Title	推定値の性質と許容限界
Sub Title	The character of estimates and the tolerance interval
Author	佐藤, 保
Publisher	慶應義塾経済学会
Publication year	1955
Jtitle	三田学会雑誌 (Keio journal of economics). Vol.48, No.3 (1955. 3) ,p.198(14)- 217(33)
JaLC DOI	10.14991/001.19550301-0014
Abstract	
Notes	論説
Genre	Journal Article
URL	https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=AN00234610-19550301-0014

三田學會雜誌　第四十八卷　第三號

推定値の性質と許容限界

佐藤　保

一

經濟諸量の統計的分析を考へる場合、ある安定的と考へられる期間について、考察の對象となつているものについて模型をつくり、現實の資料に對應させることにより、又豫測をも考慮することによつてその模型の良否を判斷し、改善すべき點があれば修正をほどこしてゆくという手段によつて漸次眞の値に近づいてゆくことができると考へる。勿論現實の動きは無數の要因からなつているから、それに全く一致した値が現れることはむしろ不思議といつてもよいはずであるが、本質的特徴をつかまへることができれば（それは少數の要因であることが想像されるならば）、動きの傾向を知ることが出來るはずである。そこで要因が少なくてよく現實にあてはまつていればそれにこした事はないが、通常はそうでないから多くの變數を導入することとなるが、なにを入れなにを捨てるかはすぐにきめることができない。勿論模型は經濟理論的に關係ありと認められるものによつて組みあげられているのであるが、資料による當嵌まりの良さということから別の結論がでることはしばしばあることである。その場合經濟理論的考察が違つているとしてその變數を取ってしまうか或は修正して時間的な差をつけて見るとかして當嵌りのよさがよくなれば、それによつて成程實はこうであつたのだとなつとくするか、或は資料に操作を加えて例へば期間を長くするとか短くするとか、ある點まで資料を切るとかしてみてそれによつて當嵌りがよくなればその範圍で理論が妥當するのだ、それ以上、以下では不規則であるから理論を當嵌めることがいけないのだとなつとくするかである。係數の符號の問題などもその一例である。

そして經濟的意味のつく式に變形すると統計的精度を落すことになる場合もあるがやはり經濟的に意味づけのできる式を採用することにならう。Klein の Simple Three-Equation System は經濟理論的な意味づけがなされていると共に大きさが手頃である所から特によく知られている。充分な分析といへばその次にある Large Structural Model の方がよいであらうが變數が多すぎるからすぐに取り上げるわけにゆかないので結果だけ眺める事になる。單純模型では消費、投資、賃銀を表す式を考へているのであるが、例へば消費を表す式で

(1)　$c=\alpha_0+\alpha_1\pi+\alpha_2\pi_{-1}+\alpha_3(w_1+w_2)$

と

(2)　$c=\alpha_0+\alpha_1\pi+\alpha_3(w_1+w_2)$

を比べて、（cは消費、πは利潤、w_1は個人契約による勞働所得、w_2は政府支拂の勞働所得）(1)の方が統計的にすぐれているが係數α_1がマイナスにでたのは經濟的意味として不合理であるとしている。ここでは相關係數の値がでていないので、一致の程度がわからないのであるが、卷末の附錄の數字からかなり高いものであることが直觀される。相關係數は式の良否をきめる際にはやはり有力な手掛りとならう。(1)の方がすぐれているのは所謂ネイマン比と呼ばれる

$$\frac{\delta^2}{s^2}=\frac{\sum(u_{xt}-u_{xt-1})^2}{\sum u_{xt}^2}\cdot\frac{N}{N-1}$$

