統計法

項目，公式等

金之助

小倉~~講~~ 述

はしがき

この講義は統計学の中で，所謂統計解析の一部としての特に数理的なる方法を，数式の計算技術よりも数式の意味に注意と批判を加へつゝ初等的に講ずるを目的とする，

統計集団	———大量観察——→	統計値	———統計解析——→	統計的結論

~~[illegible]「統計数学」[illegible]「数学」でもなければ，また単なる計[illegible]~~

参考書

蜷川虎三，~~統計[illegible]基本問題~~ 統計学概論

森田優三，統計概論

小林 新，経済統計學

改造社版，経済學全集（43，44）統計學［有澤，小倉，森，等々］，（52）本邦社会統計論［高野］．

小倉，統計的研究法

~~Bowley, Elements of statistics.~~

Mills, Statistical methods.

Yule, Theory of statistics.

~~Rietz, Hand book of mathematical statistics.~~

~~Winkler, Grundriss der Statistik.~~

~~Darmois Statistique mathematique.~~

Risser et ~~Traynard, Statistique~~ mathematique,

~~Charlier, Grundzüge d. mathematischen Statistik~~

表

對数表（四桁，五桁，七桁）

Barlow, Tables of squares, cubes, square roots, etc,

~~Pearson, Tables f. statisticians.~~

~~Glover, Tables of applied mathematics.~~

~~Crelle 又は Zimmermann, Rechentafel.~~

~~資料~~

~~日本帝国統計年鑑~~

~~天野恒太，日本國勢圖會~~

ウケ

うん

(1)

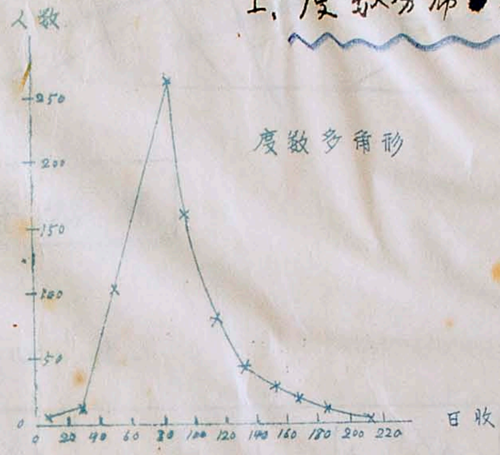
第一章　度数分布

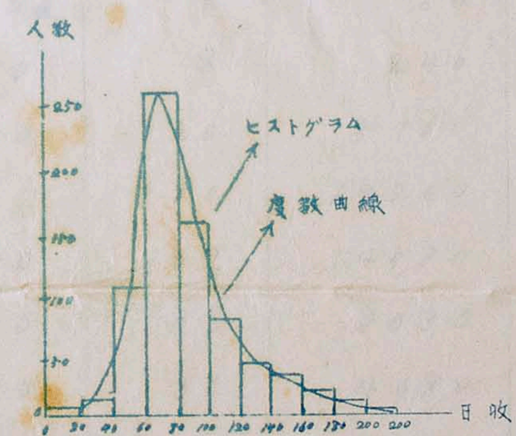
I. 度数分布●表．分布曲線

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（銭）	中央値	人数	累積人数
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
	$\Sigma f_x = N$

注意．(1)級の間隔（一定）を問題に就て適當に選ぶこと．級の數．級の限界．

(2)整正なる理想的分布曲線の存在の假定．

~~級の限界の定め方~~

(3) 比率の使用（比較の場合）

~~確率曲線．~~

~~[誤差の法則 [illegible]]~~

II. 度数分布の型

對称

非對称

J字形

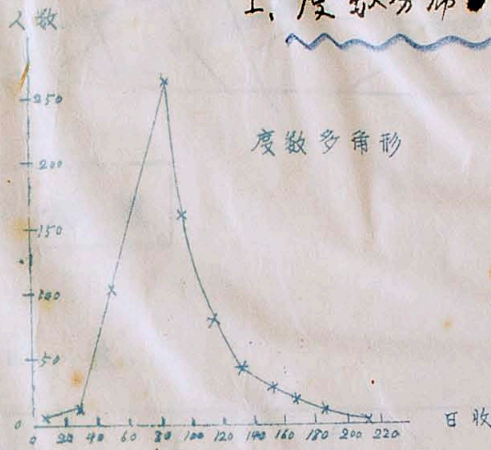
U字形

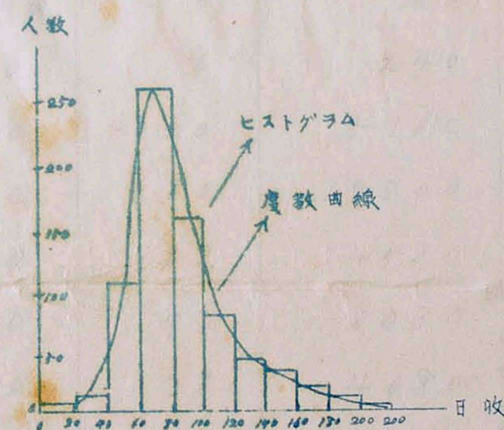
第一章　度数分布

I. 度数分布表. 分布曲線

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	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
…	…
X_n	f_n
	$\sum f_k = N$

注意. (1) 級の間隔（一定）を問題に就て適當に選ぶこと. 級の數. 級の限界.

(2) 整正なる理想的分布曲線の存在の假定.

~~級の限界の定め方~~

(3) 比率の使用（比較の場合）

~~確率曲線~~
~~[誤差の法則, ガウス法則]~~

II. 度数分布の型

對稱

非對稱

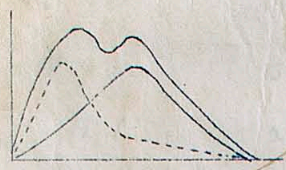
J字形

U字形

~~ピアソンの方法~~

~~ブルンス、シャルリエの方法（直交函数）~~

複雑な形



第二章　平均値

総変量の代表値.　　中心値.

I. 算術平均 (平等)

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

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

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

簡便計算

$$M = A + \frac{1}{N} \Sigma f_k \delta_k.$$

$$M = X_s + \frac{h}{N} \Sigma f_k (k-s).$$

$s = 5.$

$X_s = 90$銭

$$M = 90 + \frac{20}{704}(-159)$$

$$= 85.483\text{銭}$$

X_k	$k-s$	f_k	$f_k(k-s)$
10銭	−4	6	−24
30	−3	8	−24
50	−2	102	−204
70	−1	260	−260
90	0	162	0

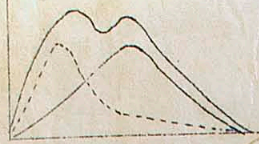
(3)

国米(4),

万平均,

(4)

27字×25＝675



第二章　平均値

総変量の代表値.　中心値.

