第 三 學 年 特 別 講 義

統 計 法

理學博士 小倉金之助述

昭 和 十 一 年 十 月

東 京 物 理 學 校

資　料

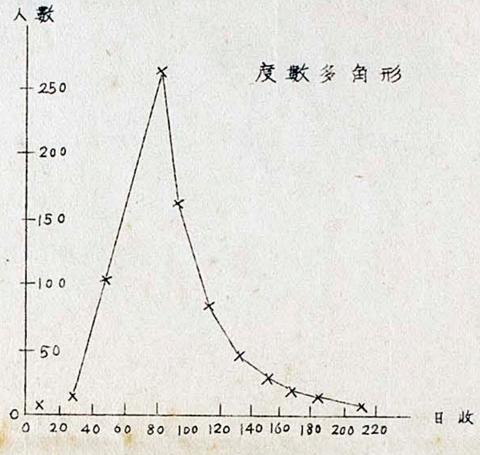
日本帝國統計年鑑

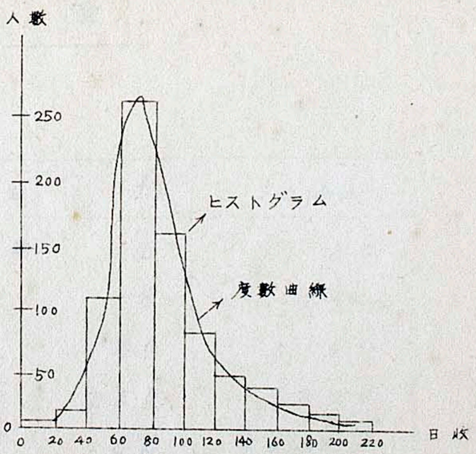
矢野恒太,　　日本國勢圖會

第一章　度數分布

昭和3年8月神戸市役所調査
神戸市マッチ軸木女工日收

日收（錢）	中央値（錢）	人數	累積人數
0 —— 20	10	6	6
20 —— 40	30	8	14
40 —— 60	50	102	116
60 —— 80	70	260	376
80 —— 100	90	162	538
100 —— 120	110	73	611
120 —— 140	130	36	647
140 —— 160	150	30	677
160 —— 180	170	18	695
180 —— 200	190	8	703
200 —— 220	210	1	704
		計704	





X	f
X_1	f_1
X_2	f_2
$\cdots$	$\cdots$
X_n	f_n
	$\Sigma f_k = N$

注意　　級の間隔（一定）を問題に就て適當に選ぶこと。
整正なる理想的分布曲線の存在の假定。

度數分布の型

對　稱　　［確率曲線，誤差の法則，ガウスの公式］　　$y = Ae^{-k^2x^2}$.

非對稱　　$f = \dfrac{N}{\sigma\sqrt{2\pi}} e^{-\frac{(X-M)^2}{2\sigma^2}}$　　英人の身長

$N = 8585,\ M = 67.46$ inch

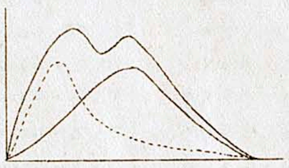
$\sigma = 2.57$

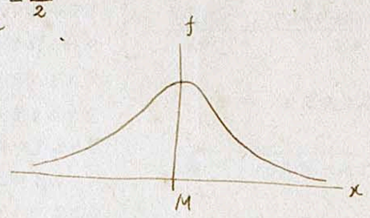
丁字形

U字形

$f = \dfrac{N}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$.

複雑な形





度數曲線の解析的表示

ピーアソンの方法

ブルンス，シャルリエの方法（直交函數）

第二章　平　均　値

總變量の代表値

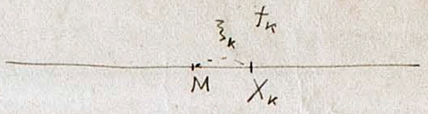
I　算術平均（平等）

$$M = \frac{\sum f_k X_k}{N} = \frac{\sum f_k X_k}{\sum f_k}$$

X_k	f_k	$f_k X_k$
10 銭	6	60
30	8	240
50	120	5100
70	260	18200
90	162	14580
110	73	8030
130	36	4680
150	30	4500
170	18	3060
190	8	1520
210	1	210
合計	704	60180

$$M = \frac{60180}{704} = 85.483 \text{ 銭}$$

偏差　$x_k - M = \xi_k$



$$\sigma = \sqrt{\frac{1}{N} \cdot \Sigma\left[f_k \cdot (x_k - M)^2\right]} = \sqrt{\frac{1}{N} \cdot \Sigma f_k \xi_k^2}$$

x_k	ξ_k	ξ_k^2	f_k	$f_k \xi_k^2$
10	−75·483	5697·683289	6	34186·099734
30	−55·483	3078·363289	8	24626·906312
50	−35·483	1259·043289	102	128422·415478
70	−15·483	239·723289	260	62328·055140
90	+4·517	20·403289	162	3305·332818
110	+24·517	601·083289	73	43879·080097
130	+44·517	1981·763289	36	71343·478404
150	+64·517	4162·443289	30	124873·298670
170	+84·517	7143·123289	18	128576·219202
190	+104·517	10923·803289	8	87390·426312
210	+124·517	15504·483289	1	15504·483289
			704	724435·795456