（uは誤差、Nは期間を示す）が(1)では（1.56）(2)では（0.98）で(1)の方が自己相關が少なくなつているからである。しかし自己關係の檢定は未だ大標本の分布をそのまゝ適用しているのであるから極端な場合は別として近似的たることはまぬかれない。Tintner[註一]は統計的時系列分析に於て他に類例をみないくらい詳細であつてそこに特色があると考へられるが、（この場合時系列分析とは、例えば昭和六年から昭和十二年迄の何かの分析といつても勿論時系列分析と言うが、ここでの意味は時系列を[註三]、傾向、振動、不規則部分の三つから成るものと考え、如何にそれを分解するかが問題となる。）その多くの例の終りに、しかしこの例では大標本と考へることができないという但し書きがついている。しかしそれを無視するとして、例へば5%、有意水準で自己相關ありとすれば、はじめの假定についても考慮されなければならなくなり、いづれの式を採用するかについてはなほ愼重を要することとなろう。つまり經濟理論第一主義でゆくか、（即ち經濟的に關係ありと考えれば當嵌りや式の性質はあまり考えない）相關係數第一主義か（即ち最も當嵌りのよい變數を選び例えそれがあまり經濟的に關係のないものであつても採用し、最小自乘法で計算する）、方程式或は係數の統計的性質第一主義で（即ち式の計算上假定された性質や推定値の條件や性質に重きをおく、例えば先の例で α_1 がマイナスになつたといつても、係數の標準誤差が係數に比較して大きいから大體係數そのものがあまりあてにならないと考える等）ゆくか、勿論現實には三者が混合しており判斷はその時々で下されるのであるが、得られた資料が母集團からの標本と考へることと、標本分布とその性質標本特性値から母數を推定するという意識がつよくなるにつれて、式或は係數の統計的精度の比重が大きくなつて相關係數第一主義の比重が小さくなる。相關係數第一主義といつても自由度の問題とからんでくれば、多元回歸の計算の場合、そのまへに從屬變數と獨立變數の單相關の計算が入つてきて獨立變數の數をきめる上に一層困難な問題を生じるが今そのことは考へず、獨立變數が何であるかはきめられているとする。つまり多元回歸の單なる當嵌めは構造係數の推定に於ける偏りの犧牲に於て高い相關を得ているのだ[註四]ということになる。しかしこれも模型を如何に見るかによつて異つてくるから簡單にそうも言へないように思はれ、結局如何に現實をはんえいしているかは豫測によつて判斷せざるを得ない場合が多いであらう。

註一　L. Klein Economic Fluctuations in the United States 1921-1941. pp. 73-74.
註二　G. Tintner Econometrics pp. 258.
註三　N. G. Kendall the Advanced Theory of Statistics vo II, pp. 369
註四　N. A. Girshick and Trygve Haavelmo. Stitastical Analysis of the Demand for Food. Example of Simultaneous Estimations of Structural Equation Econometrica vol 15, 1947. pp. 110.

二

方程式或は推定値の統計的精度が問題になつてくると、つまり標本特性値から母數を推定する際に、それがよい推定値としての條件をそなへているかどうかが問題となる。普通良い推定値の性質として四つが考へられている。即ち、不偏性、一致性、有效性、充足性でこれは好ましい順序に書かれている。充足性が最も好ましい性質であるがこのような性質をもつている推定値は殆んどないことがわかつている。そこで第一、第二の性質について多くの場合問題とせられ更に第三の性質について考へる場合もある。特に大標本の場合はあまり問題にならないが小標本では偏りが大きくなるので問題となつてくる。母數を θ とし推定値を $\hat{\theta}$ と書けば不偏性とは $E(\hat{\theta})=\theta$ 一致性とは $P(\hat{\theta}\to\theta)\to 1,\ n\to\infty$

或は $\lim^{n\to\infty} P(|\hat{\theta}-\theta|>\varepsilon)=0$ を示し有效性とは標本數が無限大へ向つて大きくなるならばθの分布は正規分布へと近づきその平均はθ、分散は、平均θをもつて近似的に正規分布する何等か他の統計量の分散よりも小さいという性質である。なほ不偏推定値のうちで最小の分散のものを最良不偏推定値、我々が取扱う推定函數は殆んど線型であるが、その場合は最良線型不偏推定値と呼ぶことになる。又最尤推定値と呼ばれるものは、$x_1 \cdots x_n$ が確密度函數 $p(x/\theta)dx$ から取り出された標本觀察値の集合とする時もしもすべての觀察値が互に獨立であれば標本の同時分布（これを尤函數という）は

$P(x_1|\theta)\ P(x_2|\theta)\cdots\cdots P(x_n|\theta)\ dx_1, dx_2\cdots\cdots dx_n$　一般に　$L(x_1 x_2\cdots\cdots x_n|\theta)\ dx_1 dx_2\cdots\cdots dx_n$