I. 算術平均 (平等)

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

X_k	f_k	$f_k X_k$
10銭	6	60
30	8	240
50	120	5100
70	210	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

$$M = \frac{60180}{704} = 85.483銭$$

(3)

簡便計算

$$M = A + \frac{1}{N} \cdot \Sigma f_k \delta_k .$$

$$M = X_s + \frac{h}{N} \cdot \Sigma f_k (k-s).$$

$s = 5.$

$X_s = 90$銭

X_k	$k-s$	f_k	$f_k(k-s)$
10銭	−4	6	−24
30	−3	8	−24
50	−2	102	−204
70	−1	260	−260
90	0	162	0
110	+1	73	+73
130	2	36	72
150	3	30	90
170	4	18	72
190	5	8	40
210	6	1	6
		704	−159

$$M = 90 + \frac{20}{704}(-159)$$

$= 85.483$銭

…の

国米(4),

与平均,

21
22
23
24
25

(4)

27字×25＝675

第二章　平均値

総変量の代表値.　　　中心値.

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

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

(3)

簡便計算

$$M = A + \frac{1}{N}\cdot \sum f_k \delta_k.$$

$$M = X_s + \frac{h}{N}\cdot \sum f_k (k-s).$$

$s = 5.$

X_k	$k-s$	f_k	$f_k(k-s)$
10銭	−4	6	−24
30	−3	8	−24
50	−2	102	−204
70	−1	260	−260
90	0	162	0
110	+1	73	+73
130	2	36	72
150	3	30	90
170	4	18	72
190	5	8	40
210	6	1	6
		[illegible]	[illegible]

$X_s = 90$銭

$$M = 90 + \frac{20}{704}(-159)$$

$= 85.483$銭

の

国米(4),

可平均,

(4)

27字×25＝675

第二章　平均値

総変量の代表値.　　中心値.

I. 算術平均 (平等)

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

X_k	f_k	$f_k X_k$
10銭	6	60
30	8	240
50	120	5100
70	210	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

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

(3)

簡便計算

$$M = A + \frac{1}{N} \Sigma f_k \delta_k.$$

$$M = X_s + \frac{h}{N} \Sigma f_k (k-s).$$

$s = 5.$

$X_s = 90$銭

$$M = 90 + \frac{20}{704}(-159)$$

$= 85.483$銭

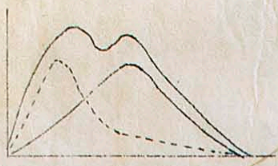
X_k	$k-s$	f_k	$f_k(k-s)$
10銭	−4	6	−24
30	−3	8	−24
50	−2	102	−204
70	−1	260	−260
90	0	162	0
110	+1	73	+73
130	2	36	72
150	3	30	90
170	4	18	72
190	5	8	40
210	6	1	6
		704	159

米(4),
平均,

(4)

27字×25＝675

複雑な形



ブルンス、シャルリエの方法（直交函数）

第二章　平均値

総変量の代表値.　　中心値.

I. 算術平均（平等）

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

X_k	f_k	$f_k X_k$
10銭	6	60
30	8	240
50	120	5100
70	210	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

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

(3)

簡便計算

$$M = A + \frac{1}{N} \cdot \Sigma f_k \delta_k.$$

$$M = X_s + \frac{h}{N} \cdot \Sigma f_k (k-s).$$

X_k	$k-s$	f_k	$f_k(k-s)$
10銭	−4	6	−24
30	−3	8	−24
50	−2	102	−204
70	−1	260	−260
90	0	162	0
110	+1	73	+73
130	2	36	72
150	3	30	90
170	4	18	72
190	5	8	40
210	6	1	6
		704	−159

$s = 5.$

$X_s = 90$ 銭

$$M = 90 + \frac{20}{704}(-159)$$

$$= 85.483\text{銭}$$

国米(4),

万平均,

(4)

27字×25=675

指数

統計値の比較を容易にするための指数

年次	価格 鋼鉄（噸）	価格 小麦（ブッセル）	価格指数（1890－1896基準） 鋼鉄	価格指数（1890－1896基準） 小麦
1890	30 弗	1.05 弗	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

綜合指数、

生活費指数

数年間の研究により、或る年の指数が：

食費	住居費	被服費	燈火燃料費	雑費
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、生活費指数（秤量平均）

$$\frac{\sum m_k X_k}{\sum m_k} = \frac{108\times 40 + 102\times 16 + \cdots}{100} = 106$$

物価指数

物価水準の一般的変化を知るを目的とす。

日本銀行、米、大麦、石油、砂糖等日用品56品の価格指数の単なる算術平均。

ダイヤモンド社、82品に重み[例、内地米(65)，外国米(4)，小麦(9)，牛肉(3)，材木(10)，…]を附した幾何平均。

郡 菊之助著，物価指数論

(4)

II、幾何平均

$$G = \sqrt[N]{X_1^{f_1} \cdot X_2^{f_2} \cdots\cdots X_n^{f_n}}$$

$$\log G = \frac{f_1 \log X_1 + \cdots + f_n \log X_n}{N} = \frac{1}{N} \cdot \sum (f_k \log X_k)$$

X_k	$\log X_k$	f_k	$f_k \log X_k$
10貫	1.00000	6	6.00000
30	1.47712	8	11.81696
50	1.69897	102	171.59597
70	1.84510	260	479.72600
90	1.95424	162	316.58688
110	2.04139	73	149.02147
130	2.11394	36	76.10184
150	2.17609	30	65.28270
170	2.23045	18	40.14810
190	2.27875	8	18.23000
210	2.32222	1	2.32222
計 —	——	704	1336.83214

