大正四年六月二十一日第三種郵便物認可（毎月一回一日發行）

京都帝國大學經濟學會

經濟論叢

第三十卷 第六號

昭和五年六月一日發行

中位數の本質

益田熊雄

第一　緒言

平均は統計的觀察の殆んど凡ての分野に於て最も多く用ひらるる所の方法であつて、統計法の最も樞要なる地位を占むるものである。[1] 蓋し統計の最根本要件たる大量觀察によりて其の大量の性質を究知せんが爲には、先づ最初に此の平均方法によるを最も便とするが故である。統計學を稱して平均の學なりとなす[2]の所以は實に此處に存するのである。

されば統計學に於ては平均法の研究は其の中樞を占め夙に多くの學者により檢討せられつゝある所であつて、I. Fisher の如きは九千種以上の平均の方法を網羅して、之より彼の所謂理想算式を抽出して居るのである。[3]

此處に取扱はんとする中位數（Median）も亦一種の平均方法であるが、此の平均値の計算の過程は甚だ原始的であつて一見甚だ粗雜なる平均値の如くであるが、それだけ他面には微妙なる性質を具有し、その利用方法にして宜しきを得ば甚だ優れたる平均値となり得るのである。

されど中位數は發見以來日尚ほ淺く、(註)之が研究は淺くして只その形式論に止り、之が本質に

1) Žižek: Die Statistischen Mittelwerte. s. 1.
2) E. Y. Edgeworth: On methods of Statistics (Journal of the Royal St. Society 1885) p. 182.
3) I. Fisher: The making of Index Number.

觸れてゐるものは甚だ少いのである。之が本質を究めんとするには、平均する事それ自體の統計學的意義、或は Gaus の誤差理論、或は Probability の究明から出發しなければならぬのであるが、此等はよく淺學の私の企て能はざる所であつて之が研究は他日を期し、本論文に於ては唯 Prologue として聊かその本質に觸れて見たいと思ふ。

註、中位數が Central tendency を測る爲に重要なりと云ふ事が發見せられたのは約五十年前であつて、其の後 gaus, Encke, Quetelet, Galton, Fechner 等の學者により開墾せられて行つたが[4]、之が初めて Median なる言葉を以て呼ばれた時は乍遺憾知るを得ない、但し1880年に於ては Galton は Medion なる語の代りに Medium なる語を以て呼びたるに拘らず Journal of the Royal Statistical Society, June, 1888 にある Edgeworth の論文 "Some New Methods of Measuring Variation in General Prices." には Median なる語を用ひて居るから、Median なる語が最初に用ひられたのは1880～1888年である事は明かである。

第二　中位數の意義

中位數 (Median) とは或る變數の各個別値を大きさの順に並べたる時にその中央の項の値を云ふ。故に中位數の上下には夫々同數の項が存在して居る譯である。之を代數的に云へば、中位數とは項數 n なる數列に於て夫等の項を大きさの順に並べたる時に $\frac{n+1}{2}$ 番目に當る項の値を云ふ[1]。故に n が奇數なる場合には最小又は最大の項より $\frac{n+1}{2}$ 番目に當る項が中位數であるが、n が偶數なる場合には $\frac{n+1}{2}$ 番目に當る項は實在しないから、中位數は $\frac{n}{2}$ 番目と $\frac{n}{2}+1$ 番目の項の間に存する値である事が知られる。從つて $\frac{n}{2}$ 番目と $\frac{n}{2}+1$ 番目との項の値が相等しき場合にはその値が直ちに中位數となるが、然らざる場合には普通には普通は $\frac{n+1}{2}$ 番目の項の値として中央の二項の算

4) H. M. Walker: Studies in the History of Statistical Method.
1) 小林新：經濟統計學 p. 187—188.

