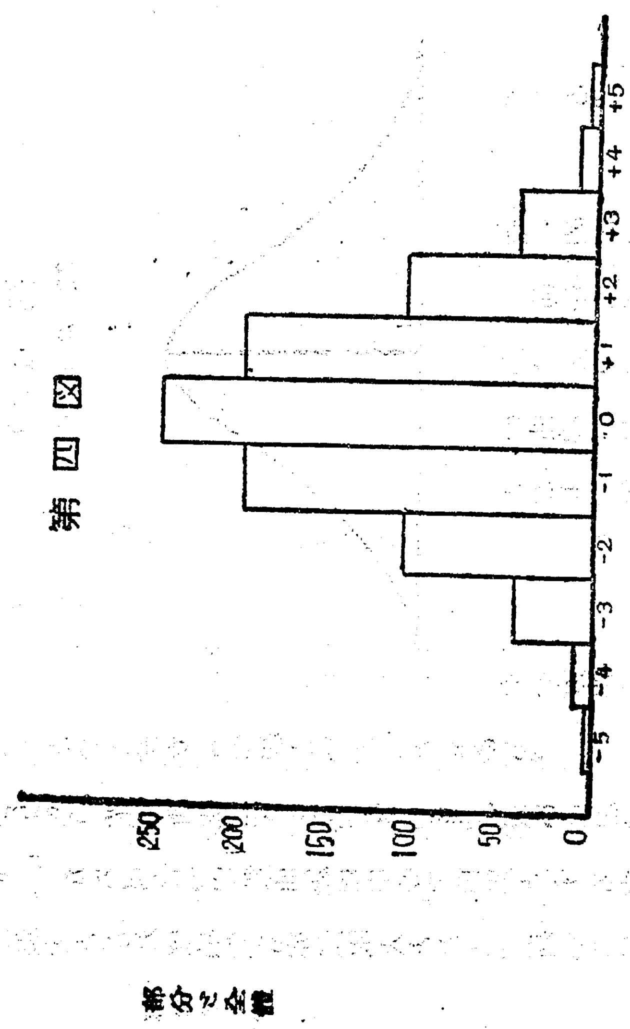
右の諸例に於ては$A=B=\frac{1}{2}$である。蓋し表の出る確率と裏の出る確率は相等しいからである。故に右の諸例は$\left(\frac{1}{2}+\frac{1}{2}\right)^n$なる一般式を以て示すことが出來る。然るにこれはn箇を一回投じたときの出現の確率であるから、もしN回投じたときのそれを求めるにはこれをN倍すればよい。即ち

$$ N\left(\frac{1}{2}+\frac{1}{2}\right)^n $$

である。いま$N=2^{10}$, $n=10$として計算すれば

$$ 1+10+45+120+210+252+210+120+45+10+1=1024 $$

となる。これは$N=2^{10}=1024$回の投錢に於ける表裏出現の確率であり、即ち千二十四回のうち十枚が全部表の場合は一回、九枚表一枚裏の場合は十回、八枚表二枚裏の場合は四十五回……五枚表五枚裏の場合は二百五十二回……一枚表九枚裏の場合は十回、全部裏の場合は一回なることを示す。即ち五枚表五枚裏の場合を中心として左右に對稱的に分布して居るのであつて、これを圖示すれば上の如くである（第四圖）。

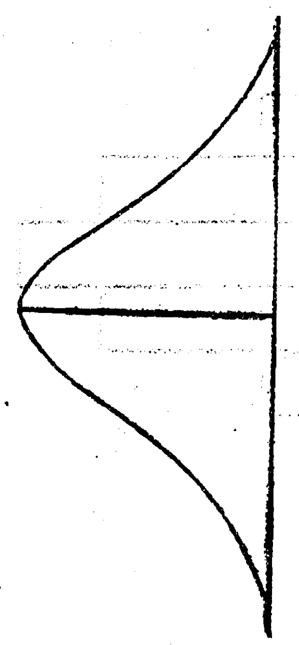


第四図

五枚表五枚裏は十枚の投錢に於ける本質形であるから、その他の場合は總べて誤差と認められる。本質形を標準とすれば六枚表四枚裏は-1、四枚表六枚裏は+1の誤差であり、同樣にして七枚表三枚表は-2、三枚表七枚裏は+2の誤差、……全部表は-5全部裏は+5の誤差である。斯くて千二十四回に於て本質形が二百五十二回出現する間に、-1及び+1の誤差は夫々二百十回、-2及び+2のそれは夫々百二十回、……-5及び+5のそれは夫々一回出現する事が判るのである。