M = 85.483 才

$$\frac{\Sigma f_k \xi_k^2}{N} = \frac{724435.795456}{704}$$

$= 1029.028119$

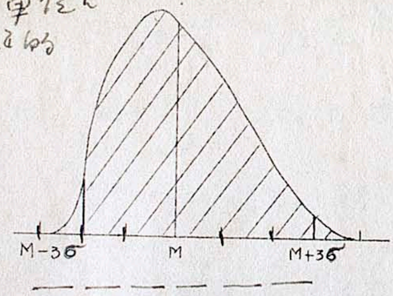
$\sigma = \sqrt{1029.028119}$

$= 32.0785$ 才

注　意　　$M - 3\sigma = 85.48 - 96.24 = -10.76$

$M + 3\sigma = 85.48 + 96.24 = +181.72$

σを新しい單位にすることが合理的である.



教育測定

故に殆んど總ての變量が　$M - 3\sigma$ と $M + 3\sigma$ との間に夾まれる.

對稱分布曲線の場合には, 上の二數の間に全變量の 99.73% が含まれる.

$$J = \frac{1}{\sqrt{2\pi}} \int_0^t e^{-\frac{t^2}{2}} dt$$

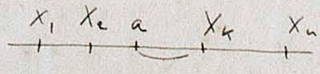
t	J
0.5	0.19146
1.0	0.34134
1.5	0.43319
2.0	0.47725
2.5	0.49379
3.0	0.49865
3.5	0.49977
4.0	0.49997

method of least squares

最小自乘法との關係

一般に a からの偏差を取り, a からの開き（分散）の程度を

$$D = \sqrt{\frac{1}{N} \cdot \Sigma\left[f_k (x_k - a)^2\right]}$$



で測る. a の位置を變ずるときは, D も變化する. そして D の最小は a が算術平均 M の場合に起る.

算術平均 M とは, 偏差の自乘の和　$\Sigma f_k (x_k - a)^2$ を最小にする a の値である, そして其の開き D の最小値は標準偏差 σ である.

$S = \Sigma f_k (X_k - a)^2$

$\frac{\partial S}{\partial a} = -2 \cdot \Sigma f_k (X_k - a)$

$= 0.$

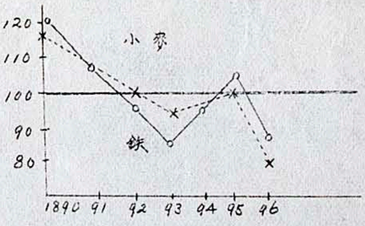
$\Sigma f_k X_k - a \cdot \Sigma f_k = 0$

$a = \frac{\Sigma f_k X_k}{\Sigma f_k} (= M)$

~~第四章~~ 指　數　index number

統計値の比較を容易にするための指數

年次	價格		價格指數（1890 — 1896基準）	
	鋼鉄（噸）	小麥（ブッセル）	鋼鉄	小麥
1890	30 \$	1.05 \$	120	117
91	27	0.96	108	107
92	24	0.94	96	104
93	22	0.83	88	92
94	24	0.88	96	98
95	26	0.92	104	102
96	22	0.72	88	80
平均	25	0.90	100	100



$X_1 \quad X_2 \cdots\cdots X_n$

指數 $I_1 = \frac{100X_1}{M}$，$I_2 = \frac{100X_2}{M}$，$\cdots\cdots I_n = \frac{100X_n}{M}$

綜合指數

生活費指數　Cost of living index

數年間の研究により，或る年の指數が

食費	住居費	被服費	燈火燃料費	雜費
108	102	110	96	104
X_1	X_2	X_3	X_4	X_5

これ等は生活費全部の内，夫々

40	16	14	6	24 %
m_1	m_2	m_3	m_4	m_5

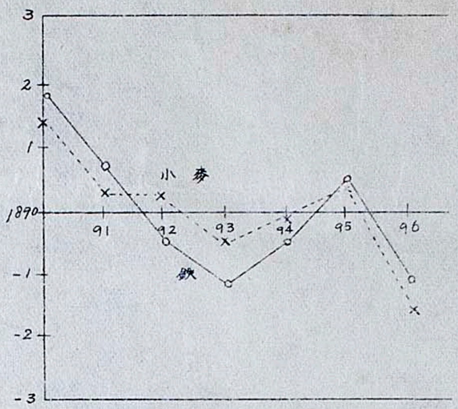
とせば，生活費指數　$\frac{\sum m_K X_K}{\sum m_K} = \frac{108 \times 40 + 102 \times 16 + \cdots\cdots}{100} = 106$