術平均を以てするのである。(幾等平均を以てしてもよい譯である。)

中位數にして此くの場き計算過程を有する以上、その外觀的形態は甚だ幼稚なる平均の觀念に基けるが如くであるが、抑々中位數は蓋然誤差 (Probable error) の算定[2]、(註)或は生物學の正常分布曲線 (Normal curve) にその端を發して居るのであつて、data が中數値の周りに密集する傾向の甚だ大なる場合に用ひられ始めたのであるから、連續函數 (Continuous Series) の場合を豫想せる平均の方法である。從つて連續函數の個別値の平均値としては true value 表はす事が出來るが、不連續函數 (Discontinuous Series)或は聯絡性少き個別値の平均方法としては不適當であり、亦意義も少い。故に data の分散程度の大小に從つて中位數の平均値としての意義も段々と減少して行く事となる。

註、Probable error は Normal curve of error に從つて對稱的に分散せられる。[3]

中位數に從屬せしめて考へ得らるるものに、四分位數 (Quartiles) 十分位數 (Deciles) 百分位數 (Percentiles) 等がある。或る變數の個別値を大きさの順に並べたる時に最小の項より數へて全體の項の四分の一の所に位する項の値を下の四分位數(Q_1)四分の三の所に位する項の値を上の四分位數(Q_3)と云ふ。同樣に全體の項を十等分する所に位する項の値を十分位數、百等分する所に位する項の値を百分位數と云ふ。又場合によつては八分位數、十六分位數、千分位數等も用ひらるる事がある。此の中で最も多く用ひらるるものは四分位數であつて、上下の四分位數と中位數とを以て全體の項を四等分する譯である。[4]「故に全體のものの半分は Q_1 と Q_3 との間にあるのである。故に範圍 $\overline{Q_1Q_3}$ の大小に依つて撒布の程度の大小を測る一つの方法とする事が出來る。……實際生物統計や教育統計に於てはこの範圍 $\overline{Q_1Q_3}$ の半分

$$Q=\frac{Q_3-Q_1}{2}$$

2) Walker: ibid. P. 83.
3) Persons: Indices of General Business Condition. p. 126.
4) 小倉金之助：統計的研究法 P. 93—94.

なる値を以て撒布の程度を測ること屢々であつて、此の値Qが所謂四分偏差と呼ばれるものである。」

中位數の統計學上の地位は既に緒言に於て述べし如く、統計學上最も重要なる平均論の一部を構成する事は勿論であるが、平均論の中に於ても比較的新らしく考案せられたる平均方法であつて、而もその性質上代數的取扱の不可能なる理由も加はり、他の平均方法に比してその研究は餘り深く行はれては居ないのである。

然しその性質上 data が密集せるものの平均方法としては秀でたるものであるから、密集性多き data を比較的に多く藏する傾向ある生物學遺傳學人類學に於ては可成り廣汎なる利用範圍を有するが故に、此等の自然科學に於ては中位數の研究は實用上よりの要求によつて相當に研究せられてゐるのである。然し精神科學に於ては教育心理學に於てその利用を見たるのみであつて、經濟統計學に於ては僅かに勞賃や物價の研究に、或は季節的變動の研究[5](註)に於て利用範圍を有するのみである。然し近時數理的經濟統計學の勃興に伴ひ、除々にその利用範圍を擴張する事と思はれる。

註、季節的變動を測定するに Median-Link-Relative Method と云ふ方法がある。此の方法は Persons により案出せられ Harvard 大學の景氣豫測に用ひられて居る方法であつて、最近季節的變動の研究によく用ひらるる樣になつて來た。此の方法の特色は、極端なる値の影響を受けざる Median の性質を巧みに應用し、普通の Link-Relative Method に Median を取入れて極端なる値の影響を避け、亦 Secular trend をも消去して、正常の季節的變動を求むる所にある。

亦 Secular trend を求むる方法としての移動平均法にも Median を取入れて極端値の影響を消去し、正常の Secular trend を求むる Method of Moving Median と云ふ方法もある。此の方法はニューヨークの聯邦準備銀行により用ひられ、Jordan の景氣豫測に引用せられて居る。

5) Persons : ibid.
Chaddock : Principles and Methods of Statistics p. 344-347.
Secrist : An Introduction to Statistical Methods p. 449-450 p. 450-455.
Rietz : Handbook of mathematical Statistics p. 152.