三、誤差曲線　さて二項式$(a+b)^n$を展開した場合の項の數は$n+1$である。故にnが大となるに從つて項數は增加し、nが

無限大となれば項數も亦無限大となる。斯かる場合には上記の棒圖表は度數曲線となる筈である。そして特に$a=b=\frac{1}{2}$なる場合の度數曲線は中心を頂點とする完全な左右對稱形をなし、これを正常曲線又は正規曲線（Normal Curve）、確率曲線（Probability Curve）或ひは誤差曲線（Curve of Errors）と呼ぶ。またこの曲線の創始者ガウスの名をとつてガウス曲線といふこともある。



第五図
誤差曲線

その方程式は

$$y=f(x)=\frac{h}{\sqrt{\pi}}e^{-h^2x^2}=\frac{1}{\sqrt{2\pi pqn}}e^{-\frac{x^2}{2pqn}}$$

となり、誤差の、延いては大數法則の種々の性質は凡てこの曲線、並びにこの曲線の包む面積から導かれるのである。如何にしてこの曲線が導出されるか。その理論は次の如くである。

前條に述べた通り誤差の起る確率は誤差の大小によつて決定されるから、數學的に言へば誤差の起る確率は誤差の函數である。即ち

$$y=f(x)$$

である。これよりして誤差xの起る確率は

$$f(x)dx$$

であり、誤差が$x=a$, $x=b$の間にある確率pは

$$p=\int_a^b f(x)dx$$

である。いま誤差x_1の起る確率を$f(x_1)dx_1$ 誤差x_2の起る確率を$f(x_2)dx_2$‥‥誤差x_nの起る確率を$f(x_n)dx_n$とすれば、この誤差の總べてが同時に起る確率Pは

$$P=f(x_1)f(x_2)\cdots\cdots f(x_n)dx,\ dx_2\cdots\cdots dx_m$$

或ひはこれを對數に書改めた

$$\log P=\log f(x_1)+\log f(x_2)+\cdots\cdots+\log(x_n)+\log dx_1+\log(x_2)+\cdots\cdots+\log(x_n)$$

である。然るにx_1 x_2 ‥‥ x_nはXを測る場合の誤差であるから何れもXの函數に外ならず、從つてこれをxについて微分すれば

$$\frac{d\log p}{dX}=\frac{d\log f(x_1)}{dx_1}\ \frac{dx_1}{dX}+\cdots\cdots+\frac{d\log f(x_n)}{dx_n}\ \frac{dx_n}{dX}$$

となる。故にpの極大値は右方程式を零と置く事によつて求められる。然るにx_1 x_2 ‥‥ x_nは夫々觀測値M_1 M_2 ‥‥ M_nと眞値Xとの差である、即ち$M_1-X=x_1$, $M_2-X=x_2$ ‥‥である。故に右式の

$$\frac{dx_1}{dX}=\frac{d(M_1-X)}{dX}=-1$$

$$\frac{dx_2}{dX}=\frac{d(M_2-X)}{dX}=-1$$

等となり、よつて

$$\frac{d\log f(x_1)}{dx_1}+\frac{d\log f(x_2)}{dx_2}+\cdots\cdots+\frac{d\log f(x_n)}{dx_n}=0$$

或ひはこれを變形して

$$\frac{d\log f(x_1)}{x_1\,dx_1}x_1+\frac{d\log f(x_2)}{x_2\,dx_2}x_2+\cdots\cdots+\frac{d\log f(x_n)}{x_n\,dx_n}x_n=0$$

然るに誤差 x_1 x_2 …… x_n は合計すれば零である。蓋し誤差は正負が相等しいからである。この事から次の比例式が成り立つ。

$$\frac{d\log f(x_1)}{x_1\,dx_1}=\frac{d\log f(x_2)}{x_2\,dx_2}=\cdots\cdots=\frac{d\log f(x_n)}{x_n\,dx_n}=K$$