（秤量平均）

標準測定法 standard measure

$$x_1 = \frac{X_1 - M}{\sigma} = \frac{\xi_1}{\sigma}, \qquad x_2 = \frac{X_2 - M}{\sigma} = \frac{\xi_2}{\sigma}$$

鋼 鐵		小 麥	
$M = 25.$ $\sigma = 2.67$		$M' = 0.9$ $\sigma' = 0.097$	
ξ	$x\left(=\frac{\xi}{\sigma}\right)$	ξ'	x'
+5	+1.87	+0.15	+1.55
+2	+0.75	+0.06	+0.62
−1	−0.37	+0.04	+0.41
−3	−1.12	−0.07	−0.72
−1	−0.37	−0.02	−0.21
+1	+0.37	+0.02	+0.21
−3	−1.12	−0.18	−1.85



標準測定値で表はされた變量 $x_1, x_2, \cdots\cdots, x_n$ の算術的平均 M' は 0 に等しい.

$$M' = \frac{1}{n}\sum x_k = \frac{1}{n}\sum \frac{X_k - M}{\sigma} = \frac{1}{n\sigma}\left(\sum X_k - nM\right) = 0$$

$x_1, x_2, \cdots\cdots, x_n$ の標準偏差 σ' は 1 に等しい.

$$\sigma'^2 = \frac{1}{n}\sum (x_k - M')^2 = \frac{1}{n}\sum x_k^2 = \frac{1}{n}\sum \left(\frac{X_k - M}{\sigma}\right)^2$$

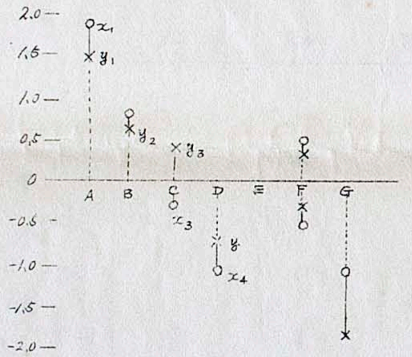
$$= \frac{1}{n\sigma^2}\cdot\sum (X_k - M)^2 = 1$$

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

第五章　相關關係 correlation

生徒	入學成績				卒業成績				xy
	X	ξ	ξ^2	x	Y	η	η^2	y	
A	90点	+5	25	+1.87	85点	+15	225	+1.55	+2.85
B	87	+2	4	+0.75	76	+6	36	+0.62	+0.46
C	84	−1	1	−0.37	74	+4	16	+0.41	−0.15
D	82	−3	9	−1.12	63	−7	49	−0.72	+0.79
E	84	−1	1	−0.37	68	−2	4	−0.21	+0.08
F	86	+1	1	+0.37	72	+2	4	+0.21	+0.08
G	82	−3	9	−1.12	52	−18	324	−1.85	+2.09
平均	$M(X)$ $=85$	0	$\sigma(X)^2$ $=7.1425$ $\sigma(X)=2.67$	0	$M(Y)$ $=70$	0	$\sigma(Y)^2$ $=95$ $\sigma(Y)=9.74$	0	

$x=\dfrac{\xi}{\sigma(X)}$,　　$y=\dfrac{\eta}{\sigma(Y)}$,　　$\sum x_k y_k=+6.25$



$$\frac{1}{n}\sum x_k y_k=\frac{6.25}{7}=+0.89$$

對應する二つの値の開きの平方の和の平均

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

が大なれば　關係が粗で　小なれば關係が密である。

（基本的考へ）

$$(x_k-y_k)^2=x_k^2-2x_ky_k+y_k^2$$

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

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

$$=2-\frac{2}{n}\sum x_ky_k$$

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

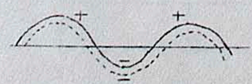
由て　　$r=\frac{1}{n}\sum x_ky_k$　　　r が小なれば関係が粗，〃 大 〃 密

とおけば　　$r=1-\frac{1}{2n}\sum(x_k-y_k)^2$　　ゆへ　$r\leqq 1$.