$$\log G = \frac{\sum (f_k \log X_k)}{N}$$

$$= \frac{1336.83214}{704}$$

$$= 1.90274$$

$G = 79.936$ 貫

数ノ比較　$b-a$ 絶對的　$\frac{b}{a}$ 相對的

時系列ノ場合ハ，比ニヨル比較ノ方ガ有意義ノ場合ガ多イ．（人口ノ増加率，物價ノ変化率）．

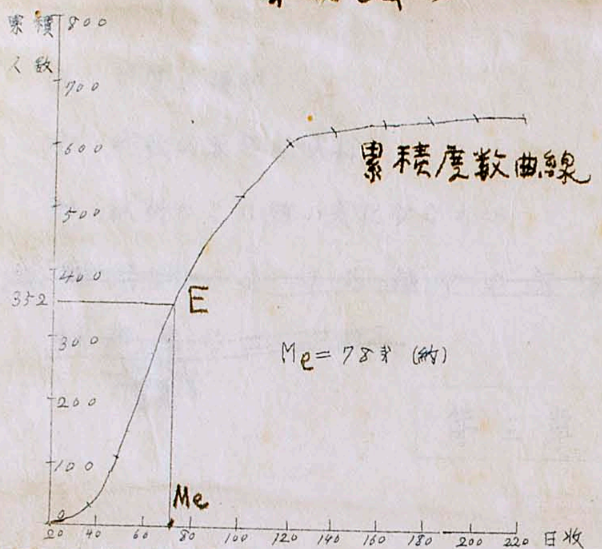
変数自身ノ平均デナク，変数ノ比率ノ平均ヲ取ルニハ，幾何平均ノ方ガ合理的デアル．

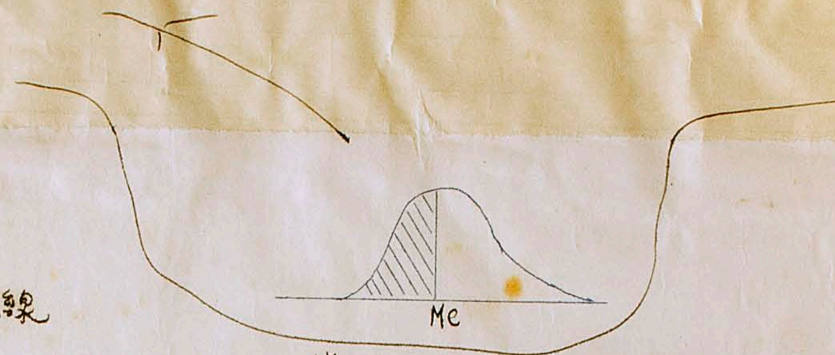
		大正10	昭和3	昭和3	昭和10
商品A		100	200	100	50
商品B		100	50	100	200
平均	M	100	125	100	125
	G	100	100	100	100

III、中央値（メディアン）

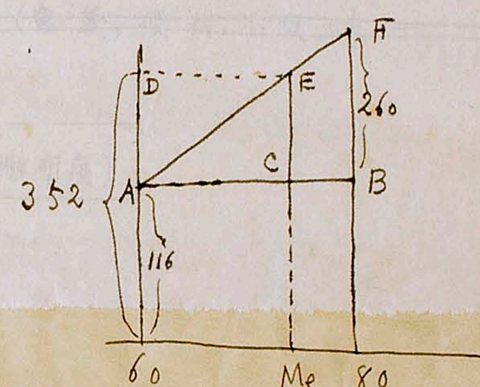
~~変数~~ 度数N が奇数ノトキハ，ソノ中央ノ値．

第一法





第二法

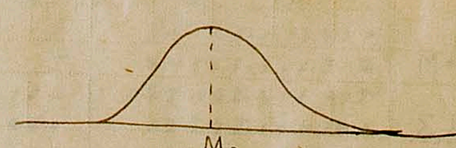


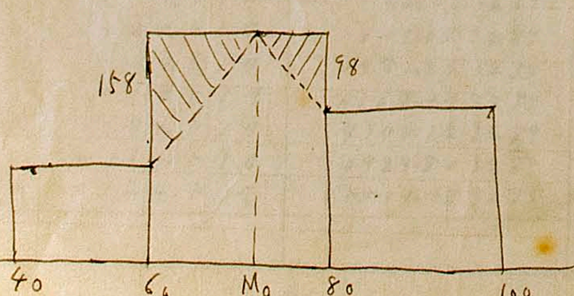
$$Me = 60^{才} + AC.$$

$$\frac{AC}{AB} = \frac{EC}{FB} \qquad AC = 20 \times \frac{352-116}{260}.$$

$$Me = 78.15^{才}$$

IV. 並数（モード）
ナミスウ





$$Mo = 60 + 20 \times \frac{158}{158 + 98}$$

平均値選擇の標準

(1) 明確に規定されるもの

(2) 変量全部の値に基くもの

(3) 計算の便利

(4) 代数的取扱の便利

(5) 試料により値の変化少きもの

~~数理統計に於ける種々の方法の比較（意義，便利，不定性——純粋学との比較），~~

第三章 分散度（撒布度）

変量の平均（中心）の周りに於ける開き（分散）の程度，

I、平均偏差

偏差 $X_k - M_i$

$\Delta = \frac{1}{N}\cdot\Sigma(f_k \cdot |X_k - M_i|)$ $M = 78$ 匁， $\Delta = 24.27$ 匁

II、標準偏差

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

$M = 85.483$ 匁

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

$$= 1029.028119$$

$$\sigma = \sqrt{1029.028119}$$

$= 32.0785$ 匁

簡便計算

$$\sigma^2 = \frac{1}{N}\cdot\sum f_k\,\delta_k^2 - d^2 \qquad M - A = d.$$

$$\delta_k = X_k - A.$$

$$\sigma^2 = \frac{h^2}{N}\cdot\sum f_k (k-s)^2 - d^2$$

X_k	$k-s$	$(k-s)^2$	f_k	$f_k(k-s)^2$
10 ギ	−4	16	6	96
30	−3	9	8	72
50	−2	4	102	408
70	−1	1	260	260
90	0	0	162	0
110	+1	1	73	73
130	2	4	36	144
150	3	9	30	270
170	4	16	18	288
190	5	25	8	200
210	6	36	1	36
			704	1847

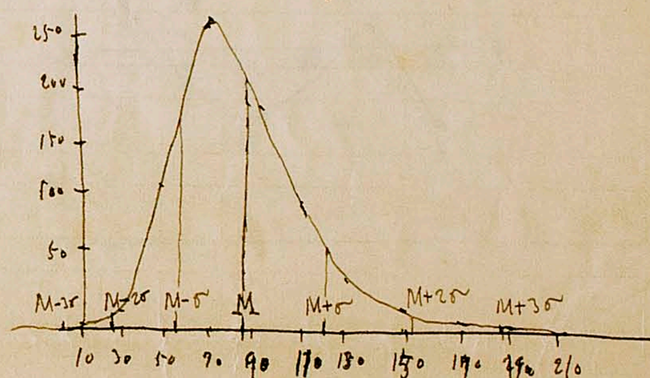
$M = 85.483$

$X_s = 90, \quad s = 5$

$$\sigma^2 = \frac{(20)^2}{704}\times 1847 - (90 - 85.483)^2$$

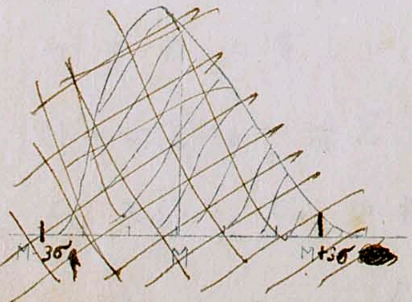
$$= 1029.028119$$

$\sigma = 32.0785$ 套



注意　$M - 3.\sigma = 85.48 - 96.24 = -10.76$

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



故に殆んど総ての変量が $M-3\sigma$ と $M+3\sigma$ との間に挟まれる。

對稱分布曲線の場合には，上の二数の間に全変量の 99.73% が含まれる。

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

最小自乗法との関係

一般に a からの偏差を取り，a からの開き（分散）の程度を

$$D = \sqrt{\frac{1}{N}\cdot\sum\left[f_h (X_h - a)^2\right]}$$

で測る。　a の位置を変ずるときは，D も変化する．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，偏差の自乗の和 $\sum f_h (X_h - a)^2$ 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．

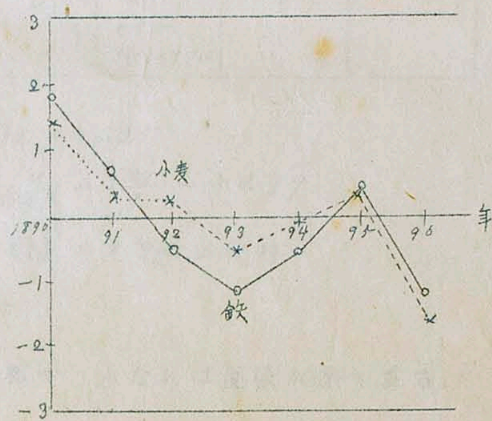
変化係数（比較に用ひる）　$V=100\dfrac{\sigma}{M}$.

