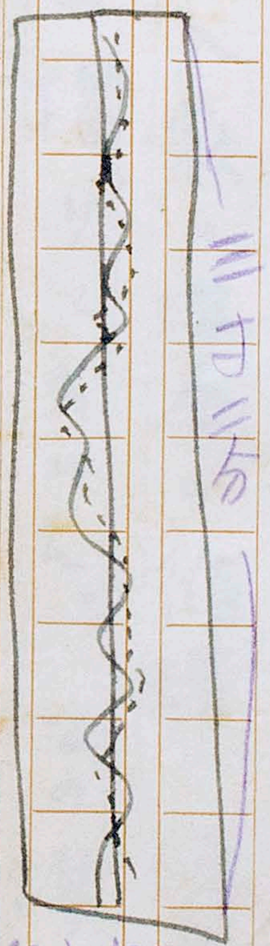
数の場合には、季節的変動が殆んど存在しないから、P=0 即ち S=100％として計算したのである。

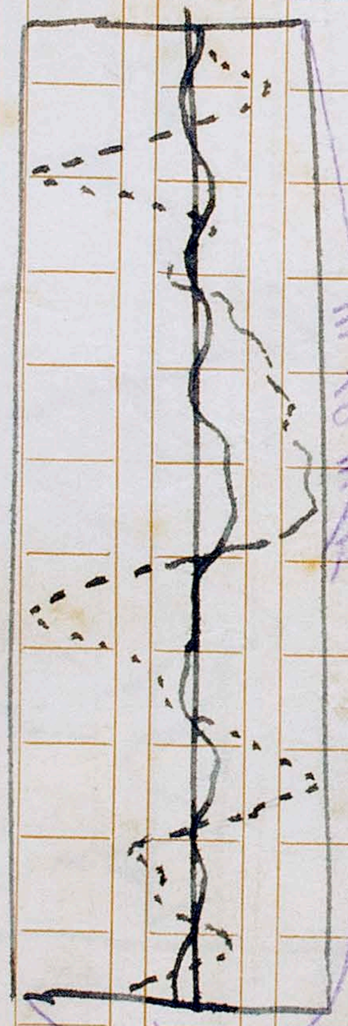
また第15図は、ブラッドストリートの物価指数と、米国の銑鉄生産額を、共に修正時系列として比較したものである。この場合には、変動の



第15図

実線は物価指数、点線は銑鉄収入

凸版



第16図

実線は物価指数、点線は銑鉄生産額

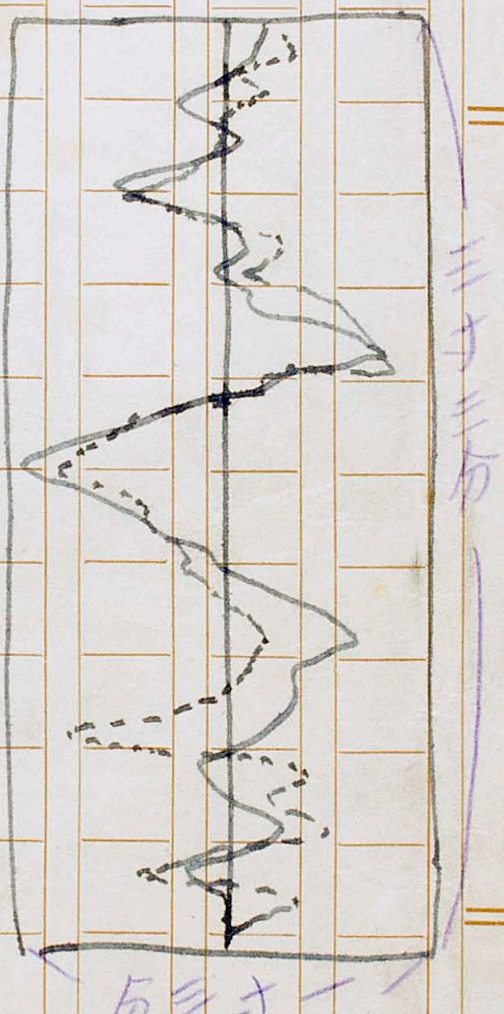
振幅が二つの曲線に於て大差あるため、曲線の形状を比較するには、不便な図となつてゐる。