若し　$(x_k+y_k)^2=x_k^2+2x_ky_k+y_k^2$ から出發すれば　$r=-1+\frac{1}{2n}\sum(x_k+y_k)^2$.　$-1\leqq r$

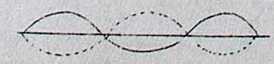
$$-1\leqq r\leqq +1$$

$r=+1$ なるための必要にして且つ十分なる條件　〔標準測定値で表はしたときに限る〕

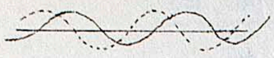


$x_1=y_1$,　$x_2=y_2$, ………… $x_n=y_n$.　相重る．

$r=-1$ なるための條件



$x_1=-y_1$, ………………… $x_n=-y_n$.　對稱．



$\gamma > 0$ 順相関

$\quad < 0$ 逆相関 $\quad \gamma = \pm 1$ 完全なる相関

$\quad = 0$ 無関係

γ を相関係數と呼び，その値の大小によつて相関の程度を測定する．

入學と卒業成績の相関 $\gamma = \frac{1}{n}\sum x_k y_k = \frac{6.25}{7} = +0.89$

英國 救貧法を布いて居る．３８地方に於ける 農業勞働者の平均一週の賃銀 Ｘと，

救貧法で救助を受けた人々の パーセンテージ Ｙとの相関 $\gamma = -0.66$

$$\gamma = \frac{1}{n}\sum \frac{\xi_k}{\sigma(X)} \frac{\eta_k}{\sigma(Y)} = \frac{\sum (X_k - M(X))(Y_k - M(Y))}{n\sigma(X)\sigma(Y)}$$

相関表 correlation table

入學成績

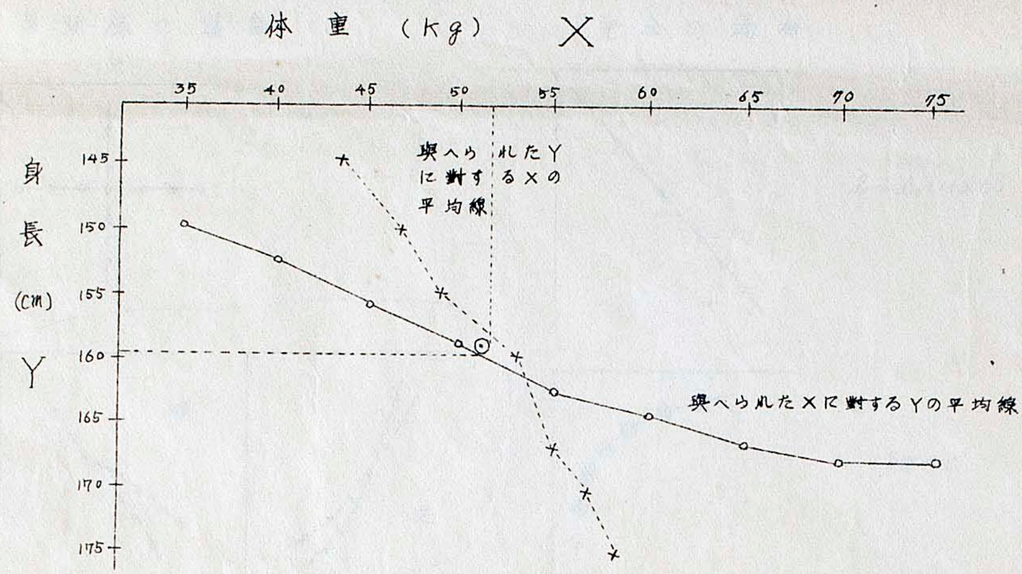
卒業成績	82	84	86	87	90
52	1	0	0	0	0
63	1	0	0	0	0
68	0	1	0	0	0
72	0	0	1	0	0
74	0	1	0	0	0
76	0	0	0	1	0
85	0	0	0	0	1

身長と体重の相関表 （昭和４年東京市隣接五郡壮丁）

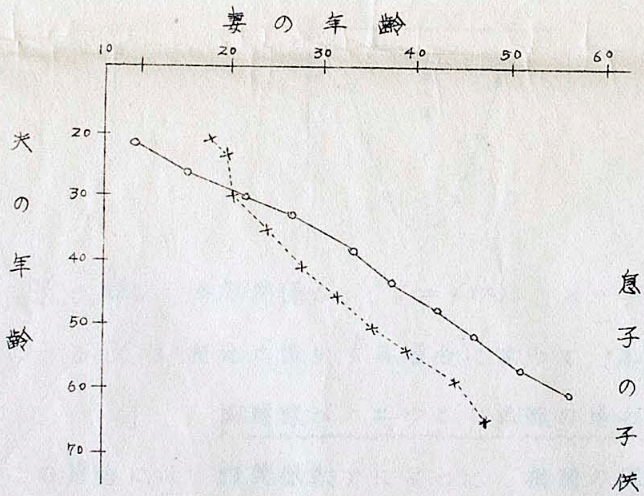
身長＼体重	35kg	40	45	50	55	60	65	70	75	合計	平均 Kg
145cm		3	5	2						10	44.5
150	1	9	29	19	5					63	46.4
155		10	53	86	37	6	1			193	49.5
160		3	36	107	103	30	5	1		285	52.5
165		1	11	48	75	42	11	1	1	190	55.0
170			1	10	22	19	8	2	1	63	57.6
175				2	4	4	2	1		13	58.5
合計	1	26	135	274	246	101	27	5	2	817	52.2kg
平均 cm	150.0	153.1	155.8	159.0	161.7	164.3	165.9	168.0	167.5	160.0cm	