第二　中位數の決定方法

中位數の決定方法は data の性質により種々異るが、その根本の考へ方は皆規を一にするものであつて、要は大きさの順に並べられたる data の中央の項の値を求むるにある。

data が度數表に集められて居らない場合には、先づその data を大きさの順に並べその項數 n が奇數の場合には $\frac{n+1}{2}$ 番目の項の値が卽ち中位數である。n が偶數なる場合には中位數は $\frac{n}{2}$ 番目と $\frac{n}{2}+1$ 番目との間にある事は明かである。此の兩項が同値なる場合には其の値が直ちに中位數であるが、然らざる時には普通は兩者の算術平均を以てする。然し此の場合算術平均を以てする事は便法であつて幾何平均を以てするも差支へは無い譯である。此等の點に就いて餘り正確にするも中位數そのものゝ性質より考ふれば餘り意義は無いが、純理論的に云へば將に補間法を以てすべきであらう。

然し此處に一段の考慮を要する場合がある。例へば data が 1, 2, 3, 4, 5, 6, 20, 30, 40, 50, 60, 70,である場合には項數は十二であつて偶數であるから、六と二十との間の任意の數はその上下に夫々同數の卽ち六個の項を有するから、凡て中位數となり得るのであつて、その可動範圍は餘りに廣きに失する。亦普通なさるる方法によりて六と二十との算術平均十三を以て中位數とするも餘り意義はない。(算術平均は二四・二五)亦二・五・六と云ふ如き數列に於て五を以て中位數となすも之亦同樣に意義なき事である。此等は只擬制的の中位數に過ぎず、十三或は五は何等此等數列の

代表値と認むべき根據が無い。此くの如く假令人工的に平均値を求め得たとしても、その平均値としての價値は甚だ低いのである。此くして中位數は不連續函數又は小なる數列の場合[1]には全く意義なき平均値となり終るのである。故に中位數の利用に當つてはよくその data を吟味する必要があるのであつて、如何なる series にも直ちに用ひ得らるる性質のものではないのである。

data が度數分布表に集められて居る場合には各階級内に於ける度數は均等に撒布せられて居ると云ふ假定に立つて、比例法を用ひて中位數を求めるのである。[2]此の場合に連續數列及び狹い階級に分類せられたる不連續數列の場合には、階級内の撒布は均等であると假定しても、此の假定から生ずる誤差は極めて小さく、neglect しても差支へはないのである。

圖法により中位數を決定するには若し分布曲線が理想的に描かれて居る場合には x 軸への垂線によつて其の面積を二等分する樣な位置が中位數である。[3] 何となれば階級に屬する個體の數はその二分せられたる各々の面積に比例するからである。然し此くの如くに面積を二等分する線を引く事は實際は甚だ困難であつて、亦曲線そのものを描く場合にも誤差を生ずるものであるから、此の方法は實際には殆んど用ひられない。度數分布表より直ちに圖法を以て中位數四分位數等を求むる方法もあるが之は結果より見れば次に述ぶる累積度數分布曲線の場合と殆んど規を一にするものであるから此處には述べないが Chaddock が詳しく說明をなして居るから[4]夫を參照せられ度い。

次に累積的度數分布曲線 (Cumulative Fequency Distribution Curve) について說明する。此の方

1) Yule: An Introduction to the Theory of Statistics. p. 116-117.
2) c. f. Secrist: ibid p. 286-290.
3) 小倉金之助: ibid p. 66-67.
4) Chaddock: ibid p. 63-63 p. 114.

法を用ふるが爲には先づ普通の分布表に與へられたる度數を順次累加して累積的度數分布表を作成し、此の表に從つて累積的度數分布圖を描き之を滑かなる曲線となし、垂直線上の累積的度數の50%の點より水平線を引き、曲線との交點より垂線を立てx軸と交はらしむれば、その交點が卽ち中位數である。四分位數に於ても同樣であつて、25%，75%の點より同樣の過程を行へば、夫々下の四分位數、及び上の四分位數を得るのである。十分位數、百分位數等も普通此の方法によつて求められる。

中位數の定義に從へば累積的度數分布表を成作する場合に、上端より計算しても下端より計算しても同一の中位數を得べき理であるから、上下より計算したる二本の累積的度數分布曲線は中位數の所に於て將に交叉すべき筈である。此く分布の上下兩端より計算せる中位數は一致すべしと云ふ根本原則を建てたのは Fechner が最初である。[5] されど中位數も四分位數も完全に一致する場合は甚だ稀であつて、寧ろ普通は多少ずれる傾向があるのである。曲線には多少の誤差は伴ふものであり、亦階級內の分布は必ずしも假定の如くに均等に分布するものではないから、從つて圖法は唯近似値をのみ表はし得るものであるから、之は寧ろ當然の事と云はねばならぬ。故に普通は下より上への累積度數分布曲線のみが用ひられる。此くの如くに中位數は比例法によりても圖法によりても求め得らるるのであるが、圖法は比例法に比してその性質上誤差の程度が大であつて、比例法よりも不正確たるを免れない。圖法は何れも近似値のみを知り得るのである。

5) Walker: ibid p. 85.

