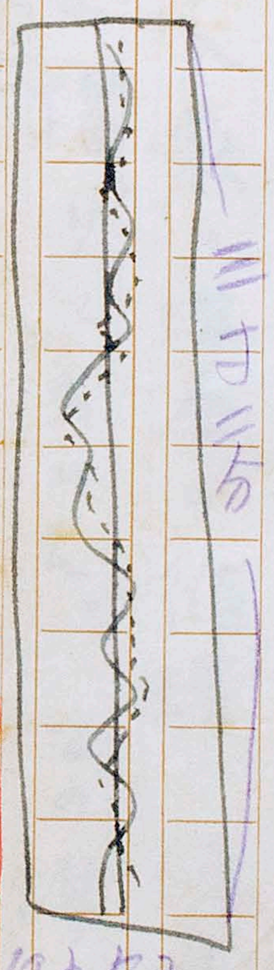
数の場合には、季節的変動が殆んど存在しないから、p=0 即ち S=100% として計算したのである。

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、点線は鉄道収入

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

6

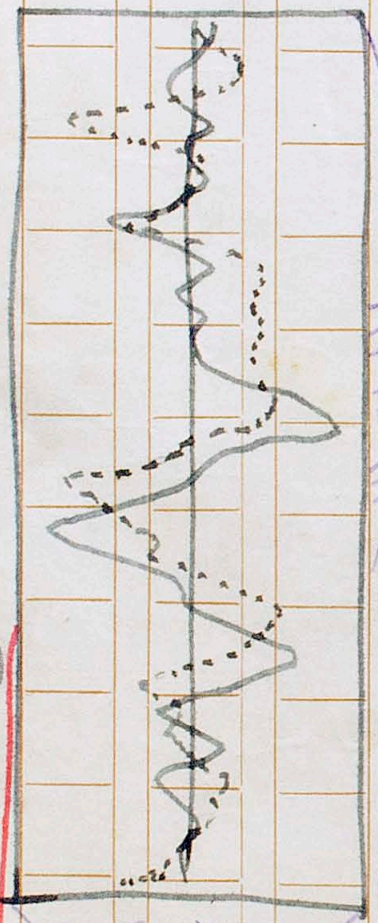
らば、縦線として標準測定値を採用する方が便利である。（第一章第四節）

パース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。この場合は、修正時系列の（%）標準偏差は 5.78（%）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、点線は鉄鋼生産額

6

6

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ない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、

$t_i$ から $t_{i+1}$ までの変化の速度 $= \dfrac{y_{i+1}-y_i}{t_{i+1}-t_i}$

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得られる。

同様に、加速度 $\alpha_i$ は速度の各項の差、即ち

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

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

従て、是の時系列の各項の第二差として與へられる。

かやうに二つの時系列(y)、(y')に於て、各項の第一差(υ)と(υ')、ない第二差(α)と(α')、――更に進んで第三差、等々――を、比較する人々もある。しかし此の比較は單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、速度や加速度は、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~~相~~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

| | | | | |
|---|---|---|---|---|
| $t_1$, | $t_2$, | …… | $t_n$ |
| $x_1$, | $x_2$, | …… | $x_n$ |
| $y_1$, | $y_2$, | …… | $y_n$ |

とする。

吾々は、このとき

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

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 $(x_k-y_k)^2 = x_k^2 - 2x_ky_k + y_k^2$ を、$k$ に就りて $1$ から $n$ まで加へ合せ、その結果を $n$ で割れば

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

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 $y$ 系列の標準偏差をそれぞれ $\sigma_x$, $\sigma_y$ 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 $\frac{1}{n}\sum x_ky_k = r$ 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 $(x_k+y_k)^2 = x_k^2 + 2x_ky_k + y_k^2$ から出発して、同様の計算を行ふならば

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

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 $\Sigma(x_k - y_k)^2$ が零なることと、従て平方数の
各が皆共に零となることを要する。即ち
$x_1 = y_1,\ x_2 = y_2,\ \ldots\ldots,\ x_n = y_n.$
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、絶対値を等しくして符号を異にするため、即ち

$x_1=-y_1,\ x_2=-y_2,\ \cdots,\ x_n=-y_n$

なるためには、類似係数（相関係数）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、時間の軸に対して対称なる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。