いま x_1, x_2, …… x_n を一般的に x_1 で示せば

$$\frac{d\log f(x_1)}{dx_1}=Kx_1$$

となり、これを積分すれば

$$\log f(x_1)=\frac{1}{2}Kx_1^2+C$$

$$\therefore\ f(x_1)=Ce^{\frac{k}{2}x_1^2}$$

又は單に $f(x)=Ce^{\frac{k}{2}x^2}$

この $f(x)$ は減少函數であるから（誤差が大となるに從つて確率は小となるから）$\frac{1}{2}K$ は負數である。いまこれを $-h$ で示せば

$$f(x)=Ce^{-h^2x^2}$$

この式の常數Cを決定するには次の如き積分法を使用する。$x=-\infty$, $x=+\infty$ との間に起る誤差の確率は1であるから

$$\int_{-\infty}^{+\infty}f(x)dx=1=\int_{-\infty}^{+\infty}Ce^{-h^2x^2}dx$$

この積分變數 x を y と改めれば

$$1=\int_{-\infty}^{+\infty}Ce^{-h_2y_2}dy$$

$$\therefore\ 1=\int_{-\infty}^{+\infty}Ce^{-h^2x^2}dx\int_{-\infty}^{+\infty}Ce^{-h^2y^2}dy$$

$$=C^2\int_{-\infty}^{+\infty} e^{-h^2x^2}dx\int_{-\infty}^{+\infty} e^{-h^2y^2}dy$$

$$=C^2\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} e^{-h^2(x^2+y^2)}dxdy$$

これを變形すれば、

$$1=C^2\int_0^{2\pi}\int_0^{\infty}e^{-h^2\gamma^2}\gamma d\gamma d\theta=C^2\int_0^{2\pi}\left[\frac{-1}{2h^2}e^{-h^2\gamma^2}\right]_0^{\infty}d\theta$$

$$=C^2\frac{1}{2h^2}\int_0^{2\pi}d\theta=\frac{C^2}{2h^2}\Big[\theta\Big]_0^{2\pi}=\frac{C^2}{2h^2}2\pi=\frac{C^2\pi}{h^2}$$

$$\therefore\quad C=\frac{h}{\sqrt{\pi}}$$

これを $f(x)=Ce^{-h^2x^2}$ に代入すれば

$$f(x)=\frac{h}{\sqrt{\pi}}e^{-h^2x^2}\quad\cdots\cdots(1)$$

いま $h=\dfrac{1}{\sqrt{2pqn}}$ と置けば右式は

$$f(x)=\frac{1}{\sqrt{2pqn\pi}}e^{-x^2/2pqn}\quad\cdots\cdots\cdots(2)$$

となる。然るに二項式 $(p+q)^n$ に於ては

$$\sqrt{pqn}=\sigma\cdots\cdot(\sigma\text{は標準偏差})$$

であるから、次の如く書改めてもよい

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2}\cdots\cdots\cdots(3)$$

更に度數をNとすれば

$$f(x)=\frac{N}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2}\cdots\cdots\cdots(4)$$

とも書ける。この(1)(2)(3)又は(4)は畢竟同じもので、これを誤差曲線といふのである。

與へられた材料に右公式を適用するには

$$y_2=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2}$$

を x/σ の種々な値について算出した計算表を使用する(別表一參照)。この表では x/σ が零なる場合、即ち算術平均の位置では $-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2=0$