非對稱度　$S=\dfrac{M-M_0}{\sigma}$.

標準測定値

$$x_k=\frac{X_k-M}{\sigma}=\frac{\xi_k}{\sigma},\quad x'_k=\frac{X'_k-M'}{\sigma'}=\frac{\xi'_k}{\sigma'}.$$

鋼鉄		小麦	
$M=25$ $\sigma=2.67$		$M'=0.9$ $\sigma'=0.097$	
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変量 $x_1, x_2, \dots, x_n$ の算術平均 M' は 0 に等しい.

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

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

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

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

第四章　相関関係

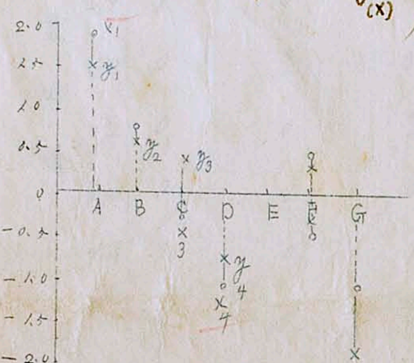
I、相関ノ概念

生徒	入	学	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	$\sigma^2_{(X)}=7.1425$ $\sigma_{(X)}=2.67$	0	$M(Y)=70$	0	$\sigma^2_{(Y)}=95$ $\sigma_{(Y)}=9.74$	0	

$x=\frac{\xi}{\sigma_{(X)}}$,　　$y=\frac{\eta}{\sigma_{(Y)}}$,

$\Sigma x_k y_k=+6.25$

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



対応する二つの値の開きの平方の和の平均

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

が大なれば　関係が粗で　小なれば関係が密である。

(基本的考へ)。

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

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

$$=1-\frac{2}{n}\Sigma x_ky_k+1$$

$$=2-\frac{2}{n}\Sigma x_ky_k.$$

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

由て　$r=\frac{1}{n}\Sigma x_ky_k$

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

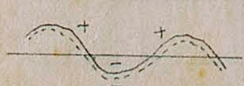
若し　$(x_k+y_k)^2=x_k^2+2x_k+y_k^2$ から出発すれば　$r=-1+\frac{1}{2n}\Sigma\cdot(x_k+y_k)^2.$

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

~~[標準化した値で表わしたとき]~~

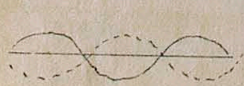
$r=+1$ なるための必要にして且つ十分なる條件

$x_1=y_1,\quad x_2=y_2,\ \dots\ x_n=y_n.$　相重る。



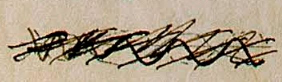
$r=-1$　なるための條件

$x_1=-y_1,\ \dots\dots\ x_n=-y_n.$　対称。



$r>0$　順相関
$r<0$　逆相関
$r=0$　無関係

$r=\pm1$　完全なる相関



γ を 相関係数と呼び，その値の大小によつて相関の程度を測定する．

入学と卒業成績の相関 $\gamma = \frac{1}{n} \sum x_K y_K = \frac{6.25}{17} = +0.89$

英國 救貧法を布いて居る 38 地方に於ける，農業労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 $\gamma = -0.66$

$$\boxed{\gamma = \frac{1}{n} \sum \frac{x_k}{\sigma(x)} \frac{y_k}{\sigma(Y)} = \frac{\sum (X_k - M(x))(Y_k - M(Y))}{n\,\sigma(x)\,\sigma(Y)}}.$$