で示される。尤度函數を最大にするようなθ

$$\frac{dL}{d\theta}=0$$

を解くことによつて得られる[註一]、推定値がこれらの性質を滿たしていれば滿たさない場合よりもすぐれているわけであるからそのような推定式を採用しようというわけである。しかし良い推定値としての條件の中でもいろ〳〵な對立がある。不偏推定値は期待値が母數に一致する性質であるから勿論望ましいが必ずしも一致推定値であるとは限らない。即ち標本數をいくら增しても母數に近づかない場合がある。これでは實際上はよい推定値とはいへないわけでたゞ不偏性があるといつただけでは安心できないわけであるが、事實上不偏推定値の多くは一致性をもつている。一致推定値は極限に於ては必ず不偏であるからこの性質をもつ推定値はもたない推定値に比べればすぐれているといへる。但し標本數が少なければ一致推定値といつても偏りがある場合のある事は當然である。一致性を滿たしている不偏推定値の中で分散の最小な最良不偏推定値は最も望ましい性質として多く使はれる。推定値の期待値が母數に一致することと、母數のまわりにできるだけ密集していることが望ましいのであるが、最良不偏推定値は不偏推定値の中で分散最小というだけであつて偏りのある推定値の中にはもつと分散の小さいものがありうるわけである。その差が大きい場合はいづれの推定値をとるべきかが問題となるが差が小さければ分散性の方を捨てて不偏性の方がとられる。最も推定値は一致性、有效性を滿たしているから大標本の場合は充分良い推定値、（この場合も近似的に正規分布するものの中で分散最小という意味であるが、その差が小さければ正規性の方がとられるだろう）と言へるが一般に不偏性は滿たしておらず小標本の場合偏りのある推定値はそれを不偏推定値へ變換しなければならない。このため推定値の分散は大きくなるかもしれないがなほその方が有效であらう。（例えば母分散の推定の場合）尤度函數を用ひて推定値の分散を求める場合も大標本という前提があるので一般的には使へないが、囘歸係數の分散を求める際には非常に有效となる。これらの事を考慮して Haavelmo の限界消費性向の分析について考へて見る[註二]。cを消費、yを所得とすれば普通消費は

$$c_t=\alpha y_t+\beta+u_t$$

の形で表はされる。そして最小自乘法によつて推定値$\hat{\alpha}$、$\hat{\beta}$を決定する。この式をもし正しいと考へれば$\hat{\alpha}$ $\hat{\beta}$はα βの最尤推定値又最良不偏推定値であり充分望ましい性質を持つている。しかし今この式は消費を表すのに未だ不滿足なものであり、又投資乘數を推定する問題も考慮することにより投資支出z_tを考へることにする。そしてzは所得（可處分所得）から消費支出を引いたものとして定義される。そこで

$$y_t=c_t+z_t$$

$$c_t=\alpha x_t+\beta+u_t$$

投資はすべて主動的なものとする。若干の理論は投資は主動的な變數であると考へる例へば Schumpeter の革新の理論等、誘導投資の概念によつて展開される他の理論も現在の投資は少くとも一部主動的變數でありその主な決定者は人口の増加、新發明、戰爭、或は若干の他の經濟變數の過去の値、例へば利潤、賣上高、資本蓄積の如き外部的要因である。そして(a)確率變數 u_t はの各々の値に對して期待値 $\varepsilon(u_t)=0$、分散 $\varepsilon(u_t^2)=\sigma_u^2$ をもつ、(b)時系列 z_t （$t=1, 2\ldots\ldots$）は c_t と y_t に對して主動的 (autonomus) である。この條件は次ぎのいづれかを滿たすことによつて成立つ。(b.1) 系列 z_t は $\varepsilon(z_t u_t)=z_t\varepsilon(u_t)=0$ (b.2) 各 z_t は u_t から確率的に獨立な確率變數である。という假定を設けることにより、（これらの假定が滿たされるかどうかにも問題があるがここでは滿たされているとする）これから

$$c_t=\frac{\alpha}{1-\alpha}z_t+\frac{\beta}{1-\alpha}+\frac{u_t}{1-\alpha}$$

$$y_t=\frac{1}{1-\alpha}z_t+\frac{\beta}{1-\alpha}+\frac{u_t}{1-\alpha}$$

の式を導きは獨立變數であるから最小自乘法を使つてもよいことになる、積率の記號を

$$m_p=\frac{1}{N}\sum_{t=1}^{t=N}p_t$$

$$m_{pq}=\frac{1}{N}\sum_{t=1}^{t=N}(p_t-m_p)(q_t-m_q)$$

として書くと、最良不偏推定値を b.u.e で表せば

$$\frac{m_{cz}}{m_{zz}}=b.u.e \text{ of } \frac{\alpha}{1-\alpha}$$

$$\frac{m_{yz}}{m_{zz}}=b.u.e \text{ of } \frac{1}{1-\alpha}$$

$$\frac{m_{zz}m_c-m_{cz}m_z}{m_{zz}}=b.u.e \text{ of } \frac{\beta}{1-\alpha}$$

又これらの推定値はNが無限大へ向う時、眞のパラメータが任意の ε 以上推定値より離れる確率は0に近づく。α と β は $1/1-\alpha$ と $\beta/1-\alpha$ の連續函數である故先の式より導かれる