二。それで、変動の振幅自身の比較が目的でなく、曲線の形状の比較を主要目的とするな

らば、縦線として標準測定値を採用する方が便利である。（第一章第四節）

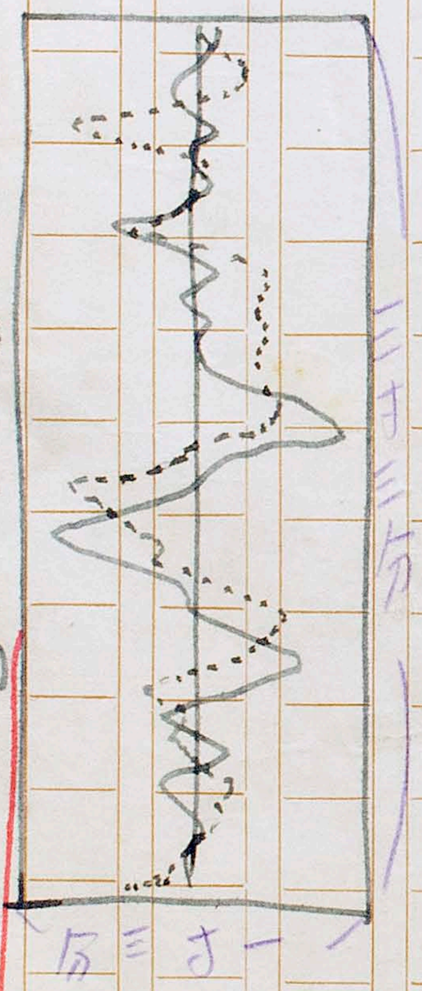
パーソンズは標準測定値で測つた修正時系列のことを、單に循環（サイクル）と呼んでゐる。

第17図は第15図を、第18図は第16図を、循環に直して書いたものである。（鉄道収入額の場合の循環の値は、一部分上掲の表の中へ掲げておいた。この場合には、修正時系列の(%)の標準偏差は5.98(%)である。



第17図

実線は物価指数、点線は鉄道収入



第18図

実線は物価指数、点線は銑鉄生産額

時系列の比較として、グラフの比較に優る方法は、今日まで未だ発見されてゐない。併し相関係数の方法も、グラフと併用して、必ずしも無用の業ではあるまいと思ふ。

注意。

時系列の比較には、なほ其の変化の速度及び加速度を比較して見るのも、参考になるであらう。—読者は力学について回想するが宜しい。

時 t_i, t_{i+1}, t_{i+2} に於ける時系列の項を y_i, y_{i+1}, y_{i+2} とせば、

t_i より t_{i+1} 迄の変化の速度 $= \dfrac{y_{i+1}-y_i}{t_{i+1}-t_i}$

であるが、時の間隔 $t_{i+1}-t_i$ は一定であるから、之を無視して宜しい。由て上の速度を v_i で表はせば、

$v_i = y_{i+1} - y_i$ となる。同様に $v_{i+1} = y_{i+2} - y_{i+1}$。

即ち変化の速度は、時系列の各項の差として現はれる。

同様に、加速度 α_i は速度の各項の差、即ち

$\alpha_i = v_{i+1} - v_i = (y_{i+2} - y_{i+1}) - (y_{i+1} - y_i)$

UTSUYAMA

$= y_{t+2} - 2y_{t+1} + y_t$

従て、某の時系列の各項の第二差として与へられる。

かやうに二つの時系列(y)、(y')に於て、各項の第一差(v)と(v')、ないし第二差(α)と(α')、――更に進んで第三差、等々――を、比較する人々もある。しかし此の比較は単なる参考に止まるのであつて、多くの場合に於て有意義ではあり得ない。何故なら、速度ないし加速度は、原の変化

一行アキ

参考書目

第一章末の邦文の諸書を見よ。

第四章 相関関係

第一節 相関係数

二つの時系列の比較に於て、それ等が(或はそれ等のグラフが)、如何程類似した形を有するかの程度を測ることを、その二つの時系列の類似関係或は相関関係を測ると称する。

吾々は前章の終に述べたやうに、先づ二つの時系列が、夫々その標準測定値で測られたものとして出発しよう。その時二つの時系列を

t_1,	t_2,	$\cdots$,	t_n
x_1,	x_2,	$\cdots$,	x_n
y_1,	y_2,	$\cdots$,	y_n

とする。

吾々は、このとき

$$r=\frac{1}{n}\sum x_k y_k=\frac{x_1y_1+x_2y_2+\cdots+x_ny_n}{n}\quad\cdots(1)$$