II、相関表

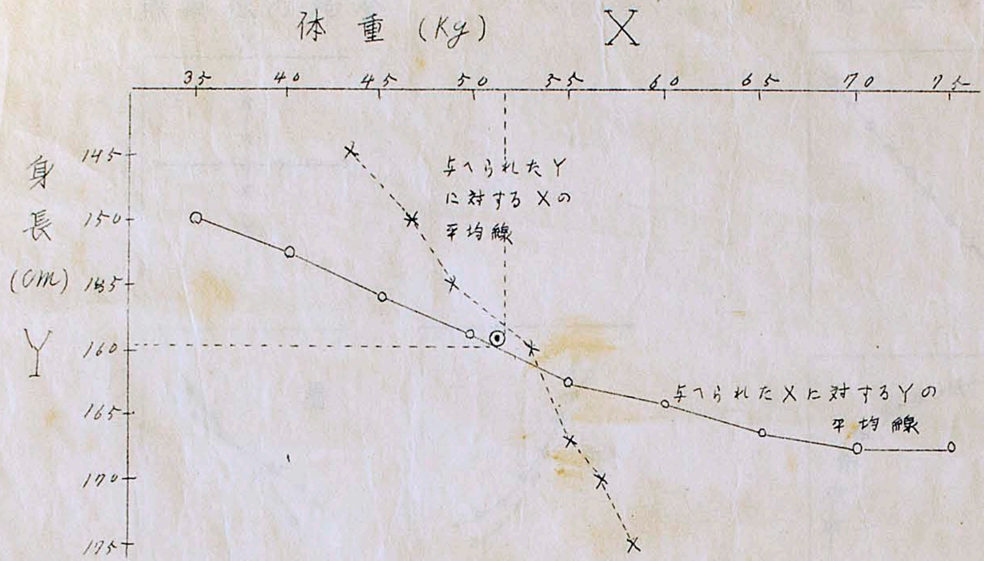
入學成績

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

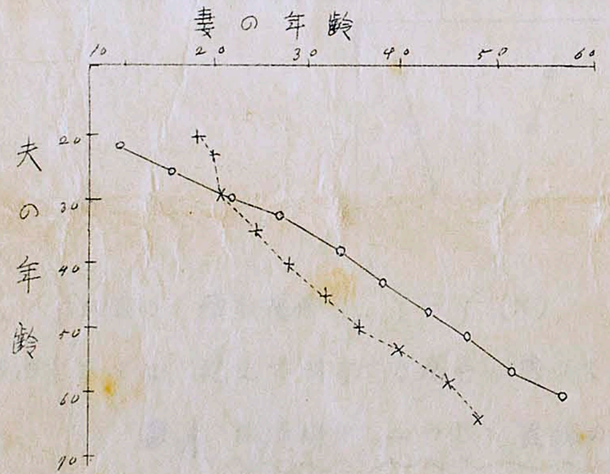
身長と体重の相関表（昭和4年東京市隣接五郡壮丁）

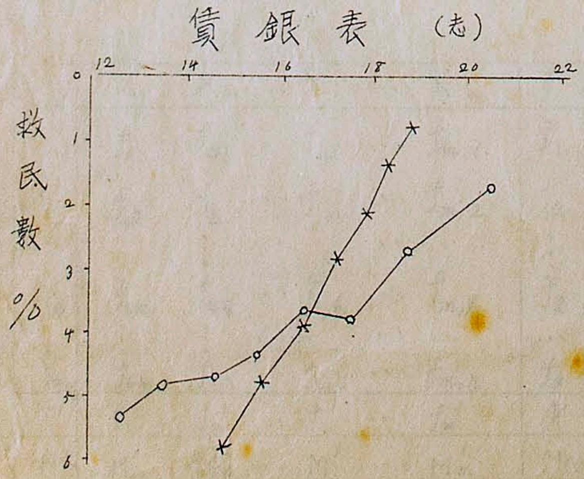
体重 / 身長	35 Kg	40	45	50	55	60	65	70	75	合計	平均 Kg
145 cm		3	5	2						10	44.5
150	1	9	29	19	5					63	46.4
155		10	53	86	37	6	1			193	49.5
160		3	36	107	103	30	5	1		285	52.5
165		1	11	48	75	42	11	1	1	190	55.0
170			1	10	22	19	8	2	1	63	57.6
175				2	4	4	2	1		13	58.5
合計	1	26	135	274	246	101	27	5	2	817	52.2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

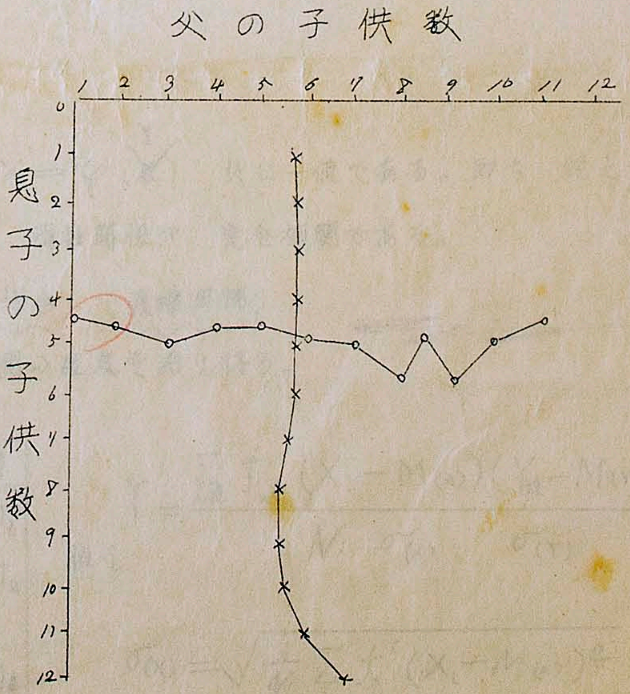
III. 相関圖　(回帰線)



昭和四年内地ノ結婚

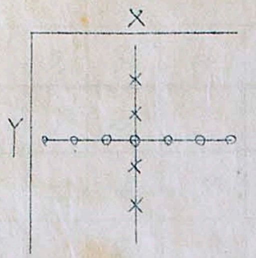




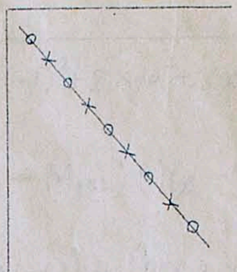


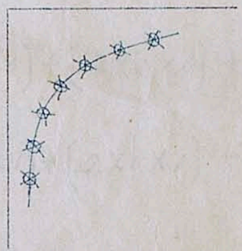
IV 相関の程度

無関係の場合

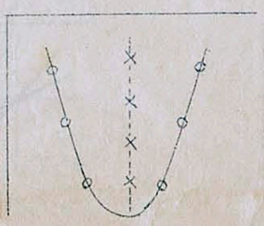


完全の場合





完全相関と函数的関係との相違. [一値函数]



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って, 函数関係が 完全相関である。

A. 回帰線が 二つとも直線の場合 (直線回帰)

この場合には 相関係数 r によって 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'	$M_{(Y)}$	

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

但し

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

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

Yの計算の例

X

Y	2	3	4	f	M
1	3	2	1	6	2.66
2	2	4	2	8	3.00
3	1	2	3	6	3.33
f'	6	8	6	20	3.00
M'	1.66	2.00	2.33	2.00	

$$\sigma_{(X)} = \sqrt{\frac{1}{20}\left\{6\times(-1)^2+8\times0^2+6\times1^2\right\}} = \sqrt{\frac{12}{20}},\quad M_{(X)}=3,$$

$$\sigma_{(Y)} = \sqrt{\frac{1}{20}\left\{6\times(-1)^2+8\times10^2+6\times1^2\right\}} = \sqrt{\frac{12}{20}},\quad M_{(Y)}=2,$$

$$\sum_{i,k} f_{i,k}(X_i - M_{(X)})(Y_k - M_{(Y)})$$

$$= \{3\times(-1)\times(-1)\} + \{2\times0\times0\} + \{1\times1\times(-1)\}$$

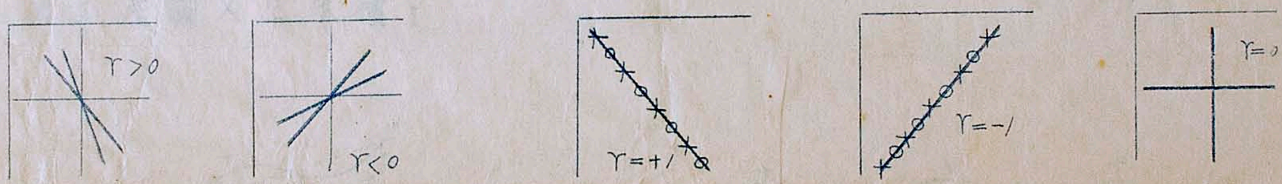
$$+ \{2\times(-1)\times0\} + \{4\times0\times0\} + \{2\times1\times0\}$$

$$+ \{1\times(-1)\times1\} + \{2\times0\times1\} + \{3\times1\times1\}$$

$$= 3-1-1+3=4.$$

$$r = \frac{4}{20\sqrt{\frac{12}{20}}\sqrt{\frac{12}{20}}} = \frac{4}{12} = \frac{1}{3}$$

$= +0.333$



身長ト体重 $r=+0.534$.

債銀と救民数 $r=-0.66$

父の子供数と息子の子供数 $r=+0.066$

母の子供数と娘の子供数 $r=+0.213$

B. 非直線回帰の場合

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.