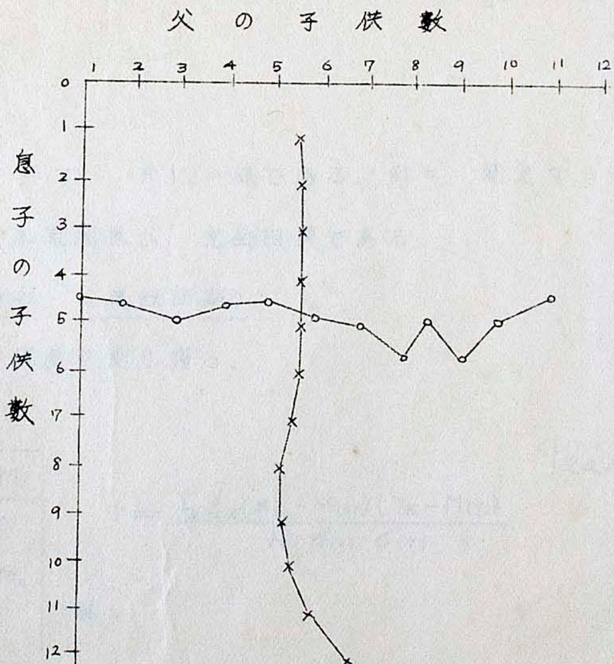
相関圖　（回歸線）　lines of regression

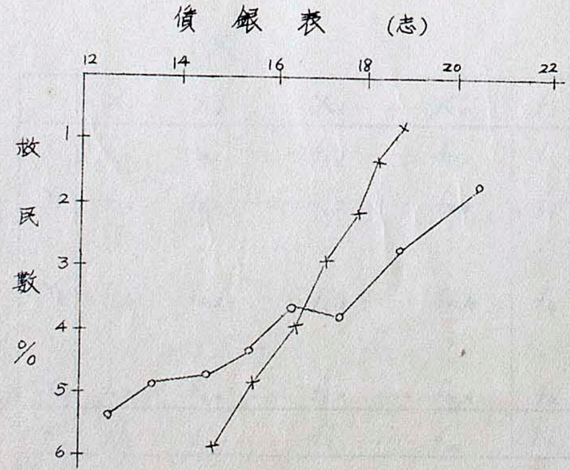
Galton



昭和四年内地の結婚

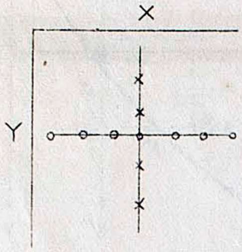




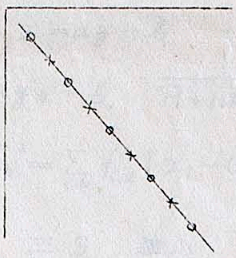


相関の程度

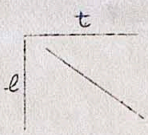
無関係の場合

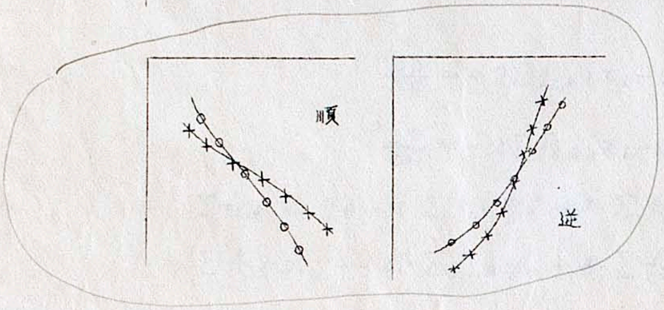


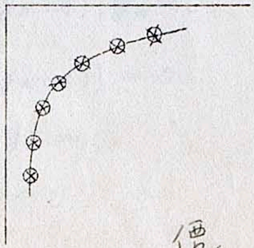
完全の場合



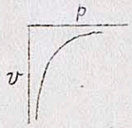
$l = l_0(1+\alpha t)$



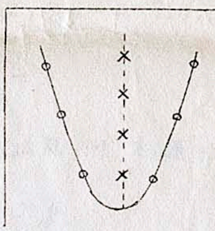




$pv = c$



完全相関と函数的関係との相違.〔一~~値~~価函数〕



X, Y の間の　函数関係が, $Y = f(X)$, $X = \varphi(Y)$　共に一値である. 即ち　絶えず増加する 又は 絶えず減少する場合に限つて, 函数関係が　完全相関である.