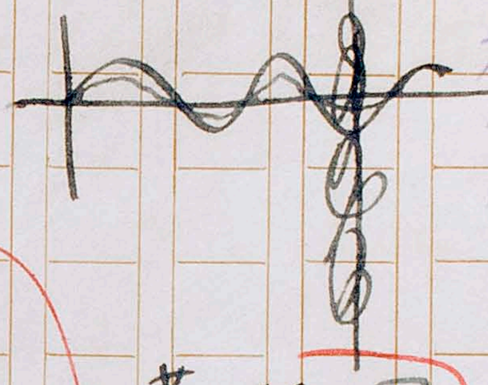
それ故に $r$ が $+1$ に近い値を取るとき、二つの循環はよく類似した——そのまゝの位置で（位置を変へずに）、全等に近い形——を有する。（第一図）

$r$ が $-1$ に近い値を取るときは、よく類似した形であるが、一方を時間の軸に沿いて折り返した位置を取る。（第二図）

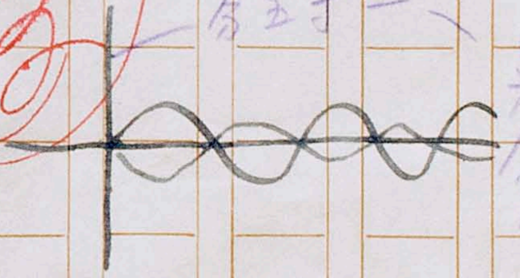
この意味に於て、類似係数（相関係数）は時系列の類似性又は逆類似性を計るに用ひられる。

$r$ 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

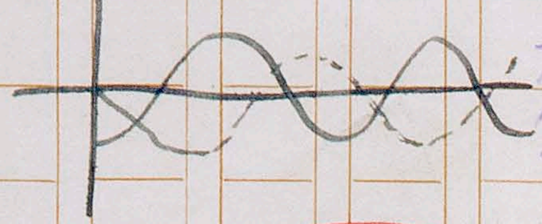
るときは、二つの時系列は無関係なりと定義される。(第21図)。



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 $x_1, x_2, \dots, x_n$ 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

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

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類似性の高ければ高い程、平均的に、小M$(x_k+y_k)^2$は小さくなる。従てrの絶対値は大きくなる。

これ即ちrの正負、大小によって、時系列の類似性（又は逆類似性）を測る所以である。

説明の便宜上、簡単な例を取らう。次表に於ては、物価の循環(サイクル)をx、雇傭数の循環をyとして、その間の相関係数が求められてゐる。

~~図[illegible]例 相関図と[illegible]~~

| 年月 $t_k$ | 物価 $x_k$ | 雇傭数 $y_k$ | $x_k y_k$ |
|---|---|---|---|
| 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収入額の二つの循環の比較に於ては、$r=+0.98$を得る。

最後に、時系列$X_1, X_2, \ldots, X_n$又は$Y_1, Y_2, \ldots, Y_n$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 $\sigma_x$, $\sigma_y$ とすれば、

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

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

そこで吾々は、二系列の類似係数（相関係数）を

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

によって定義しよう。

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

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

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

となるから、(4)は

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

となる。しかるに、(4)によって、その値は標準測定値によって測られた系列の類似係数、即ち $r_{x,y}$ 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。