$$\alpha=\frac{m_{cz}}{m_{yz}}$$

$$\beta=\frac{m_{yz}m_c-m_{cz}m_y}{m_{yz}}$$

は α、β それぞれの一致推定値である。そして最小自乘法によつて得られる推定値を a、b で示せば

$$a=\frac{m_{cy}}{m_{yy}}$$

$$b=\frac{m_{yy}m_c-m_{cy}m_y}{m_{yy}}$$

となるがこれは一致推定値ではない。そして

$$a=\frac{\alpha m_{zz}+(1+\alpha)m_{zu}+m_{uu}}{m_{zz}+2m_{zu}+m_{uu}}$$

で　$N\to\infty$　$m_{zu}\to 0$　$m_{uu}\to\sigma_u^2$　$0<\alpha<1$　の時

$$\underset{N\to\infty}{p\lim}\, a=\frac{\alpha+\frac{\sigma_u^2}{m_{zz}}}{1+\frac{\sigma_u^2}{m_{zz}}}>\alpha$$

$$p\lim b=\frac{\beta-\frac{\sigma_u^2}{m_{zz}}\,\overline{m_z}}{1+\frac{\sigma_u^2}{m_{zz}}}<\beta\quad \overline{m_z}>-\beta$$

の如き偏りを生ずるというのである。これは單一方程式でなくて條件をつけたのだから當然とも言へるわけである。Haavelmo はアメリカの資料で一九二二年から四一年迄の二十年間にわたるもので、單一方程式による最小二乗法と同時推定方式によるものとでそれぞれ計算している。

	Y_t	C_t	Z_t
1922	433	394	39
23	483	423	60
24	479	437	42
25	486	434	52
26	494	447	47
27	498	447	51
28	511	466	45
29	534	474	60
30	478	439	39
31	440	399	41
32	372	360	22
33	381	364	17
34	419	392	27
35	449	416	33
36	511	463	48
37	520	469	51
38	477	444	33
39	517	491	46
40	548	494	54
41	629	529	100

Y_tは生計費指數によつて割つた一人當り可處分所得、C_tは同じく生計費指數で割つた一人當り消費支出、Z_tはY_tからC_tを引いたものである。單位は弗。これより計算の結果は

(1)　$C_t=0.672\ Y_t+113.1$

$\hat{\alpha}=0.672\quad \hat{\beta}=113.1$

(2)　$C_t=0.732\ Y_t+84.0$

$a=0.732\quad b=84.0$

となつている。αに對する95%信頼區間

$0.57\leqq\alpha\leqq0.73$

を見れば兩者の違ひはかなりあることがわかる。なほ $\hat{\alpha}=0.67$ は信頼區間の中點ではない。Haavelmo は單に式の結果だけをだして、それ以上は何もふれていないが、更に實際値と對應させてその誤差を計り誤差の平方和の計算では(1)式よりは $\sum u_t^2=1306$　(2)式より $\sum u_t'^2=1306$ となつておりいはば當然のことながら最小自乗法による方がかなり當嵌りがよくなつている。しかし推定値の偏りをも考へた場合(1)式の方がすぐれているとする考へ方もある。[註三] 更に相關係數を計算してみよう。相關係數r^2は從屬變數の變動（variation）のうち回歸によつて説明される部分を示すのであるから逆に1から回歸によつて説明されない部分を引いたものと思つてもよい。今 Mcc のN倍は 35886.80 と計算されるから

$$r^2=1-\frac{\sum u^2}{\sum(c-\bar{c})^2}$$

にあてはめrを計算すれば　(1)に對して $r=0.98$　(2)に對して $r=0.919$ を得る。共に非常に高いが相關係數がわずかに下ると誤差の平方和はかなり增加するものである。前に戻つて推定値の偏りのあるなしから(1)と(2)と比べて(1)をとることは前節で述べた相關主義を捨てることになるが、實際の當嵌りは悪いのであるから、これだけでは$\hat{\alpha}$ $\hat{\beta}$が一致推定値であるといつ方がa bに比べて現實の値、或は眞の値をよく表しているとはいへないであらう。