I　回帰線が　二つとも直線の場合　(直線回歸)

この場合には　相関係數 r によつて　相関の程度を測り得る.

X

Y

	X_1	X_2 ……	X_i ……	X_m	f	M
Y_1	$f_{1,1}$	$f_{2,1}$ ……	$f_{i,1}$ ……	$f_{m,1}$	f_1	M_1
Y_2	$f_{1,2}$	$f_{2,2}$ ……	$f_{i,2}$ ……	$f_{m,2}$	f_2	M_2
⋮	⋮	……	……	⋮	⋮	⋮
Y_k	$f_{1,k}$	$f_{2,k}$ ……	$f_{i,k}$ ……	$f_{m,k}$	f_k	M_k
⋮	⋮	……	……	⋮	⋮	⋮
Y_n	$f_{1,n}$	$f_{2,n}$ ……	$f_{i,n}$ ……	$f_{m,n}$	f_n	M_n
f'	f'_1	f'_2 ……	f'_i ……	f'_m	N	$M_{(X)}$
M'	M'_1	M'_2 ……	M'_i ……	M'_m	$M_{(Y)}$	

Pearson

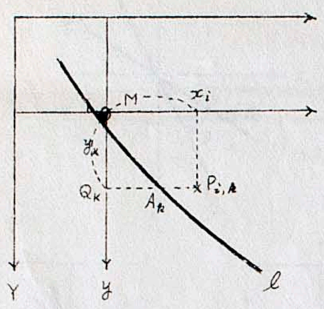
$$r = \frac{\sum_{ik} f_{i,k}(X_i - M_{(X)})(Y_k - M_{(Y)})}{N\,\sigma_{(X)}\,\sigma_{(Y)}}$$

但し

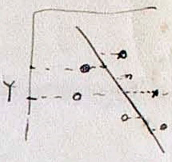
$$\sigma_{(X)} = \sqrt{\frac{1}{N}\sum_i f'_i (X_i - M_{(X)})^2}$$

$$\sigma_{(Y)} = \sqrt{\frac{1}{N}\sum_k f_k (Y_k - M_{(Y)})^2}$$

X, Y の標準測定値 x, y を用ひれば $M_{(X)}=0$, $M_{(Y)}=0$



最小偏差線 line of minimum deviation



ℓ: $x = ay + b$

$\overline{Q_k A_k} = ay_k + b$, $\overline{P_{i,k}A_k} = x_i - (ay_k + b)$

$$\sum_{i,k} f_{i,k}\overline{P_{i,k}A_k}^2 = \sum_{i,k} f_{i,k}\left[x_i - (ay_k + b)\right]^2 \quad \begin{pmatrix} i = 1, 2, \cdots\cdots, m \\ k = 1, 2, \cdots\cdots, n \end{pmatrix}$$

$\equiv S$ 極小

$$\frac{\partial S}{\partial a} = -2\sum f_{i,k}\left[x_i - (ay_k + b)\right]y_k = 0.$$

$$\frac{\partial S}{\partial b} = -2\sum f_{i,k}\left[x_i - (ay_k + b)\right] = 0.$$

$$\sum f_{i,k}\, x_i y_k - a\cdot\sum f_{i,k}\, y_k^2 - b\cdot\sum f_{i,k}\, y_k = 0.$$

$$\sum f_{i,k}\, x_i - a\cdot\sum f_{i,k}\, y_k - b\cdot\sum f_{i,k} = 0$$

$(x) = \frac{\sum f_{i,k} x_i}{N \sum f_{i,k}}$

$\sum f_{i,k}\, x_i = 0$ $(M_{(x)} = 0)$, $\sum f_{i,k}\, y_k = 0$ $(M_{(Y)} = 0)$, $\sum f_{i,k} = N$.

$b = 0.$ $\sum f_{i,k}\, x_i y_k - a\cdot\sum f_{i,k}\, y_k^2 = 0$

$\frac{1}{N}\sum f_{i,k}\, y_k^2 = \sigma_{(y)}^2 = 1.$ $a = \frac{\sum f_{i,k} x_i y_k}{N} = r.$

故に 最小偏差線 ℓ は点 M を通り，その方程式は

$$x = ry.$$

一般に

$$\frac{X - M_{(X)}}{\sigma_{(X)}} = r\cdot\frac{Y - M_{(Y)}}{\sigma_{(Y)}}. \qquad \ell:\ X - M_{(X)} = r\,\frac{\sigma_{(X)}}{\sigma_{(Y)}}\cdot\left(Y - M_{(Y)}\right).$$

もーつの最小偏差線 ℓ' は ℓ': $Y - M_{(Y)} = r\,\frac{\sigma_{(Y)}}{\sigma_{(X)}}\cdot\left(X - M_{(X)}\right).$