なる数値を計算して、之を二つの時系列の類似係数—或は世間並の用語に従って、相関係数と呼ぶことにする。なぜなら、この類似係数 r は、或る一種の意味に於て、二

つの時系列の数値関係を測る数値であるから。

私は簡単に、その理由を説明したい。

代数の公式 $(x_k - y_k)^2 = x_k^2 - 2x_k y_k + y_k^2$ を、それぞれ就りて1からnまで加へ合せ、その結果をnで割れば

$$\frac{1}{n}\sum(x_k - y_k)^2 = \frac{1}{n}\sum x_k^2 - \frac{2}{n}\sum x_k y_k + \frac{1}{n}\sum y_k^2$$

となる。しかるに時系列は標準測定値で測られてあるから、（x系列及びy系列の標準偏差を夫々 σ_x, σ_y で表はす）。

$$\frac{1}{n}\sum x_k^2 = \sigma_x^2 = 1, \quad \frac{1}{n}\sum y_k^2 = \sigma_y^2 = 1$$

であり、その上に吾々は $\frac{1}{n}\sum x_k y_k = r$ と置いたのであるから、上の式から

$$r = 1 - \frac{1}{2n}\cdot\sum(x_k - y_k)^2 \quad \cdots\cdots (2)$$

を得る。若し公式 $(x_k + y_k)^2 = x_k^2 + 2x_k y_k + y_k^2$ から出発して、同様の計算を行ふならば

$$r = -1 + \frac{1}{2n}\cdot\sum(x_k + y_k)^2 \quad \cdots\cdots (3)$$

に達する。

さて平方数の和は常に正数であるから、(2)から、rは+1より大きくないことが分り、(3)から、rは−1より小さくないことが知れる。即

ち類似係数（相関係数）rの値は -1 と $+1$ との間にある。

また(2)から、r が $+1$ なるためには、平方数の和 $\Sigma(x_i - z_i)^2$ が零なることと、従て平方数の各が皆夫々零となることを要する。即ち

$$x_1 = z_1,\ x_2 = z_2,\ \ldots\ldots,\ x_n = z_n$$

逆に、これ等の式が同時に成立すれば、(2)から r は $+1$ となる。由て

二組の時系列の標準測定値が順々に全く一致するためには、類似係数（相関係数）が $+1$ なることを要する。逆に、類似係数（相関係数）が $+1$ ならば、二組の時系列は順々に全く等しい標準測定値を有する。

これを時系列のグラフに就いて言へば、次の如くなる。

標準測定値で測られた二つの時系列のグラフ（例へば二つの循環）が、全く重なり合ふために、必要にして且つ十分なる條件は、類似係数（相関係数）が $+1$ なることである。