第四　中位數の性質

中位數は既に述べたるが如くに、大きさの順に並べられたる數列の中央の項であるから、大小の順位と中央の項の値に變化なき限り、變化しない。而して凡ての項を度數一と見做す以上には極端なる項の存在を無視するのであるから、極端値の影響は消去せられるのである。中位數は此の性質あるが爲に、或る數列を平均する場合にその數列内の極端なる値の影響を好まざる場合には平均として中位數は甚だ便利である。例へば南北戰爭中の木綿及煙草の如く極端に變動した る商品が存在する場合に、他の平均方法によるならば正常的の物價指數を得られない[1]が、中位數を用ふるならば斯る特殊の商品の極端なる變動より影響を受くる事なく、正常的物價指數を得る事が出來る。中位數に此の性質あるが爲に相當の利用範圍が存するのである。

中位數は凡ての項を考慮する點に於ては算術平均と同樣であるが、後者が極端なる値に直接の影響を受くる點とは異なつて居る。又極端なる値に影響せられない點は並み數(mode)と同樣であるが並み數が most probable value のみを採りて他の凡ての項を全然考慮せざる點とは異る。又算術平均と中位數との形式的の根本相違點は、前者の代數的取扱可能なるに對して後者は夫が不可能なりと云ふ所に存する。此の點に關しては並み數と性質を同じふするものであつて、代數的取扱の可能なる算術平均、幾何平均、調和平均等を凡て mean と稱し、代數的取扱不能なる中位數、並み數を average と稱して、兩者を區別して居るのは妥當である。

1) Edgeworth: Some New Methods of Measuring Variation in General Prices. (Journeal of the Royal Statistical Society. June 1888) p. 360.

分布曲線が完全に對稱的なる bell-shape をなす時には算術平均、中位數、並み數等の凡ての平均値が一致する事は論を俟たない。然し完全に對稱的であつても、分布曲線がU字形をなす場合には算術平均と中位數は一致して最中央に來るが、並み數は最端に行くであらう。

非對稱形の場合には勿論夫等の位置は異なつて來る。而して夫等の位置に就いては確定的の相互關係は存在しないが、少し偏よれる分布曲線の場合に於ては、[2] 並み數と中位數と算術平均とはそれぞれ特色を示すのである。即ち並み數は曲線の最頂點の直下にある、中位數は極端なる項の値には影響を受けないがその項の存在には度數一としての影響を受けるのであるから、極端なる項の存在する方の側へ並み數よりは少し偏よつて居る、又算術平均は極端なる項の値に直接の影響を受けるのであるから中位數よりも更に極端なる項の存在する側へ偏よるのである。故に少し偏よれる bell-shape の分布曲線に於ては中位數は常に並み數と算術平均との間に存在し、經驗的にはその間の間隔に就いては

$$\text{Mean} - \text{mode} = 3(\text{mean} - \text{median})$$ [3]

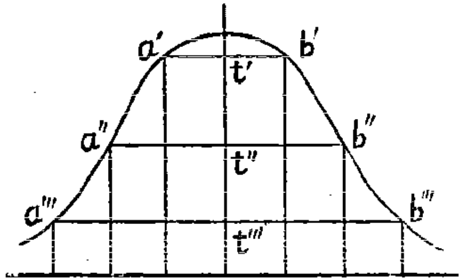
なる關係の存する事が知られて居る。

Edgeworth は中位數の愛用者であるが、中位數が算術平均よりも幾何平均に接近して居るらしく見えるのは不思議であると云つて居るが、Walsh の云ふ所によれば不思議ではない。[4] 即ち中位數は、data が geometric law of dispersion に從つて配列されて居る時には幾何平均にヨリ接近し、Arithmetic law of dispersion に從つて配列されて居る時には算術平均にヨリ接近するのであ

2) Chaddock: ibid p. 136.
3) Chaddock: ibid p. 137.
4) Walsh: The Problem of Estimation. p. 132–133.