ても、それは現在たてた式に於て一致推定値であるというのであつて、更に變數を增やしたり、複雜な關係を考慮すれば今の$\alpha$$\beta$は一致推定値ではなくなつてしまうであらう。丁度先の最小自乘法による推定値が一致推定値でなくなつたように。そこで一致推定値であるかないかといつても考へられている式のみについていへることであり、二つの式を見ていづれが現實に近いものであるかは簡單にきめることはできない。複雜な式をつくつたのは經濟理論的にその方がすぐれていると考へてつくつたのであるが、眞にそうであるかどうかはその式を見ただけではわからない。今の場合單一方程式最小自乘法による推定値の方がより現實をはんえいしているかもしれない。（單一方程式なら最小自乘法による推定値が最良不偏推定値であることは前にのべた）結局一つの安定的な期間と考へられる期間であれば豫測によつてそれを確めることになろう。しかし經濟的安定期間或は同質的と考へられる期間は短いから自由度等の問題もあつて多くは期間一杯をとつてしまうから豫測については考へ得ないことになる。例へば現在の例で最初の十五年間でパラメーターの推定を行ひ、それで後の五年間の豫測をしてみる。得られた式は

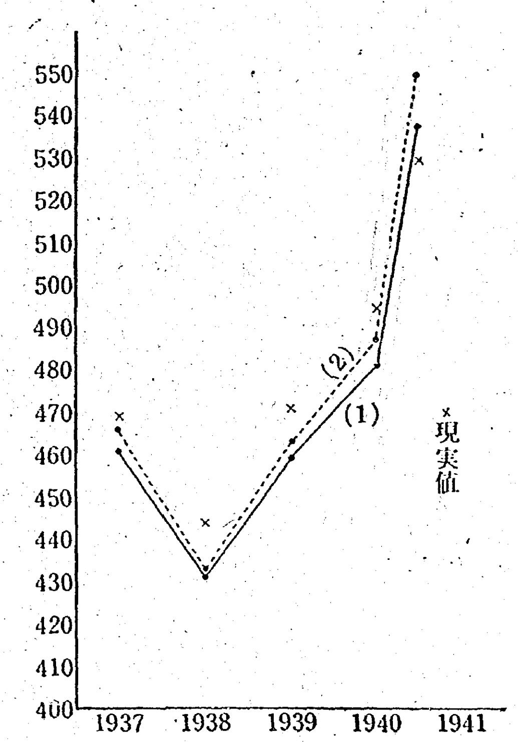
$$(1)'\quad C_t = 0.699\ Y_t + 97.58$$

$$(2)'\quad C_t = 0.764\ Y_t + 68.26$$

となつた。先の(1)(2)に比べてやゝ傾斜が大きくなつている。誤差の平方和は(1)では 603 (2)では 441 である。これより後の5年間の理論値と實際値とを比べてみると

(1)	理論値	實際値	u	u^2
1937	461	469	−8	64
38	431	444	−13	169
39	481	464	−12	144
40	481	494	−13	169
41	537	529	8	64
				610

(2)				
1937	466	469	−3	9
38	433	444	−11	121
39	463	471	−8	64
40	487	494	−7	49
41	539	529	20	400
				634



これを見れば五年間の誤差の平方和は(1)では 610 (2)では 643 で(1)の方が小さい。そこでこの結果から$\alpha$$\beta$の方が$a$ bよりよいということになるが、圖表からわかるように最初の4年間は最小自乘法によるものの方が壓倒的に當嵌り方がよいのである。5年目になつて大きくはずれているため總計に於て大きくなつている。しかしこの五年目の年は所得が大きく增加しているにもかゝはらず消費はわずかしか增えていない。即ち 0.4ぐらいの割合である。そこでこの年は異常な年であるともいへる。既に戰時體制に入つたのかもしれない。そこでこれを除外しようという考へも起るであろうし、最初の三年間も最小自乘法によるものより惡い推定であれば、方程式の組み方自體が惡いと考へこれを修正しようという氣になるであろう。即ち一九二二年から三六年迄の十五年間で考へれば $C_t=\alpha y_t+\beta$, $y_t=C_t+Z_t$ は單に $C_t=\alpha y_t+\beta$ よりも劣つている。だからこれを組み變へて豫測値のよいものにしなければならない。自由度の問題を考へればなかなか困難であろうがともかくやり直す氣になれば五年目迄まつてよく當嵌つた所は見ないことになる。しかし5年目によく當嵌つたことは同時推定式の方が經濟變動に耐え得たことになる。この變動が經濟體制の根本的變化であるかどうかの解釋にもよるが、その點からすれば(1)′式の方がすぐれているということに

なる。この間の様子はあと二、三年の資料があればより明瞭になると思はれるが現在の場合やむを得ない。

註一　L. Klein A Textbook of Econometrics pp. 52-53.