(2) 完全相関でも 必ずしも $r=\pm1$ にならないことがある。

相関比

$$\eta_{(x)} = \frac{\sqrt{\frac{1}{n}\sum_{h} f_h (M_h - M_{(x)})^2}}{\sigma_{(x)}},$$

$$\eta_{(Y)} = \frac{\sqrt{\frac{1}{n}\sum_{i} f_i' (M_i' - M_{(Y)})^2}}{\sigma_{(Y)}}.$$

この二つの相関比が

共に 0 なるときは無関係

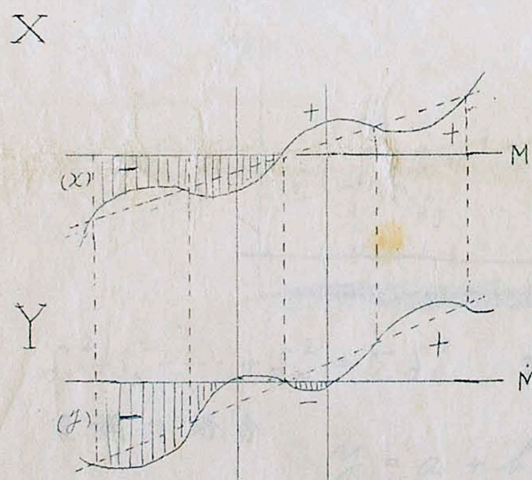
共に $+1$ なるときは完全相関、

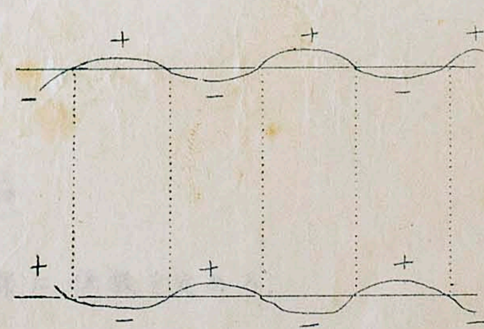
多変量の相関

相関関係の測定に関する注意、（二つの事象の数値の中，何と何とを比較するか，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。

第五章 時系列

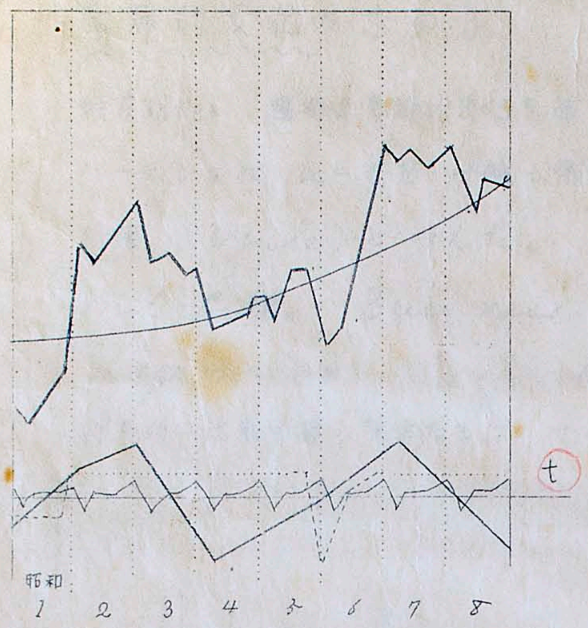
米の産額Xと米價Y





XとYとを其の値で比較すれば，短期間では負の相関係数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 $r>0$。

一般的趨勢を除いて考へれば，明かに $r<0$.



時系列の分析

I. 趨勢

II. 季節的変化

III. 循環的変化

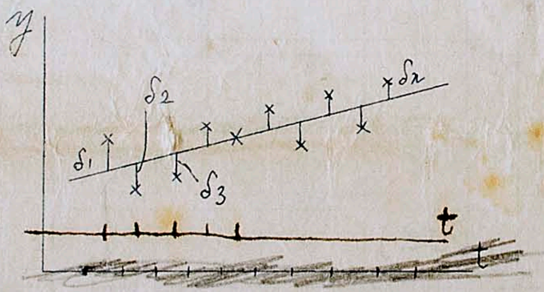
IV 不規則変化

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.



$\delta_1^2+\delta_2^2+\cdots+\delta_n^2=\sum\delta_k^2$ を最小にする様に係數を定める.

直線の場合

$$y=a+bt. \qquad \cancel{\sum\left[y_k-(a+bt_k)\right]^2=\min. \quad \frac{\partial S}{\partial a}=0,\ \frac{\partial S}{\partial b}=0}$$

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 $y=a+bt+ct^2,$

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

季節的変化の求め方.

時系列から　趨勢と季節的変化を除き去つたものは循環的変化と　不規変化の結合である.

パーースンズは　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数値を

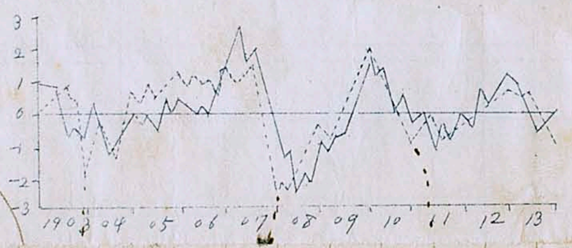
循環　(cycle) と呼んだ.

~~(Lacombe,)~~

Mitchell,　Business cycles

Wagemann,　Konjunkturlehre (小島昌太郎訳 景氣変動論)、

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・[大戦前]

物價指數のサイクル (実線)、

銑鐵生産額のサイクル (点線)、

恐慌の週期 約40個月、

(絵)

(17)

第三學年特別講義

統計法

理學博士 小倉金之助述

昭和九年十月

東京物理學校

Crelle 又は Zimmermann, Rechentafel.

資　料

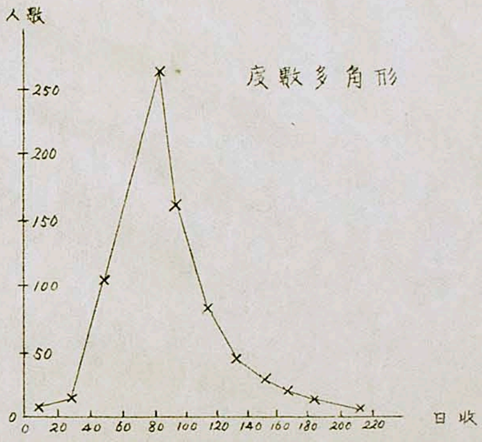
日本帝國統計年鑑

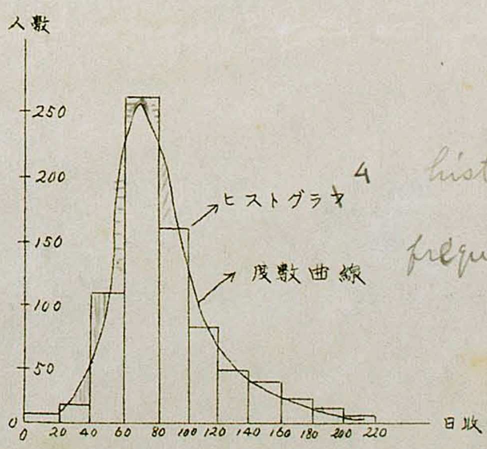
矢野恒太,　　日本國勢圖會.