第一表

$$y=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2}$$

e	0	1	2	3	4	5	6	7	8	9
0.0	0.39894	39892	39886	39876	39862	39844	39822	39797	39767	39733
0.1	0.39695	39654	39608	39559	39505	39448	39387	39322	39253	39181
0.2	0.39104	39024	38940	38853	38762	38667	38568	38466	38361	38251
0.3	0.38139	38023	37903	37780	37654	37524	37391	37255	37115	36973
0.4	0.36827	36678	36526	36371	36213	36053	35889	35723	35553	35381
0,5	0,35207	35029	34849	34667	34482	34294	34105	33912	33718	33521
0,6	0.33322	33121	32918	32713	32506	32297	32086	31874	31659	31443
0.7	0.31225	31006	30785	30563	30339	30114	29887	29659	29431	29200
0.8	0.28969	28737	28504	28269	28034	27798	27562	27324	27086	26848
0.9	0.26609	26369	26129	25888	25647	25406	25164	24923	24681	24439
1.0	0.24197	23955	23713	23471	23230	22988	22747	22506	22265	22025
1.1	0.21785	21546	21307	21069	20831	20594	20357	20121	19886	19652
1.2	0.19419	19186	18954	18724	18494	18265	18037	17810	17585	17360
1.3	0.17137	16915	16694	16474	16256	16038	15822	15608	15395	15183
1.4	0.14973	14764	14556	14350	14146	13943	13742	13542	13344	13147
1.5	0.12952	12758	12566	12376	12188	12001	11816	11632	11450	11270
1.6	0.11092	10915	10741	10567	10396	10226	10059	09893	09728	09566
1.7	0.09405	09246	09089	08933	08780	08628	08478	08329	08183	08038
1.8	0.07895	07754	07614	07477	07341	07206	07074	06943	06814	06687
1.9	0.06562	06438	06316	06195	06077	05959	05844	05730	05618	05508
2.0	0.05399	05292	05186	05082	04980	04879	04780	04682	04586	04491
2.1	0.04398	04307	04217	04128	04041	03955	03871	03788	03706	03626
2.2	0.03547	03470	03394	03319	03246	03174	03103	03034	02965	02898
2.3	0.02833	02768	02705	02643	02582	02522	02463	02406	02349	02294
2.4	0 02239	02186	02134	02083	02033	01984	01936	01889	01842	01797
2.5	0.01753	01709	01667	01625	01585	01545	01506	01468	01431	01394
2.6	0.01358	01323	01289	01256	01223	01191	01160	01130	01100	01071
2.7	0.01042	01014	00987	00961	00935	00909	00885	00861	00837	00814
2.8	0.00792	00770	00748	00727	00707	00687	00668	00649	00631	00613
2.9	0.00595	00578	00562	00545	00530	00514	00499	00485	00471	00457
3.0	0.00443	00430	00417	00405	00393	00381	00370	00358	00348	00337
3.1	0.00327	00317	00307	00298	00288	00279	00271	00262	00254	00246
3.2	0.00238	00231	00224	00216	00210	00203	00196	00190	00184	00178
3.3	0.00172	00167	00161	00156	00151	00146	00141	00136	00132	00127
3.4	0.00123	00119	00115	00111	00107	00104	00100	00097	00094	00090
3.5	0.00087	00084	00081	00079	00076	00073	00071	00068	00066	00063
3.6	0.00061	00059	00057	00055	00053	00051	00049	00047	00046	00044
3.7	0.00042	00041	00039	00038	00037	00035	00034	00033	00031	00030
3.8	0.00029	00028	00027	00026	00025	00024	00023	00022	00021	00021
3.9	0.00020	00019	00018	00018	00017	00016	00016	00015	00014	00014

即ち

$$e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2}=1$$

となるから、

$$\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2}=\frac{1}{2.5066}=0.39894$$

である。確率曲線のY座標は右の點を頂點として次第に低くなつてゆくから、xが大となるに從つてY座標（y_2）の値は小さくなるのである。

この公式の適用例として既述の十枚の投錢 $1024\left(\frac{1}{2}+\frac{1}{2}\right)^{10}$ を採つて見やう。その場合の表裏出現の狀態は次の第二表の如くであつた

第二表

誤差	度數
−5	1
−4	10
−3	45
−2	120
−1	210
0	252
+1	210
+2	120
+3	45
+4	10
+5	1

先づσを求むるに、二項式$(p+q)^n$のσは$\sqrt{npq}$であるから、本例では$\sigma=\sqrt{10\times\frac{1}{2}\times\frac{1}{2}}=\sqrt{2.5}=1.58$である。故に誤差が-5のときは$\frac{x}{\sigma}=\frac{-5}{1.58}=-3.16$となり、從つて第一表から$y_2=0.00271$を得る。これに$\frac{N}{\sigma}$

を乗ずればよいのであつて、$\frac{N}{\sigma}=\frac{1024}{1.58}=648.1$ であるから、誤差-5 に於ける「求める確率曲線のY座標」は $648.1\times0.0271=2$ となる。同様の計算を誤差-4、-3……の全部について行へばよいのであつて、これを表示すれば次の第三表となる。

第　三　表

誤差(xi)	度數	$\frac{N}{\sigma}$	$t=\frac{(xi-A)}{\sigma}$	y_2	修正された度數 $\left(y_2\frac{N}{\sigma}\right)$
−5	1	648.1	−3.16	0.00271	2
−4	10	〃	−2.53	0.01625	11
−3	45	〃	−1.90	0.06562	43
−2	120	〃	−1.27	0.17810	115
−1	210	〃	−0.63	0.32713	212
0	252	〃	0	0.39894	259
+1	210	〃	0.63	0.32713	212
+2	120	〃	1.27	0.17810	115
+3	45	〃	1.90	0.06562	43
+4	10	〃	2.53	0.01625	11
+5	1	〃	3.16	0.00271	2