註二　T. Haavelmo, Methods of Measuring the Marginal Propensity to Consume. Studies in Econometric Method, chapter IV pp. 75-91.

註三　山田勇　經濟の計量　pp. 165.

三

豫測の問題に關聯して常に附隨しているのは許容限界の問題である。即ち回歸分析に於て

$$y=\alpha_0+\alpha_1 z$$

なる形を考へ、即ち尤度として

$$L=\overset{n}{\underset{i=1}{\pi}}\frac{1}{\sqrt{2\pi\sigma}}exp-\left(\frac{1}{2\sigma^2}\right)[y_i-(\alpha_0+\alpha_1 z_i)]$$

を考へることによりこれより、直接的にも求まるが、尤度函數をα_0、αについてそれぞれ二度微分しつくられた行列の逆行列の要素を求めるという方法によつて係數の分散及び共分散を求めることができる。その結果は

$$\sigma^2\alpha_0=\frac{\sigma^2\sum z^2}{n\sum(z-\bar{z})^2}$$

$$\sigma^2\alpha_1=\frac{\sigma^2}{\sum(z-\bar{z})^2}$$

$$\mathrm{cov}(\alpha_0\alpha_1)=\frac{\sigma^2\bar{z}}{\sum(z-\bar{z})^2}$$

となり、又母分散の最尤推定値

$$\sigma^2=\frac{1}{n}\sum(y_i-\hat{\alpha}_0-\hat{\alpha}_1 z_i)^2$$

を得る。$\dfrac{n\hat{\sigma}^2}{\sigma^2}$が自由度 $n-2$ のχ^2分布をすることから、[註一] 不偏推定値とするためには

$$M\frac{n\sigma^2}{\sigma^2}(\theta)=(1-2\theta)^{\frac{n-2}{2}}$$

〔$M(\theta)$ は積率母函數の記號を示す〕

$$M\hat{\sigma}^2(\theta)=\left(1-\frac{2\sigma^2}{n}\right)^{-\frac{n-2}{2}}$$

これより$\hat{\sigma}^2$の分布は $\alpha=\dfrac{n-2}{2}-1$ $\beta=\dfrac{2\sigma^2}{n}$ のガンマー分布であることがわかる。

$$M'\sigma^2(\theta)_{\theta=0}=\frac{n-2}{n}\sigma^2$$

$$E(\sigma^2)=\frac{n-2}{n}\sigma^2$$

不偏推定値とするためには

$$\frac{n}{n-2}E(\sigma^2)=\sigma^2,$$

即ち、

$$\frac{\sum(y-\alpha_0-\alpha_1-\alpha_1 x)^2}{n-2}=\hat{\sigma}_2^2$$

とすれば$\hat{\sigma}_1^2$は母分數の不偏推定値である。推定値の分數は原點のまわりの二次の積率から一次の積率の二乘を引くことによつて求まり積率は積率母函數より求められるから

$$\sigma^2_{\hat{\sigma}_2^2}=\frac{2(n-2)\sigma^4}{n^2}$$

$$M_{\hat{\sigma}_1^2}=\left(1-\frac{2\sigma^2}{n-1}\right)^{-\frac{n-1}{2}}$$

と考へることにより

$$\sigma^2_{\hat{\sigma}_1^2}=\frac{2\sigma^4}{n-2}$$

その差を考へれば

$$\frac{2\sigma^4}{n-2}-\frac{2(n-2)\sigma^4}{n^2}=\frac{4\sigma^4(2n-2)}{n^2(n-2)}>0 \quad n>2 \quad p>0$$

となる。一般にP變數では

$$\frac{2\sigma^4\cdot p(2n-p)}{n^2(n-p)}>0 \quad n>p \quad p>0$$

となつて常に母分散の最尤推定値は最良不偏推定値よりも小さな分散をもつことなるが、その差の小さいことを考へて不偏分散を使うことになる。$y=\alpha_0+\alpha_1 x$ に於てある任意の x_0 をきめた場合の（$y_0=\alpha_0+\alpha_1 x_0$）$y_0$を豫測することが許容限界をつくることになる。（$y_0$をとらへることが信頼限界の問題である）即ち例へば任意の x_0 をきめた時、標本を抽出する以前に於て任意の y がその中にあるような確率 95% の區間を作ることなどがその例である。大標本では回歸線上の値から ±1.96σ² をとればよいが小標本では x_0 の値以何で變つてくる。

$$u=y-\hat{\alpha}_0-\hat{\alpha}_1 x_0$$

とすれば

$$\sigma_u^2=\sigma_y^2+\sigma\hat{\alpha}_0^2+x_0^2\sigma\hat{\alpha}_1^2+2x_0\,\mathrm{cov}(\hat{\alpha}_0-\alpha_0)(\hat{\alpha}_1-\alpha_1)$$