体重と身長の場合

$N = 817$, $M_{(X)} = 52.2^{kg}$, $M_{(Y)} = 160.0^{cm}$

$\sigma_{(X)} = 5.81^{kg}$, $\sigma_{(Y)} = 5.79^{cm}$

$r = +0.534$

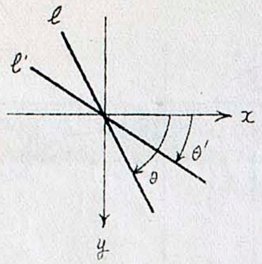
ℓ: $X - 52.2^{kg} = +0.536\,(Y - 160.0^{cm})$

ℓ': $Y - 160.0^{cm} = +0.532\,(X - 52.2^{kg})$

直線回歸の場合には，二つの最小偏差線は夫々二つの回歸線と一致する。

（證明は自ら試みよ）。

標準測定値を用ひれば



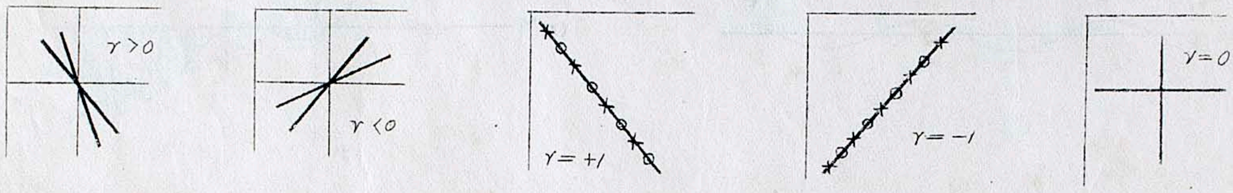
$l \qquad x = ry \qquad \tan\theta = \frac{1}{r}$

$l' \qquad y = rx \qquad \tan\theta' = r$

$r = +1, \qquad \theta = \theta' = 45^\circ$

$r = -1, \qquad \theta = \theta' = -45^\circ$

$r = 0. \qquad \theta = 90^\circ, \quad \theta' = 0$



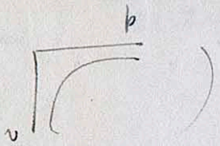
賃金と救民數	$r = -0.66$
父の子供數と息子の子供數	$r = +0.066$
母の子供數と娘の子供數	$r = +0.213$

Ⅱ　非直線回歸の場合 non-linear regression

この場合には

(1)　$r = 0$　なればとて必ずしも　無関係とは断定し得ない．　(例

(2)　完全相関でも　必ずしも　$r = \pm 1$　にならないことがある．　(例



rrelati
ratio

相　関　比

$$\eta_{(X)} = \frac{\sqrt{\frac{1}{N}\sum_{k} f_{k}(M_{k} - M_{(X)})^{2}}}{\sigma_{(X)}},$$

$$\eta_{(Y)} = \frac{\sqrt{\frac{1}{N}\sum_{i} f'_{i}(M'_{i} - M_{(Y)})^{2}}}{\sigma_{(Y)}},$$

この二つの相関比が

共に　0　なるときは無関係

共に　+1　なるときは完全相関

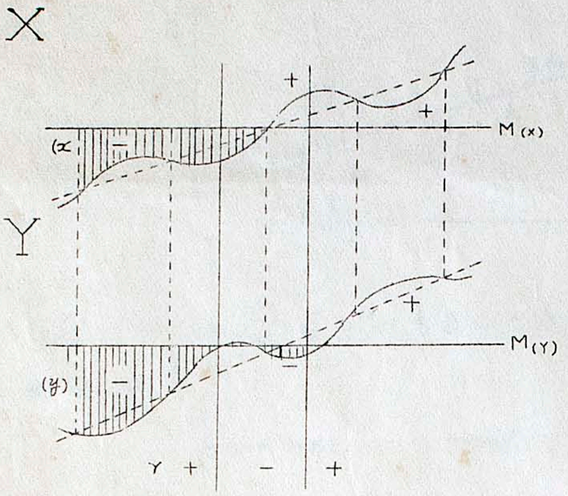
多變量の相関

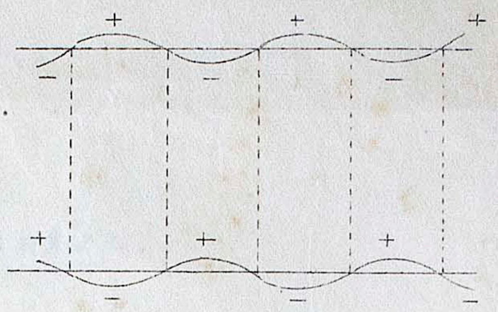
れは何等かの
論的基礎
而つ説明す
はねばならない

相関関係の測定に関する注意.（二つの事象の數値の中，何と何とを比較するが，本質的にその関係を明にする所以であるか）．それは特に時系列の場合に於て最も重要性を帯ぶ。