四、誤差曲線の包む面積　誤差の起る確率は誤差の函數即ち $f(x)dx$ であるから、誤差が $x=-a$ と $x=+a$ との間にあることの確率 p は

$$p=\int_{-a}^{+a}f(x)dx$$

である。然るに

$$f(x)=\frac{h}{\sqrt{\pi}}e^{-h^2x^2}$$

であるから、

$$p=\frac{h}{\sqrt{\pi}}\int_{-a}^{+a}e^{-h^2x^2}dx$$

と記すことが出來る。この式に於て $hx=t$ とおけば

$$p=\frac{1}{\sqrt{\pi}}\int_{-ah}^{+ah}e^{-t^2}dt=\frac{2}{\sqrt{\pi}}\int_{0}^{ah}e^{-t^2}dt$$

となり、p の大さは ah の如何によつて決定されることが判る。換言すれば p は ah の函數である。故に二つの觀察に於ける p の値が等しい場合、誤差 a の値が小さければ小さいほど h は大でなければならぬ。この理由から h をば觀察の精密度又は精度(Precision)といふのである。次の第六圖はこれを示したものである。

このhは前述の如く

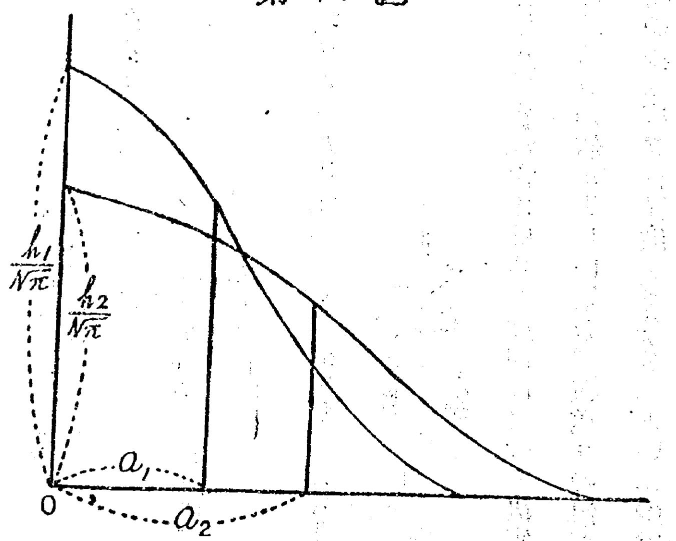
$$h=\frac{1}{\sqrt{2pqn}}$$

であるが、これを直接に算出することは不可能である。然るに標準偏差σは正常分布に於ては $\sigma=\sqrt{pqn}$ であるから、これよりして

$$\sigma^2=pqn=\frac{1}{2h^2}$$

なる式が成り立つ。斯くて實際には標準偏重は精密度に反比例すると見てよいのであつて、誤差曲線に於ける標準偏差を特に標準誤差(Standard Error)といふのはこれが爲である。正常分布に於て標準偏差だけを算術平均の左右にとればその範圍即ち $Ma\pm\sigma$ の間に全度數の約六八・二七%を含むが、これと全く同様に標準誤差を算術平均の左右にとれば、起る誤差全體(即ち誤差曲線の包む總面積)の六八・二七%を含む。また標準偏差の場合と同じく、算術平均の左右に2σをとれば全體の九五・四五%を、また3σをとれば九九・七三%を含むのである。

第六図



然るに誤差曲線は中央を境として左右全く對稱的であるから、求める面積の半分だけを計算すれば、それを二倍する事によつて容易に求める面積を知ることが出來る。即ち $Ma\pm\sigma$ を知るには單に $Ma+\sigma$ を知ればよく、その

値は前記の六八・二七%の半分即ち三四・一三四%である。同様にして $Ma\pm2\sigma$ は $Ma+2\sigma$ 即ち四七・七二五%(九五・四五%の半分)から、また $Ma\pm3\sigma$ は $Ma+3\sigma$ 即ち四九・八六五%から求めることが出來るのである。σの各種倍數の範圍内に含まれる斯かる面積は詳しく計算され數學表に纒められてゐるが、この若干を表示すれば次の如くである。