第一章　度數分布

frequency distribution

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（銭）	中央値（銭）	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	



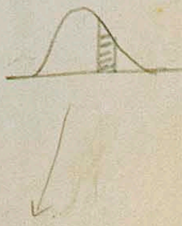


histogram

frequency curve

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。



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

II 幾何平均

$$G = \sqrt[N]{X_1^{f_1} \cdot X_2^{f_2} \text{-----------} X_n^{f_n}}$$

$$\log G = \frac{f_1 \log X_1 + \text{-------} + f_n \log X_n}{N} = \frac{1}{N} \cdot \Sigma (f_k \log X_k)$$

X_k	$\log X_k$	f_k	$f_k \log X_k$
10歳	1.00000	6	6.00000
30	1.47712	8	11.81696
50	1.69897	102	171.59597
70	1.84510	260	479.72600
90	1.95424	162	316.58688
110	2.04139	73	149.02147
130	2.11394	36	76.10184
150	2.17609	30	65.28270
170	2.23045	18	40.14810
190	2.27875	8	18.23000
210	2.32222	1	2.32222
合計 —	———	704	1336.83214

$$\log G = \frac{\Sigma (f_k \log X_k)}{N} = \frac{1336.83214}{704} = 1.90274$$

$$G = 79.936\text{歳}$$

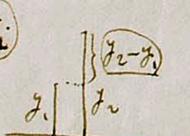
$y = a \cdot b^x$

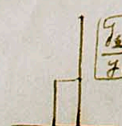
$\log y = \log a + x \cdot \log b$

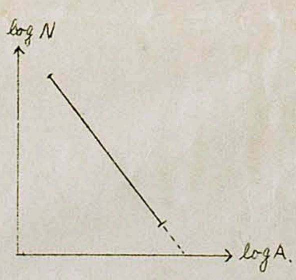
$\log y_2 - \log y_1 = \log \frac{y_2}{y_1}$

幾何平均は半對數方眼紙上に畫ける縱線の長さの算術平均である。

絶對的變化と相對的變化。(差と比)。







對數方眼紙

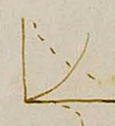
A円以上の所得を得る人數をNとする。

パレト線

$\log N = a + b \cdot \log A$.

汐見三郎外三氏著

國民所得の分配。



$y = a x^b$.

$\log y = \log a + b \cdot \log x$

$\eta = \log a + b \cdot \xi$

$A = \pi r^2$

コノニツノ lim ノ existence ハ 認メナイ。

median ノ 場合

$n = 2 \quad x = \frac{x_1 + x_2}{2}$.

$n = 4 \quad x = \frac{x_4 x_3 - x_2 x_1}{(x_4 + x_3) - (x_2 + x_1)}$.

n=6 以上ハ 計算ガ 不便トナル

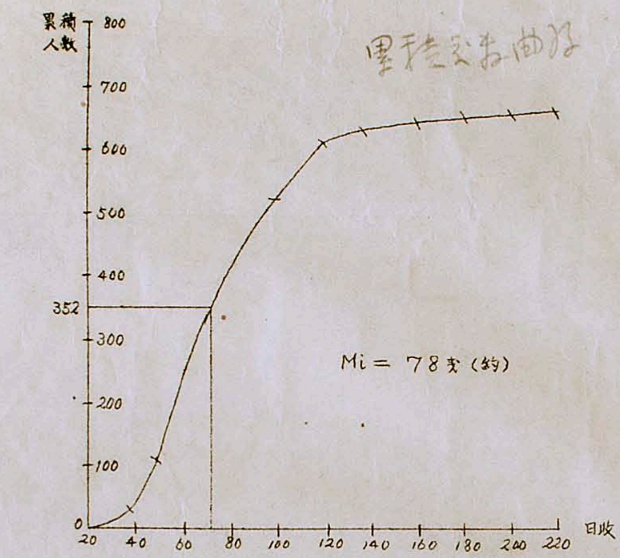
empirical formula

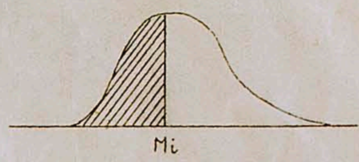
平均値選擇の標準

median

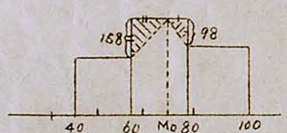
Ⅲ 中央値（メディアン）

累積度数曲線





Ⅳ 並　数（モード） mode



$$M_0 = 60 + 20 \times \frac{260-102}{(260-102)+(260-162)} = 72.43\text{式}$$

平均値選擇の標準

(1)　明確に規定されるもの

(2)　變量全部の値に基くもの

(3)　計算の便利

(4)　代數的取扱の便利

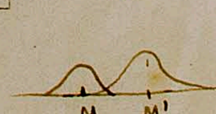
(5)　試料により値の變化少きもの

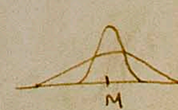
數理統計に於ける種々の方法の比較（意義，便利，不安性——

純粋數學との比較）

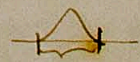
第三章　分散度（撒布度） dispersion

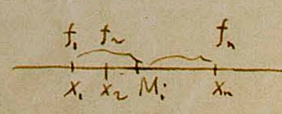
變量の平均（中心）の周りに於ける開き（分散）の程度，





Ⅰ 平均偏差 mean deviati





deviation —— 偏差　$X_k - M_i$

$$\Delta = \frac{1}{N} \cdot \Sigma (f_k \cdot |X_k - M_i|) \qquad M_i = 78\text{式}, \qquad \Delta = 24.27\text{式}$$

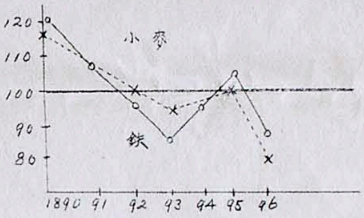
Ⅱ 標準偏差

standard deviation

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)	小麦 (ブッセル)	鋼鉄	小麦
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　$I_1 = \dfrac{100X_1}{M},\quad I_2 = \dfrac{100X_2}{M},\ \cdots\cdots I_n = \dfrac{100X_n}{M}$

單一指數

綜合指數

生活費指數

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　$\dfrac{\sum m_K X_K}{\sum m_K} = \dfrac{108\times 40 + 102\times 16 + \cdots\cdots}{100} = 106$

(秤量平均)

Quetelet

isher

物價指數

物價平準の一般的變化を知るを目的とす。

日本銀行　　米，大麥，石油，砂糖等日用品５６品の價格指數の單なる算術平均

ダイヤモンド社．　８２品に重み〔例，内地米(65)，外國米(4)，小麥(9)，牛肉(3)，材木(10)---------〕を附した幾何平均．

郡菊之助著．　物價指數論．

大正元年　大正八年　昭和五年

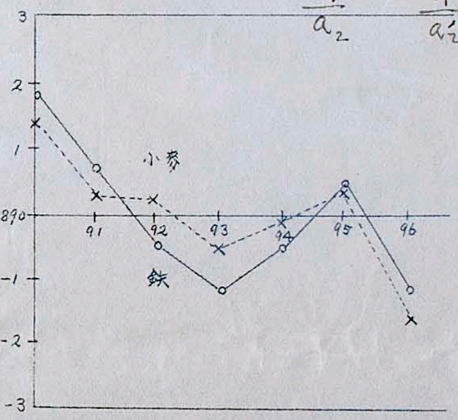
二年间の比

$\frac{a_1}{a_2}$　　$\frac{a_1'}{a_2'}$

標準測定値

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

鋼鉄		小麥	
$M = 25.$ $\sigma = 2.67$		$M' = 0.9$ $\sigma' = 0.097$	
ξ	$x (= \frac{\xi}{\sigma})$	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。