同様の推論を、式(3)に就いて行へば、次の結

果に到達する。

二組の時系列の標準測定値が、絶對値を等しくして符号を異にするため、即ち

$x_1=-y_1$, $x_2=-y_2$, …, $x_n=-y_n$

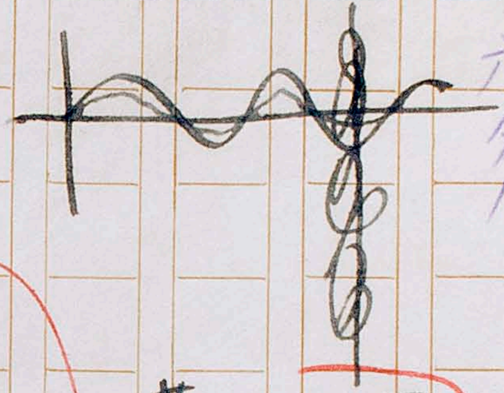
なるためには、數似係數（相関係數）が−1なることを要する。その逆もまた成立つ。

標準測定値で測られた二つの時系列のグラフが、時間の軸に對して對称なるために必要にして且つ十分なる條件は、數似係數（相関係數）が−1なることである。

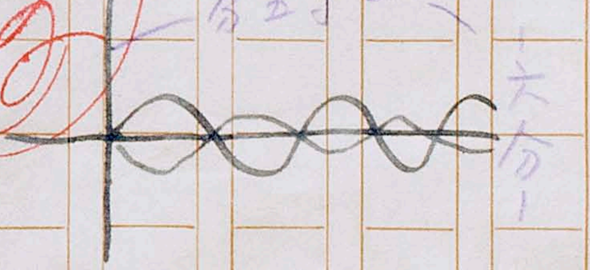
それ故にrが+1に近い値を取るとき、二つの循環はよく數似した—そのまゝの位置で（位置を変へずに）、全體に近い形—を有する。（第二図）rが−1に近い値を取るときは、よく數似した形であるが、一方を時間の軸に於いて折り返した~~形~~位置を取る。（第三図）。

この意味に於て、數似係數（相関係數）は時系列の數似性又は逆數似性を計るに用ひられる。rが正ならば順相関、負ならば逆相関をなすと呼ばれる。rが零な

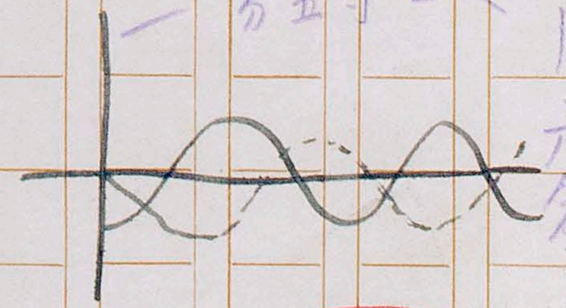
るときは、二つの時系列は無関係なりと定義される。（第21図）。



第19図



第20図



第21図

r を正数なりとせよ。この場合には、(2)から、

$$1-r=\frac{1}{2n}\cdot\Sigma(x_k-y_k)^2$$

であるから、二つの系列 $x_1,\dots,x_n$ と $y_1,y_2,\dots,y_n$ の開きの程度を平均的に測る値と見做し得られ、$\frac{1}{n}\Sigma(x_k-y_k)^2$ の大（小）は、r の小（大）に伴ってゐる。換言すれば、開きの程度が平均的に小なる場合——即ち二つの系列の類似性が高い場合——には、r の値は大きくなる。

次に、r を負数なりとせよ。(3)から

$$1-(-r)=\frac{1}{2n}\cdot\Sigma(x_k+y_k)^2,$$

若し x_k と y_k とが全く逆——即ち絶対値を等しくして符号を異にする——ならば、$x_k+y_k=0$ である。

由て二つの系列の逆數似性が高い程、平均的(けれは高)に、$\frac{1}{n}M\left(\frac{x}{\sigma_x}+\frac{y}{\sigma_y}\right)^2$は小くなる。従てrの絶對値は大きくなる。

又rは即ちrの正負、大小によって、時系列の數似性(或は逆數似性)を測る所以である。

説明の便宜上、簡單な例を取らう。次表に於ては、物價の循環(サイクル)をx、雇傭數の循環をyとして、その間の相關係數が求められてゐる。

圖表 相關圖と對

年月 t_k	物價 x_k	雇傭數 y_k	$x_k y_k$
1920年			
1月	0.75	0.59	0.443
2月	0.94	0.85	0.799
3月	0.91	0.51	0.461
4月	0.88	1.10	0.968
5月	0.89	0.68	0.605
6月	0.57	0.51	0.291
7月	0.37	0.34	0.126
8月	0.18	0.42	0.076
9月	−0.14	−0.09	0.013
10月	−0.53	−0.34	0.180
11月	−0.98	−0.59	0.578
12月	−1.74	−1.27	2.210
1921年			
1月	−2.10	−2.71	5.691
計	0	0	12.443

$$r=\frac{\sum x_k y_k}{n}=\frac{12.443}{13}=+0.96$$

同様の計算によって、第三章第六節に示した例、即ち銑鉄生産額と鉄道收入額の二つの循環の比較に於ては、$r=+0.88$を得る。

最後に、時系列$X_1, X_2, \cdots, X_n$或は$Y_1, Y_2, \cdots, Y_n$が、標準測定値で測られてゐない場合に就いて

考へよう。

この場合には、二つの系列の算術平均をそれぞれ M_x, M_y とし、標準偏差をそれぞれ σ_x, σ_y とすれば、

$$M_x=\frac{1}{n}\Sigma X_k,\quad \cancel{\xi_k = X_k - M_x},\quad \sigma_x=\sqrt{\frac{1}{n}\Sigma(X_k-M_x)^2},$$

$$M_y=\frac{1}{n}\Sigma Y_k,\quad \sigma_y=\sqrt{\frac{1}{n}\Sigma(Y_k-M_y)^2}.$$