第六章　時　系　列　time series

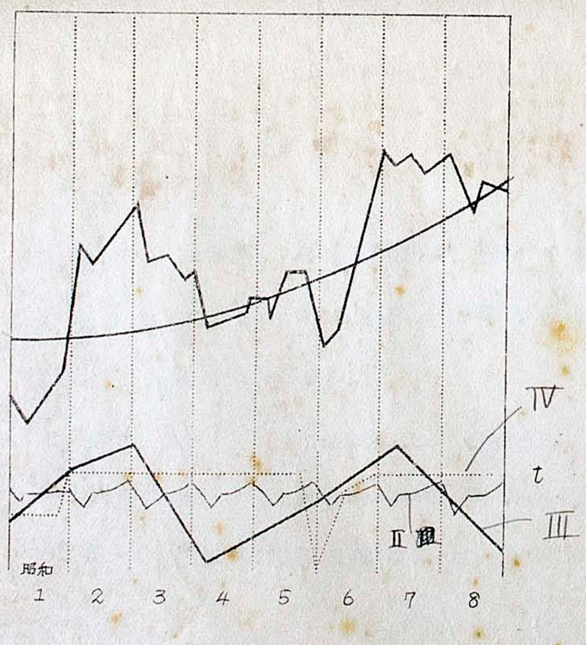
米の産額Xと米價Y





XとYとを其の儘で比較すれば，短期間では負の相関係數を得ることもあれど，長期間では，共に趨勢的に増して居る．結局 $\gamma > 0$.

一般的趨勢を除いて考へれば，明かに $\gamma < 0$.



時系列の分析 Persons Wagemann } Anderson, Math. statistik

I 趨勢 trend

II 季節的變化 seasonal variat.

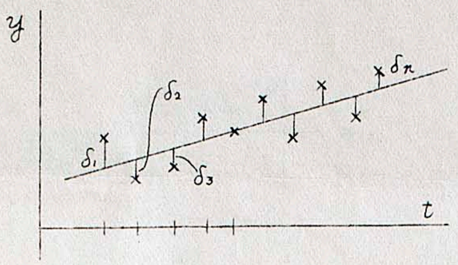
III 循環的變化 cyclic ,,

IV 不規則變化 irregular ,,

趨勢の求め方に就ての二問題

(1) 最も適當な函數の形を選ぶこと． (2). 次に函數の中の未定の係數を定めること．

この第二の問題は，最小自乘法によつて解かれる．



$\delta_1^2+\delta_2^2+\cdots\cdots+\delta_n^2=\sum\delta_k^2$ を最小にする様に係数を定める。

直線の場合

$$y=a+bt \qquad S\equiv\sum[y_k-(a+bt_k)]^2=Min \qquad \frac{\partial S}{\partial a}=0,\ \frac{\partial S}{\partial b}=0.$$

a, b は次の二式から求める.

$$\begin{cases} na+b\cdot\sum t_k=\sum y_k \\ a\cdot\sum t_k+b\cdot\sum t_k^2=\sum t_k y_k \end{cases}$$

拋物線の場合　　$y=a+bt+ct^2$,

$$\begin{cases} na+b\sum t_k+c\sum t_k^2=\sum y_k, \\ a\sum t_k+b\sum t_k^2+c\sum t_k^3=\sum t_k y_k, \\ a\sum t_k^2+b\sum t_k^3+c\sum t_k^4=\sum t_k^2 y_k. \end{cases}$$

季節的變化

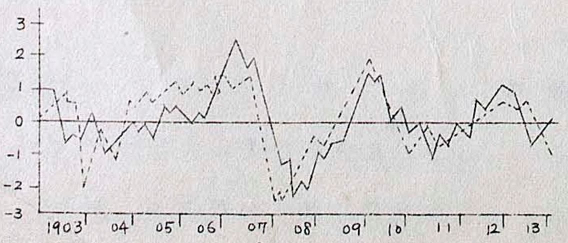
循環的變化

時系列から　趨勢と季節的變化を除き去つたものは循環的變化と，不規則變化の結合である．パースンズは　此二者を　分離し得ないと考へ，その結合を標準測定値に直した數値を循環（cycle）と呼んだ．

Mitchell,　　Business Cycles.

Wagemann,　　Konjunkturlehre　（小島昌太郎譯　景氣變動論）

時系列の比較中最も重要なるは　サイクル　の比較である．



アメリカ〔大戰前〕

物價指數のサイクル（實線）

銑鐵生産額のサイクル（点線）

恐慌の週期　約４０個月．

—終—

理論と実証との統一、

[illegible] 外 内