第四表

σの倍數	包む面積(%)
0.1	3.983
0.2	7.926
0.3	11.791
0.4	15.542
0.5	19.146
0.6	22.575
0.7	25.804
0.8	28.814
0.9	31.594
1.0	34.134
1.5	43.319
2.0	47.725
2.5	49.379
3.0	49.865
3.5	49.977

右表に就て見るに $0.6\sigma=22.575\%$ であり、$0.7\sigma=25.804\%$ である。よつてこの二つ(〇・六と〇・七)の間の或る數に於てはその含む面積は25%となる筈で、これを二倍すれば50%即ち起る誤差全體の丁度半分を含む範圍となることが判る。この或る數は計算によれば〇・六七四五であつて、從つて $0.6745\sigma=25\%$ と記す事が出來る。斯くて標準誤差の〇・六七四五倍を特に確率誤差(Probable Error)といひ、一般にP.E.と記す。即ち $P.E=0.6745\sigma$ である。確率誤差よりも小さい誤差の確率と大きい誤差の確率とは全く相等しく、何れも$1/2$である。なほこの〇・六七四五は約$2/3$であるから、近似的には

$$P.E=\frac{2}{3}\sigma$$

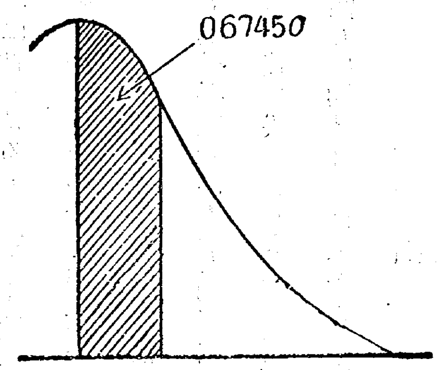
と記すことも出來る。

部分的觀察の値と全域の値との間には斯く種々なる程度の誤差が介在するから、この誤差の範圍は必ず明瞭に記されて居らねばならぬ。然るに右の如く誤差の範圍は標準誤差又は確率誤差によつて示されるから、この何れかを觀察値に添へて置けばよいのであつて、假りに觀察値をAとすれば

$A\pm\sigma$ 又は $A\pm$P.E.

と記すことによつて誤差の範圍を限定するのである。標準誤差と確率誤差とはその値が異るから、右の二つの記し方は勿論その意味を異にする。例へば大なる全域から幾多の部分を抽出してその各々について平均値を求めたとすれば、これら平均値は決して全域の眞の平均値と一致するものではなく、且つこれら平均値相互の間に偏差のあることは免れない。併し前述の理由によつて、これら平均値は誤差法則に從つて分布するから、これら平均値の平均値は全域の眞の平均値と合致する傾きがある。

第七図



例へば十歳の男子の平均身長を求める場合、その全域を百萬人とすれば、到底その全部を調査する事は不可能であらうから、假りに百人づゝの組を百だけ抽出してその各々の組につき平均値を求めれば、この百の平均値は素より一致するものではないが、これら平均値を更に平均すれば大體に於て全域の平均値と一致するものと見てよいのである。然るにこれら部分的平均値の平均を以て直ちに全域の平均値と見做すことは、起るべき誤差を無視したことで、決して正しい方法ではない。然るにこれら幾多の平均値は前述の如く誤差法則に從つて分布するから、起るべき誤差の範圍は標準誤差又は確率誤差として計算する事が出來る。いまこれら平均値の平均をA、その標準誤差をσとし

$A\pm\sigma$

と記せば、これは觀察された百の平均値の約六八%は右の範圍内に在ることを意味し、また

$A\pm3\sigma$

と記せば、この百の平均値の殆ど全部はこの範圍内に在ることを意味する。則ちσが小さければ小さいほど、觀察された諸平均の平均は相互に大なる差がなく、從つてAが全域の眞の平均値に近似してゐる證據と見られるのである。

右は標準誤差を用ひた場合であるが、いまその代りに確率誤差 P.E. 即ち 0.6745σ を用ひ

$A\pm$P.E.

と記せば、起る誤差が右の範圍内にある確率は$\frac{1}{2}$なることを意味する。一般には寧ろこの表現法が用ひられるやうである。

時には標準誤差の場合の如く、

$A\pm3$(P.E.)