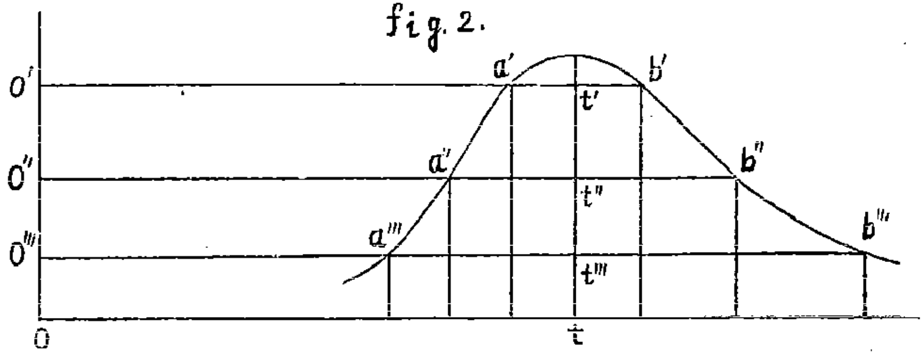
る。而して観測 (observation) は後者の法則に、測定 (estimate) は前者の法則に従つて分配するものである。價格は亦前者の法則により配列される事が知られてゐる。(註) Edgeworth は此の價格

fig. 1.



at = tb

fig. 2.



oa : ot = ot : ob

註、Fig. 1. が Arithmetic law of dispersion であり、Fig. 2. が geometric law of dispersion である。[5]

に對しては幾何平均を薦め中位數を夫に次ぐものとして薦めて居る。然し一八六二年の木綿に於けるが如くに或る貨物が極端なる値を示し、幾何平均が無理にその影響を受くるが如き場合には却つて中位數が薦めらるべきものとして居る。Walsh は之に對して次の如く述べて居る。卽ち價格に對する Edgeworth の說が誤つて居り normal probability curve 卽ち Arithmetic law of dispersion に從つて價格が分散するものとすれば、中位數は算術平均にヨリ接近し、亦若し Edgeworth の說が正しく價格は logarithmic probability curve 卽ち geometric law of dispersion に從つ

5) Walsh : ibid p. 40-41.

て分散するものとすれば、中位數は幾何平均にヨリ接近する事となる。故に Walsh は Edgeworth の説は寧ろ「價格は場合によつて此等二通りの法則に從つて分散する性質のものであり、中位數は價格の此の兩樣の姿に對して自身自働的に調整するから、中位數は價格の平均としては最もよい方法である。故に中位數は最も包含性多きものであり、且妥協性の多いものである。」と改むべきものとして居る。

前に屢々繰返し述べたる所であるが、中位數は Prabable error にその端を發して居るのであるから data が多數である事を要し、亦度數分布表に集められて居る場合には階級の狹き事を必要條件とするのである。階級内の分布は均等なりと云ふ假定に立つて始めて中位數を求め得るのであるから、連續函數に於てもその階級が餘りに廣きに失し又は data が少數なる場合には、その假定は根據を失ふ、假令此の假定に基き中位數を算定し得ても、夫は擬制的の平均値であつて決して true value を示して居るものでは無い。又不連續函數に於ては、觀測が正確且つ、多數であり、目盛小さく階級が狹い時に始めて中位數はよい平均値となり得るが、然らざる場合には實に擬制的の平均値となり終るのであるから、中位數の利用に當つては大いに注意を要するのである。[6]

次に中位數の長所或は短所と考へられる點を總括して列擧し、今後中位數を利用する場合への指針としたい。

先づ中位數の長所と考へらるる點を擧ぐれば次の六つを數へる事が出來る。

(一)極端なる項に影響せられざる事。此の性質あるが爲めに賃銀統計、物價統計に中位數がよく用

6) Secrist: ibid p. 294.
Yule: ibid p. 116-117.