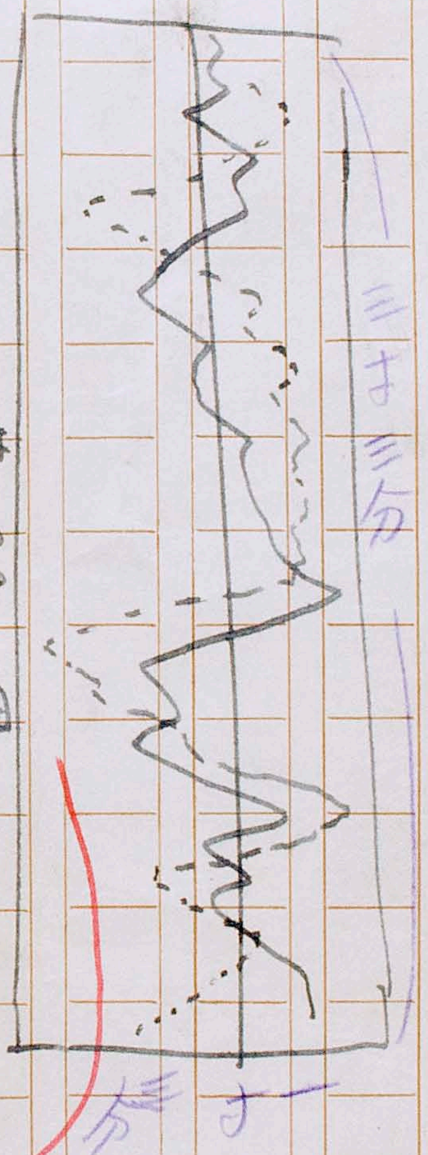
## 第二節　時の遅れ。

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、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後には、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にする。

例へば米国に於ける銑鉄生産~~高~~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。　Z、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~~銑鉄~~銑鉄生産額

に對應するものと見做して、r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、點線は銑鐵生産額

| 時の遅れ(月) | r |
|---|---|
| 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、rの最大値から判断して、時の遅れは七個月である。

(或は七個月と八個月との間にある)と見做すのが、至当であらう。

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。併しながら、rは、結局の所、単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知るよりも、グラフの上で全面的に比較する方が遥かに正当である。時系列の

間の或る特種なる、そして精密なる性質が、係数X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。)

# 第三節　相関係數（類似係數）の他の解釋

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、これより他の解釋を考へよう。

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$(x)$と雇傭數の循環$(y)$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。之を相関圖と称する。（第23圖）

相関圖に於ては、時とず全然逐出されてゐるから、それは時系列の性質を失つてゐる。しかし吾々は形式的に、この圖形によつて、相関係數の他の意味を解釋し得るのである。これまでは時系列$(x)$、$(y)$を時の函数と見做し、

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

の意味に考へて来たのであったが、今後はたゞ

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

なるやうな對應する二つの變量そのものとして、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単な方法の一つは先づ$x$、$y$の間の關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

$$y = a + bx$$

の係數$a$、$b$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$x_k$、$y_k$は標準測定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

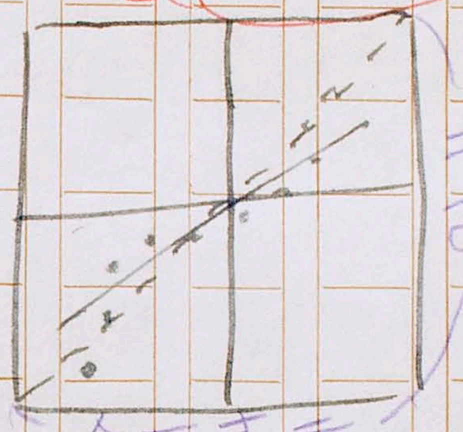
$$\boxed{y = rx}$$

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だけ増加するときに、平均的に$y_k$の増加する値は、$r$に等しい。

この図前頁の最初に入れ

挿



第23図 6

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

6

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(対角線)となったであらうし、また若し $r=-1$ であったならば、直線は今一つの対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 $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ 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、如何にすべきであらう？

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）

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

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

を以て、撒布度を測ることにしよう。之を$x$に於ける$y$の回帰直線からの撒布度と呼ばう。

そして

さて

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

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

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

故に

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

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 $(x_1, y_1), \ldots, (x_n, y_n)$ が、回帰直線上に来るのである。

これまで吾々は、xの方の値を与へるものと見做して来た。しかし若しy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、yに於けるxの回帰の場合には、上と同様の計算によって、[redacted]直線の式として

$x = r y$

を得る。従て、yが1だけ増加するとき、平均的には[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ゐない場合には、先づ之をそれに直[redacted]して考へる。Yはこの変更によって不変である（本章第一節の終りを見よ）Xを自変数と見做せば、$y = r_x x$ は

$$\frac{Y-M_Y}{\sigma_Y} = r \cdot \frac{X-M_X}{\sigma_X}$$

となる。[redacted]由て[redacted] Xに於けるYの回帰直線の方程式として

$$\boxed{Y - M_Y = r\frac{\sigma_Y}{\sigma_X}(X - M_X)}$$

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 $a_{YX}$ 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