が用ひられることがある。$3(P.E)=3\times0.6745\sigma$ 即ち 95.698% に該當するから、その面積は 3σ 即ち99.7% には及ばぬとしても、略々全面積と等しいと見てよからう。斯くて $A\pm\sigma$ の代りに $A\pm3(P.E)$ を取つても大なる誤りはないわけで、これを言葉で言へば、確率誤差の三倍に達する偏差は殆ど起らぬと考へてよいといふ事になる。最近では斯かる場合には確率誤差よりも標準誤差がより多く用ひられるやうであるが、何れにしても標準誤差と確率誤差との關係を熟知してゐれば別に不都合を來す惧れはないのである。

右の理を若干例に適用して見やう。

（一）　百枚の錢を投げたとき、表の現はれる枚數は如何。理論的には五十枚であるが、起りうる誤差は確率誤差の三倍であるから、求める答は $50\pm3(P.E.)$ である。然るにこの場合の確率誤差は標準偏差即ち

$$\sqrt{npq}=\sqrt{100\times\frac{50}{100}\times\frac{50}{100}}=5$$

の〇・六七四五倍、即ち三・三七であるから、求める枚數は

$$50\pm3\times3.37$$

即ち四十枚乃至六十枚である。

尤も起りうる誤差の範圍は標準誤差の三倍なりとすれば、右の答は $50\pm3\times5$ 即ち三十五枚乃至六十五枚と改められねばならぬ。この二つの答の間には著しい相違があるが、これは小數現象を扱つた爲であつて、もし百枚の代りに千枚を取れば、標準偏差は一五・八となり從つて $3\sigma=47$, $3(P.E.)=53$ となるから、起りうる誤差の範圍として標準誤差の三倍をとれば四五三乃至五四七枚となり、確率誤差の三倍をとれば四四七乃至五五三枚となつて、二つの答の相違は比較的に少なくなるのである。

（二）　一石炭商が十噸の石炭から百封度の見本を抽出したところ、そのうち五十封度は噸價五弗のA級品、三十封度は噸價四弗のB級品、二十封度は噸價三弗のC級品であつた。右十噸の價格は如何なる限界内に於て決定さるべきか（この問題は Davis & Nelson, Elements of Statistics, p. 194 よりとる）。

この答を求めるに當り、假りに右の見本が正確にその母域たる十噸と同性質なりとすれば、單に右の見本につき加重平均を算出すれば足りる。即ち

$$\frac{50\times\$5+30\times\$4+20\times3\$}{100}=\$4.3$$

である。然るに事實に於ては斯かる假定は許容されないから、起るべき誤差の範圍を確定せねばならぬ。

いま三者のそれぞれの確率誤差を求めるに、

A級品の P.E. は $0.6745\sqrt{100(50/100)(1\text{-}50/100)}=3.37$

B級品の P.E. は $0.6745\sqrt{100(30/100)(1\text{-}30/100)}=3.09$

C級品の P.E. は $0.6745\sqrt{100(20/100)(1\text{-}20/100)}=2.70$

起りうる誤差の範圍をば確率誤差の三倍とすれば、A級品は右十噸中50%±3×3.37%即ち約六〇%乃至四〇%である。同様にしてB級品は三九%乃至二一%、C級品は二八%乃至一三%である。よつて右十噸が最も高い價格を持つのはA級品が最も多く、C級品が最も少い場合であるから、

A級品——六〇%

C級品——一二%

（従つて）B級品——二八%

として計算すればよい。即ち

$5×6噸+$4×2.8噸+$3×1.2噸=$44.80

となる。これが最高價格である。同様にして最低價格は右十噸中A級品が最も少く、C級品が最も多い場合で、その時の割合は

A級品——四〇%

C級品——二八%

（従つて）B級品——三二%

であるから、求める價格は

$5×4噸+$4×3.2噸+$3×2.8噸=$41.20

となる。即ち右十噸は四十一弗二十仙乃至四十四弗八十仙の間で定められねばならないのである。

誤差の範圍は、部分的調査に於ける総べての統計的結果について確定されねばならぬが、その主たるものは次の如くである

(1) 算術平均の標準誤差　$\sigma/\sqrt{N}$

同　確率誤差　$0.6745\ \sigma/\sqrt{N}$

(2) 標準偏差 σ の標準誤差　$\sigma/\sqrt{2N}$

同　確率誤差　$0.6745\ \sigma/\sqrt{2N}=0.4769\ \sigma/\sqrt{N}$

(3) 確率 P_1 の標準誤差　$\sqrt{P_1(1-P_1)/\sqrt{N}}$

同　確率誤差　$0.6745\sqrt{P_1(1-P_1)/\sqrt{N}}$

(4) 二つの獨立變數 x 及び y の和又は差の標準誤差（ε_x 及び ε_y は標準誤差）　$\sqrt{\varepsilon_x^2+\varepsilon_y^2}$