ひられ、又極端値の影響を好まざる他の統計にも將來に用ひらるるであらう。

(二)多くの計算を用ひずして簡單に求め得らるる事、完全なる bell-shape の對稱形に近き分布曲線に於ては各種の平均値は殆んど合致するものである。故に正確なるを要せざる場合には中位數は計算が簡單であるから便法として廣き利用範圍を持つ。

(三)性質が簡單明瞭で一般に理解し易き事。[7]

(四)完全なる報告の必要なき事。[8] 各項に關する調査資料が正確且完全ならざる場合に於ても猶中位數は正確に見出す事を得る。例へば賃銀調査に於て賃銀が平均より遙かに下位にある十萬人の勞働者が全然申告に表はれない樣な場合には、之より算術平均を求むる事は出來ないが、中位數は求め得らるるのである。

(五)計量し得ざる量に適用し得る事。[9] 例へば少年の智能の程度の平均を求むる場合に、その智能の程度を數量的に各人に就いて調査する事は甚だ困難であつて、從つて此くの如き場合に算術平均を求むる事は最難事とせねばならぬ。然るに少年を智能の程度に從つて配列する事は左程困難でないから、その平均として中位數を求むるならば、その配列の中央の少年の一人の智能の程度を知ればよいのであつて、斯る計量し得ざる量の平均としては中位數は甚だ便利である。之は Galton が最初に "Natural Inheritance" 又は "Hereditary Genius" に於て用ひたる方法であつて、此の研究の必要上より百分位數は生れたのである。

(六)他の平均方法では計量方法の異なるに從つて平均値に相違を來す事があるのに、中位數は如何

7) 小林新：ibid. p. 193.
8) Bowley: Elements of Statistics p. 103.
9) Bowley: ibid p. 103-104.

なる計量方法を用ひても常に同一である。[10]

次に中位數の短所と考へらるる點を列擧すれば次の如くである。

(一)一列の觀察の中央値は其の典型とは全然離れて居り、觀察された種々の對象物の何れにも接近して居ないかも知れぬと云ふ事、[11] 故に中位數は主として大部分の data が可成り接近して集まつて居る場合卽ち連續函數の場合に使用してこそ始めて有効なのである。

(二)代數的取扱の不可能なる事、抽象的に取扱ふ場合に不便なる事が多い。

(三)利用範圍が制限せられて居る事、中位數は既に述べたるが如くに data の性質によつては利用し得ざる場合が多く、その利用に當つては考慮を要する。特に經濟統計學の立場より見れば中位數の利用範圍は相當の制限を受けて居るのである。

(四)蓋然誤差の大なる事、計算の過程は甚だ原始的であるから蓋然誤差は相當に大であつて正確なる平均値を要するものには餘り適しない。蓋然誤差が大である爲めに、可成り正確なものを得るには他の平均値に比して data が多い事を必要條件とするのである。

第五　中位數の誤差

中位數には夫と諸個別値との差の絶對値の總和は最小なりと云ふ性質がある。[1]

此の性質は Fechner が最初に證明したものであつて、彼は中位數は夫と諸個別値との差の絶對値の總和が最小なりと云ふ性質を有し、算術平均は夫と諸個別値との差の平方の總和が最小なり

10) 小林新：ibid p. 194.
11) Bowley：ibid p. 104.
1) Walsh：ibid p. 46.

と云ふ性質を有し、又他の或る平均（彼は此の平均の型は發見し得なかつたが夫を得る方法は盡して居る）は夫と諸個別値との差の三乘の總和が最小なりと云ふ性質を有する事を指摘した。此くして Fechner は算術平均を中位數に優るものとし、從つて最小の三乘値を有する平均は算術平均に優り、最小の四乘値を有する平均は三乘値の平均より優るとした。此くして羃數の多き最小値を有するもの程、勝れたる平均なりと推論したのである。故に中位數は觀察が不正確なる場合には充分であるが、觀察がより正確なる場合には算術平均を要すとなしたのである。[2] Fechner は亦簡單なる例を掲げて之を說明してゐる。

任[3]意に選擇せる七個の値、零、二、四、六、七、八、八の數列に於いては、其の算術平均は五であり、中位數は六である。中位數に於いては其の偏歪の總和は一七であるが、中位數以外の値よりの偏歪の總和の方が大である。それは數列中の一項であつてもよいし、又數列中の二項間の平均であつてもよいのであつて、例へば四と六との算術平均五よりの偏歪の總和は一八であり、四と七との算術平均五・五よりの偏歪の總和は一七・五であり、又二よりの偏歪の總和は二三である。之に反して(算術平均)よりの偏歪の平方の總和は五八で最小である。例へば六(中位數)よりの偏歪の平方の總和は六五である。
(Über den Ausgangswert der kleinsten Abweichungssumme, S. 20)