そこで吾々は、二系列の類似係数（相関係数）を

$$\boxed{\gamma_{x,y}=\frac{\Sigma(X_k-M_x)(Y_k-M_y)}{n\cdot\sigma_x\cdot\sigma_y}}\quad\cdots\cdots(4)$$

ゴ

によつて定義しよう。

かやうに定義してから、系列を標準測定値 $x_1, x_2, \cdots\cdots, x_n$ および $y_1, y_2, \cdots\cdots, y_n$ に直せば、

$$x_1=\frac{X_1-M_x}{\sigma_x},\ \cdots\cdots,\ x_n=\frac{X_n-M_x}{\sigma_x},$$

$$y_1=\frac{Y_1-M_y}{\sigma_y},\ \cdots\cdots,\ y_n=\frac{Y_n-M_y}{\sigma_y}$$

となるから、(4)は

$$\gamma_{x,y}=\frac{1}{n}\Sigma x_k y_k$$

となる。しかるに、(1)によつて、その値は標準測定値によつて測られた系列の類似係数、即ち $\gamma_{x,y}$ に外ならないのである。由て

$r_{x,y} = r_{x'y'}$.

言へ換へれば、二つの時系列の間の類似係数（相関係数）の値は、時系列を標準形に直しても、不変に止まる。

第二節　時の遅れ。

一つの時系列の他の一つの時系列に對する時の遅れを測るには、先づ一方のグラフを他方のグラフの上に重ね、二つの曲線の形状が最もよく類似する位置を、視察によって判断する。

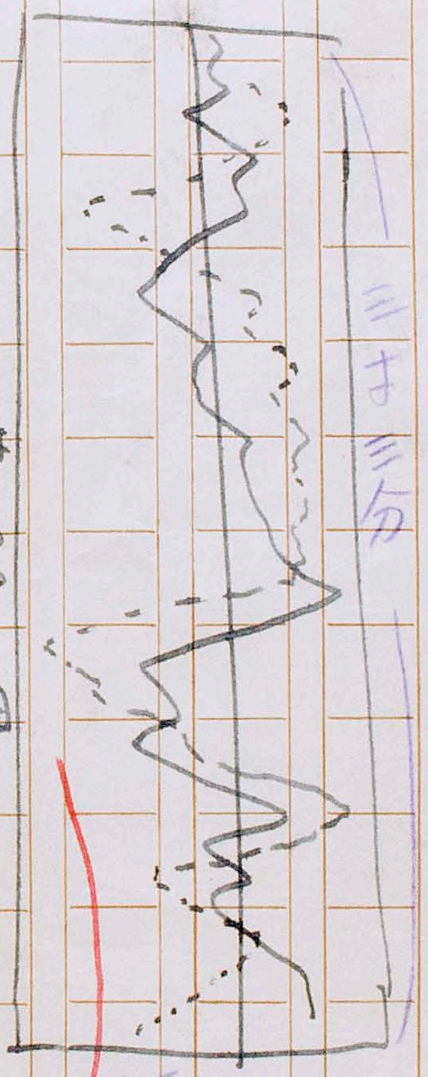
次に視察によって、類似性の大きい位置が大凡見出された場合には、その各位置に就いて、二つのグラフの間の類似係数（相関係数）を求め、その係数の値の最大なる位置を以て、時の遅れを測ることとする。

例へば米国に於ける銑鐵生産額の循環と金利（Interest rate on 60-90 day commercial paper）の循環とを比較するに、次の結果を得る。ここに四個月の遅れとあるは、「一九〇三年五月、六月、七月、……の金利子を、夫々一九〇三年一月、二月、三月、……の銑鐵生産額

No.