同　確率誤差　$0.6745\sqrt{\varepsilon_x^2+\varepsilon_y^2}$

(5) 相關係數 r の標準誤差　$(1-r^2)/\sqrt{N}$

同　確率誤差　$0.6745[(1-r^2)/\sqrt{N}]$

×　×　×　×

如上の誤差法則によつて吾人は部分と全體との關係を明確ならしめうることを知つた。即ち任意無差別に選擇された部分即ち任意試料は、誤差法則の示す誤差範圍內に於て全體即ち母域の縮圖と認められるといふことであつて、換言すれば誤差の範圍內に於て部分から全體を推して差支へないのである。唯だ注意すべきは誤差法則の完全に作用するのは、抽出された部分が全體に對して比較的大なる割合を示す場合であり、且つ全體そのものが極めて大なる集團たる場合に限られることである。小なる部分については誤差法則の作用する理由がないから、特別の場合を除いては統計的方法に對象たるを得ないし、また小なる全體はこれを大なる母域の一部分として取扱ふ以外には適當な方法はない。故に如何なる場合にも統計的方法の對象が大數的現象たる可きことは依然として眞である。

いま如上の前提が充たされたとすれば、部分より求めた結果も極めて大なる眞實性を持つのである。唯だこの場合にも誤差の範圍は常に明確に記されねばならぬ。この事については既に述べた。また試料と母域との間に起りうる誤差は、標準誤差乃至は確率誤差の三倍を超えざることから、吾人は同種の部分に於ける統計的結果に相違ある場合、この相違が全く偶然的なものか又は根本的なものかを判定することが出來る。例へば十年前に全國壯丁から抽出したａ人の試料に於て平均體重がW_1であつたとする。いま本年度の壯丁についてｂ人の試料を抽出したところ、平均體重はW_2となつたと假定すれば、この平均體重の差即ち W_1-W_2 は單に抽出調査に於ける偶然的結果であるか、又は事實に於てこの十年間に壯丁の體重そのものが變化した爲であるかを決定することは極めて必要である。然るにW_1に於ける標準偏差をσ_1、W_2に於けるそれをσ_2とすれば、平均の差即ち W_1-W_2 の標準誤差σは

$$\sigma=\sqrt{\frac{\sigma_1^2}{a}+\frac{\sigma_2^2}{b}}$$

によつて計算される。故にもしこのσが平均の差の三倍以上ならば、本年度の壯丁と十年前の壯丁とは實際に體重に相違を來したと結論してよいのであり、反對にもしこのσが平均の差の三倍に達しないならば、體重が變化したのではなくて單に試料抽出に基く偶然的要素の作用に基くと結論して差支へないのである。同樣の判斷が同種の二つ又はそれ以上の構成比や相關係數等についても下せることは言ふ迄もない。これらに就ては別の機會に論及するであらう。

最後に特に注意せねばならぬことは、試料からの推定は母域の構造に對して行へるのみであつて、母域の大さそのものを試料から決定する事は不可能だといふことであ。るこの點についてはワーゲマンの次の一文が甚だ印象的である。

Durch die gegenseitige Vertretung zwischen der Gesamtheit und ihren Teilen und damit dieser Teile untereinander entstehen darüber muss man sich klar sein nur Aussagen über die Massenstruktur, niomals über den Umfang der Massen. Inklusion und Repräsentation können z. B. die Bevölkerungsdichte, die Geburtenziffer, den Durchschnittspreis veranschlagen, nicht aber die absoluten Bevölkerungszahlen, die Anzahl der Geburten, die Umsatzsummen angeben. Nur wenn von der zu vertretenden Masse weitere Anhaltspunkte gegeben sind, kann von struktur auf ihrem Umfang geschlossen werden. So lässt sich z..B. aus der Bevölkerungsdichte die Gesamtbevölkerung errechnen, wenn man ihre Gebietsfläche kennt; so kann die Grösse der Bevölkerung auch aus der Besetzung einer Altersklasse, etwa der schulpflichtigen Kinder, bei kenntniss des Altersaufbus erschlossenwerden. (Wagemann, Narrenspiegel der Statistik. p. 176)

（一九三九・一・一四）