先に述べた結果を入れれば

$$\sigma_u^2=\sigma^2+\frac{\frac{1}{n}\sigma^2\sum x^2}{\sum(x-\bar{x})^2}+\frac{x_0^2\sigma^2}{\sum(x-\bar{x})^2}+\frac{2x_0\sigma^2\bar{x}}{\sum(x-\bar{x})^2}=\sigma^2\left(1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{\sum(x-\bar{x})^2}\right)$$

このままでは末知の母分散σ^2が殘るから

$$t=\frac{u-E(u)/\sigma_u}{\sqrt{n\hat{\sigma}^2/(n-2)\sigma^2}}$$

より

なら

$$P(\hat{a}_0+\hat{a}_1 x_0-A<x<\hat{a}_0+\hat{a}_1 x_0+A)=0.95$$

$$A=t_{0.05}\,\hat{\sigma}\sqrt{\frac{n}{n-2}\left(\frac{n+1}{n}+\frac{(x_0-\bar{x})^2}{\sum(x_i-\bar{x})^2}\right)}$$

を得る。[註二] 不偏分散をはじめから使う場合は

$$A=t_{0.05}\,\hat{\sigma}_1\sqrt{\frac{n+1}{n}+\frac{(x_0-\bar{x})^2}{\sum(x_i-\bar{x})^2}}$$

となつて計算上はこの方が便利であろう。なを回歸線が原點を通る場合、卽ち母集團に於て回歸線が原點を通ることがわかつており且つ標本に於ても原點を通らなければならないような場合、（そのような例はあまりないが）回歸係數の分散の性質から、つまり原點については誤差0なのであり、この場合一般に回歸線は平均値を通らないから平均値はそのもつ意味をなくするわけで原點から遠ざかるにしたがつて許容區間は廣くなるという結果を生ずる。原點をはずすことができればそれにより推定の精度を増すことができる場合もあろう。このような考へに對して更に、例へば母集團の少くとも95％が入ることが95％の確からしさで言へるという區間をもつて許容區間とするという考へもある。後者の方が嚴密な考へであろうし近年では經濟的豫測にも後者の考へが使われている。[註三] その方法によれば兩限界

$$\hat{y}-k\hat{\sigma}_1 \qquad \hat{y}+k\hat{\sigma}_1$$

（$\hat{y}$はxをx_0ときめたときの回歸線上の値へσ_1は不偏分數の平方根、kはx_0の値と共に變り各Z_0に對するkの値をきめることが問題になる）によつて示されれば