に對應するものと見做して、γを計算して見た

との意味であり、他の遅れについても同様である。



第22図　實線は金利、点線は鉄鋼生産額

時の遅れ(月)	γ
0	0.11
……	……
4	0.50
5	0.52
6	0.57
7	0.58
8	0.57
9	0.57
10	0.55

この場合には、γの最大値から判断して、時の遅れは七個月である。

(或は七個目と八個目との間にある)と見做すのが、至当であろう。

相関係数の使用に関する注意

これまで吾々は類似係数(相関係数)の使用に就いて語って来た。併し乍ら、γは、結局の所、單なる一種の平均値に過ぎないのである。それ故に二つの時系列の間の特殊な関係は、斯様な平均値たる唯一つの係数に依頼し、その値によってただ類似の程度を認るよりも、グラフの上で比較する方が遙かに正当である。時系列の

全面的に

間の或る特種なる、そして精密なる性質が、係数γによって表はし得ない場合でも、グラフの上では明瞭に示されることが屡々あるのである。

（なほ篇末の第六章第五節を見よ）

いらん

第三節 相関係数（類似係数）の他の解釈

これまで吾々は二つの時系列をグラフを、その儘の形で比較する場合の類似の程度を測る値として、rの意味を定めて来たが、（或はその

これより他の解釈を考へよう。

説明の便利のため、簡単な例、即ち、さきに第一節で取扱った物價の循環(x)と雇傭数の循環(y)との比較について再論しよう。今x、yを一つの直角坐標として考へて見れば、吾々は平面上にn個の点（この例では十三個の点）、即ち

$(x_1=0.75,\ y_1=0.54),\ (x_2=0.94,\ y_2=0.85),\ \ldots\ldots,\ (x_{13}=-2.10,\ y_{13}=-2.71)$

を得る。之を相関図と称する。（第23圖）

相関図に於ては、時もず全然逐出されてゐるから、それは時系列の性質を失ってゐる。しかし吾々は形式的に、この図形によって、相関係数の他の意味を解釈し得るのである。これまでは時系列(x)、(y)をtの函数と見做し、

$x_1=x(t_1),\quad x_2=x(t_2),\ \ldots;\quad y_1=y(t_1),\quad y_2=y(t_2),\ \ldots$

の意味で考へて来たのであったが、今後はただ

$$\left\{\begin{array}{l} x=x_1,\ x_2,\ \ldots\ldots \\ y=y_1,\ y_2,\ \ldots\ldots \end{array}\right.$$

なるやうに對應する二つの変量そのものとして、取扱はんとするのである。

今一つの相関図に於て、xの方の値が与へられたものと見做さう。そしてxの或る一つの値に對して、yが平均的に如何なる値を取るかを考へよう。この問題を解く最も簡單な方法の一つは、先づx、yの間の関係を近似的に直線的であると假定し、最小自乗法によって

$$y = a + bx$$

の係数a、bを定めることである。この問題は既に第二章第一節に於て解かれたが、その結果によれば、a、bは

$$na + b\sum x_k = \sum y_k, \quad a\sum x_k + b\sum x_k^2 = \sum x_k y_k$$

によって與へられる。しかるにx_k、y_kは標準化値によって與へられてゐるから

$$\sum x_k = 0, \quad \sum y_k = 0, \quad \frac{1}{n}\sum x_k^2 = 1, \quad \frac{1}{n}\sum x_k y_k = r.$$

由て

$$a = 0, \quad b = r$$

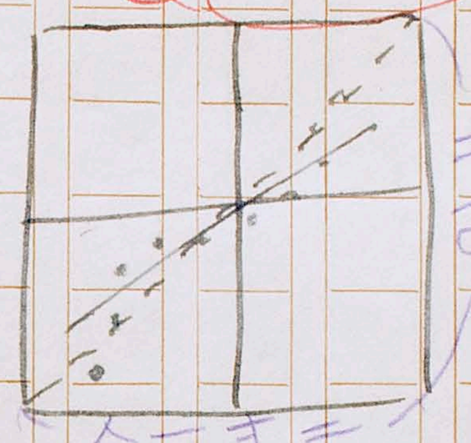
となる。従て

$$\boxed{y = rx}$$

を得る。故に次の結果を得る。

(x_k)、(y_k)が共に標準化されてゐるとき、x_kの値が1だけ増加するとき、平均的にy_kの増加する値は、rに等しい。

しかるにrは-1と$+1$との間にある値であるから、rが$+1$に近づけば近づくほど、x_kの変化によって起るy_kの変化の量が(平均的に)x_kの変化の量に近接して来る。またrが-1に近づけば近づくほどx_kの変化によって起るy_kの変化の量は、(平均的には)x_kの変化の量の符号を反對にした値に