既に述べたるが如くに中位數を用ふるが爲には多くの data を必要條件とする。既に緒言に於て述べし如く、中位數の計算の過程は原始的であつて且代數的取扱不可能なるが爲めに、中位數の蓋然誤差は自ら大ならざるを得ないのである。中位數の Probable error は算術平均の夫よりも大であつて、算術平均と同樣の確かさを持たしむるためには、算術平均に一〇〇の觀測をなしたとすれば中位數には二一七の觀測をしなければならぬ。[4] Edgeworth も此くの如く中位數の誤差は算術平均の夫よりも少し大なる事を認め、中位數に於て觀測の回數を增すならば算術平均と同樣

2) Walsh : ibid : p. 50.
3) 岡崎文規譯：統計的中數値論 p. 303.
4) Edgeworth : Some New methods of Measuring Variation in general Prdis. p.350.

の確からしさを持つ事が出來ると云つてゐる。[5]

Encke は中位數の近似的蓋然誤差の値として

$$\frac{0.786716}{\sqrt{m}}r$$

但し、r は中位數
m は觀測の回數

を出して居る。之は標準誤差の Probable error より約一・一二三倍だけ大である。[6] 小倉氏の書に掲げられて居る公式を示せば次の如くである。[7]

中位數の標準誤差

$$\sigma_{mi} = 1.253\frac{\sigma}{\sqrt{N}}$$

中位數の確率誤差

$$P.E_{mi} = 0.8454\frac{\sigma}{\sqrt{N}}$$

「之等の公式は或る假定の下に計算せられたのであつて近似的に正しいと見做し得るが決して嚴密なものではない。之等の公式に於て特に注意を要するのは之等の統計値の正確度が測定數Nの平方根に比例する事である。」

中位數を計算するに當つて度數分布表に集められたる data より比例法によつて求むる場合に階級内の分布は均等なりとの假定に立つたが、實際に於ては決して均等ではなく中數値に向つて集中する傾向がある。[8] 故に一つの階級内に於ても、その階級を折半すれば中數値に近き方の半分に他の遠き方の半分によりも多くの度數を含んで居るものである。故に中數値より下の階級に於

5) Edgeworth: The Doctrine of Indexunmber (Economic Journal, June 1918) p. 194.
6) Walker: ibid p. 83.
7) 小倉金之助: ibid p. 659.
8) Chaddock: ibid. p. 120.

ては上の階級に行くほど度數の密度が濃厚であるが、中數値を越へて上の階級に行けば、前と反對に上の階級に行くほど度數は稀薄になるのである。故に階級の廣さが倍になれば夫だけ階級內の分布が均等なりとの假定は事實に遠ざかり、從つて夫より求めたる中位數は不正確となり、半分の階級の廣さにしたる度數分布表より計算したる中位數とは正確には合致せぬ事となる。故に中位數を求むる場合には出來得る限り Class interval を小にすべきである。

又此の比例法に比すれば圖法に於ては更に確からしさは薄らいで來る。[9] 曲線は誤差を伴ひ易きものであり、又累積的度數分布曲線の傾斜の度弱き時には、中位數四分位數等を求むる水平線の交點は非常にズレ易く、誤差は大とならざるを得ない。尙前に述べたる所であるが、上下より描ける二本の累積度數分布曲線が中位數の直上で交叉せざる場合多きは明かに圖法の誤差大なるを物語るものである。

亦 Fechner は ungrouped series の場合に data の數が偶數なる時、中央の二項の算術平均を以て中位數となす計算の過程を把へ、此くの如き便法を用ふる事により計算の方法による誤差が生じて來ると述べて居る。[10]

之を要するに、中位數の誤差は他の平均に比して相當に大であつて、中位數を求むる場合に圖法は比例法に比して更に誤差が大である。

中位數はその外形は餘りに簡單に過ぎて、その微妙なる性質は粗野なる外形に蔽ひかくされ、その本質は甚だ把握し難い。從つて本論文は中位數の本質に關する將來の研究への準備として、その存在の理由を有してゐる。(五・五・一二)

9) Chaddock: ibid. p 118.
10) Walker: ibid p. 86.