$$k=\gamma\sqrt{\frac{n}{\chi_{\gamma}(n)}}$$

$$\gamma=K_p\left(1+\frac{1}{2N'}-\frac{2K_p^2-3}{24N'^2}\right)$$

$$N'=\frac{N\sum(x-\bar{x})^2}{\sum(x-\bar{x})^2+N(x_0-\bar{x})^2}$$

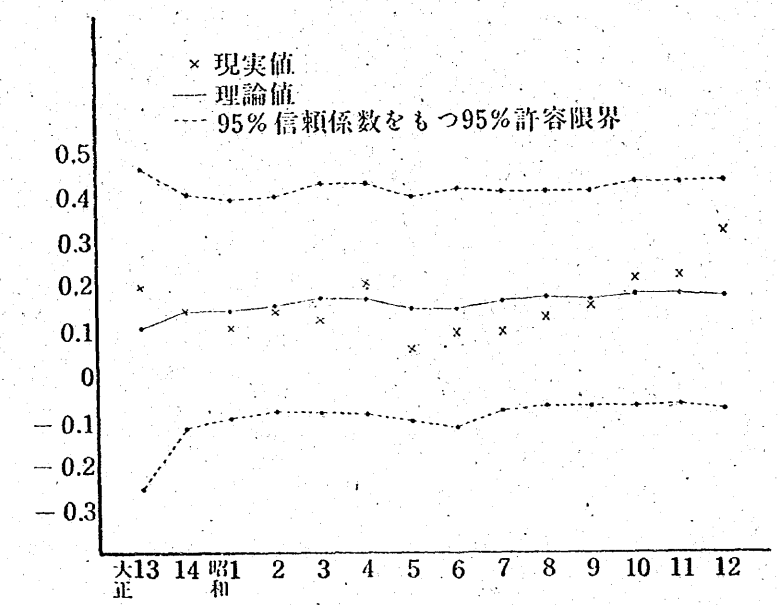
γは信賴度を示す、nは自由度、Nは標本數K_pは標準正規分布に於てそれ以上の値がでる確率が（1−P）である點、pは要求される確率。今95％の信賴度がほしければK_pは正規分布の表から1.96と得ることができN'は標本値より計算できるからγを計算できる。Nは N−2 であるからこれより $\chi^2_{0.94}(N-2)$ を表より求めれば各Z_0に對するKを求めることができる。[註四] 更に簡單には Techniques of Statistical Analysis for Scientific and Industrial Research and Production and Management Engineering by Statistical Research Group, 1947 の中にある許容限界の表を使うことによつて求められる。許容限界の問題は元來品質管理に利用されて、例へば95％許容限界を引いておき生產品の特性値がその中に入つていれば偶然的變動と見なし、許容限界の範圍外に出ればなにか變つた事があつたと考へて檢査を行う等である。卽ち區間の中にあれば良し、外に出れば惡いというわけであるが、回歸分析で例へばZ_0がある商品の卸賣價格、Yが需要と考へて、すべての點が許容限界の中にあるからといつて必ずしも良いとはいへない。つまりそれを使つてよい豫測ができるとはいへないのである。つまり當嵌りがわるければ誤差分散の値が大きくなれば許容區間の中に95％あるといつてもあまりそれが廣ければ實際的な意味をもたなくなるであろう。そのよ

うな場合回歸の線型性に對する檢定係數の同時檢定或は同時信頼區域も問題になつてくるであろう。[註五] 次の例は貯蓄率の一つの推定の場合であるが

$$S/Y = a_0 + a_1 \log \frac{Y}{N} + a^3 \frac{L-1}{Y} + \frac{Y-Y-1}{Y}$$

許容限界の表より 95% 信頼係數をもつ 95% 許容限界		理論値	實際値	
0.462	−0.254	0.104	0.195	大正13
0.405	−0.119	0.143	0.145	14
0.391	−0.397	0.147	0.106	昭和 1
0.399	−0.083	0.158	0.143	2
0.430	−0.082	0.174	0.124	3
0.430	−0.088	0.171	0.203	4
0.400	−0.100	0.150	0.059	5
0.418	−0.118	0.150	0.093	6
0.411	−0.077	0.167	0.099	7
0.413	−0.063	0.175	0.130	8
0.412	−0.020	5.171	0.155	9
0.432	−0.068	0.182	0.214	10
0.434	−0.066	0.184	0.222	11
0.435	−0.077	0.179	0.323	12

$$\frac{S}{Y} = -0.75 + 0.487 \log \frac{Y}{N} - 0.049 \frac{L-1}{Y} - 0.148 \frac{Y-Y-1}{Y}$$



Sは貯蓄、Yは可處分所得、Nは人口、Lは流動資産を示す。資料は大正十三年から昭和十二年迄の時系列、結果と圖表はS/Yに就てこの圖表からもわかる通り標本點はすべて許容限界内に入つているがその幅が廣いために豫測については殆ど何も言へないといつてもよいであろう。多元相關係數を求めてみれば0.237となる。つまりこのような低い相關であつても、許容限界の廣くなることからよほど大きな變動のない限り許容限界外に標本點が落ちることはないであろう。そこで許容限界云々が問題とされる前に先ず多元相關係數に關する檢とう、更にそれに伴う回歸係數の個々の或は同時的な研究がなされなければならないことがわかる。今の場合自由度10で0.237では相關0という假説はすてることができないのは容易にわかる。從つて當然回歸係數の値自體もあいまいなものとなり、それより許容限界を引いて豫測してみても無駄な事となろう。回歸分析に於ては相關係數の値は一番手掛りとなるものであり、それをもとにして次の手段がとられなければ意味のない結果を生むことになろう。

註一　ウィルクス小河原正巳譯數理統計學 pp. 252

註二　A. Mc. Mood, Introduction to the Theory of Statistics pp. 298-299.

註三　L. Klein A Textbook of Econometrics pp. 257.

註四　W. Allen Walls, Tolerance Interval for Linear Regression. Second Berkeley Symposium on Mathematical statistics and Probability pp. 43.

註五　David Durand, Joint Confidence Regions for Multiple Regression Coefficients, Journal of the American Statistical Association 1954.