第23図

この図前頁の最初に入れ

接近して来る。これ即ち相関係数によって、~~関係~~ (x_k)、(y_k)の間の変化の類似(の程度)——普通の用語で言へば ~~通例~~ (x_k)、(y_k)の間の相関関係の程度——を測る第二の理由である。

第23図の実線は、上の方法で求めた直線で、$r=+0.96$はその直線の勾配を示してゐる。かやうに定めた直線を、xのyに於ける回帰直線と呼ぶ。それは坐標の原点を過ぎり、そのx軸となす角の正切はrに等しい。

若し$r=+1$であったならば、直線は図中点線で示した直線(對角線)となったであらうし、また若し$r=-1$であったならば、直線は今一つの對角線となったであらう。

さて斯様に回帰直線を定めたとき、この直線に對して、与へられた相関図の平面上にあるn個の点 $(x_1,y_1),(x_2,y_2),\dots,(x_n,y_n)$は、如何なる程度に撒布されてゐるか？ その撒布の程度を測るには、如何にすべきか？

これに答へるには、一直線上に於ける諸点の（その平均値の周りの）撒布度を測るに、標準偏差を用ひることを、回想するがよい。それで吾々の問題に於ては、回帰直線について、（第3図の記号による）

$$\sigma_{y,x} = \sqrt{\frac{1}{n}\sum \overline{P_kL_k}^2}$$

$$= \sqrt{\frac{1}{n}\sum (y_k - r x_k)^2}$$

を以て、撒布度を測ることとしよう。そして之をxに於けるyの回帰直線からの撒布度と呼ばう。

さて

$$\sigma_{y,x}^2 = \frac{1}{n}\sum (y_k^2 - 2r x_k y_k + r^2 x_k^2)$$

$$= \frac{1}{n}\sum y_k^2 - 2r\cdot\frac{1}{n}\sum x_k y_k + r^2\cdot\frac{1}{n}\sum x_k^2$$

$$= 1 - 2r^2 + r^2 = 1 - r^2.$$

故に

$$\boxed{r^2 = 1 - \sigma_{y,x}^2.}$$

この結果から見れば、rが$+1$又は-1なるときに、そして其の場合に限って、相関図中の總ての点$(x_1,y_1),\dots,(x_n,y_n)$が、回帰直線上に来るのである。

これまで吾々は、xを~~自変数~~の方の値を与へるものと見做して来た。しかし若しyを~~自変数~~の値を与へるものと見做したならば、如何であらうか。この場合に、上と同様の計算によって、yに於けるxの回帰直線の式として

$$x = r y$$

を得る。従て、y_tの値が1だけ増加するとき、平均的に[redacted]x_tの変化する値はrに等しい。

一行アケ

次に若し二つの時系列(X_t)、(Y_t)が、標準測定値で与へられていない場合には、先づ之をそれぞれ[redacted]して考へる。rはこの変更によって不変である（第一章第一節の終りを見よ）。

~~X を自変数と見做せば、~~

$y = r x$ は

$$\frac{Y-M_y}{\sigma_y} = r\,\frac{X-M_x}{\sigma_x}$$

となる。由てXに於けるYの回帰直線の方程式として[redacted]

$$\boxed{Y-M_y = r\,\frac{\sigma_y}{\sigma_x}(X-M_x)}$$

を得る。故にX_tの値がX全体の平均値M_xよりも1だけ大なるときは、Y_tの値は（平均的には）Y全体の平均値M_yよりも、$r\frac{\sigma_y}{\sigma_x}$だけ大きい。

この数値$r\frac{\sigma_y}{\sigma_x}$を$b_{yx}$と書き、之を$X$に於ける$Y$の回帰係数と呼ぶ。

同様に、~~Yを自変数と見做せば~~、Yに於けるX