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

$\underline{\sum x_k = 0}$

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

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

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

$\frac{1}{n}\sum x_k^2 = 1$

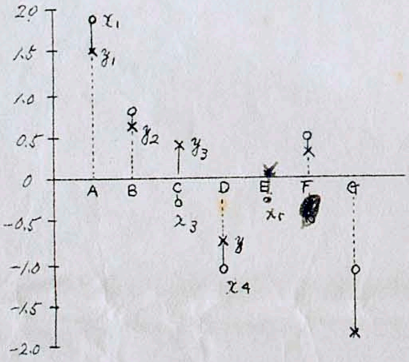
$$\left(\because \sqrt{\frac{\sum(X_k - M)^2}{n}} = \sigma.\right)$$

第五章　相関関係　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	

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

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



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　関係が粗で　小なれば関係が密である。

（基本的考へ）

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

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

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

$$= 2 - \frac{2}{n}\sum x_k y_k$$

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

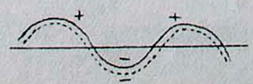
由て　$r = \frac{1}{n}\sum x_k y_k$

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

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

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

$r = +1$　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]

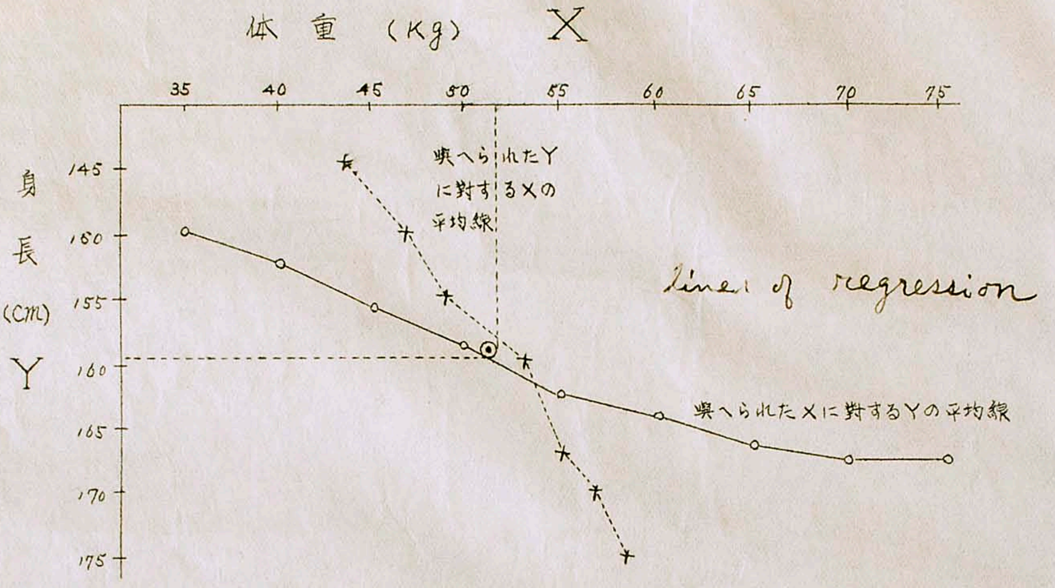


$x_1 = y_1,\quad x_2 = y_2, \text{------} \; x_n = y_n.$　相重る.

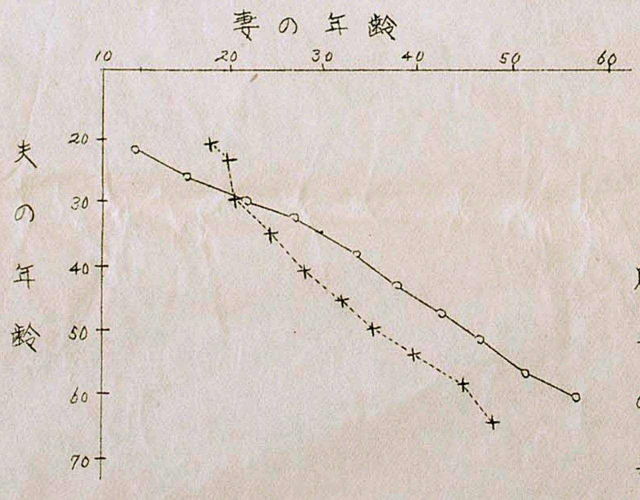
$r = -1$　なるための條件

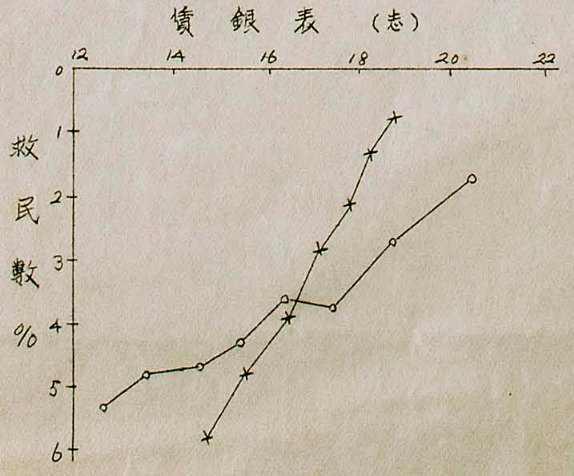
$x_1 = -y_1$ ---------------- $x_n = -y_n$.　對称.

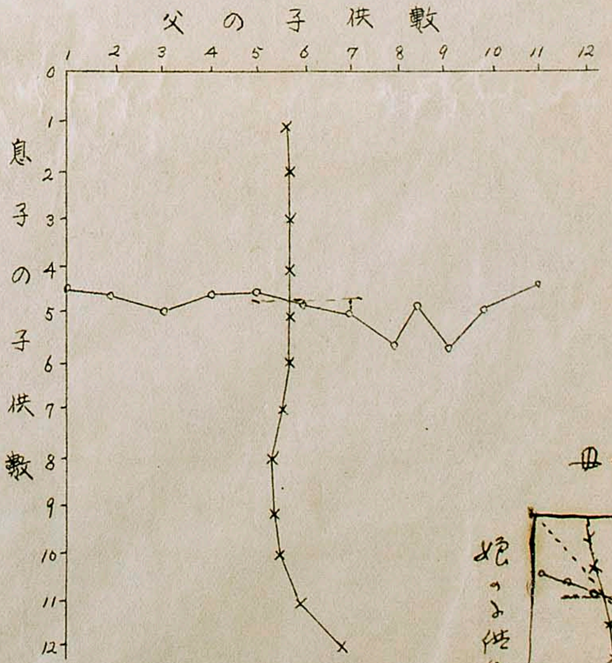
相関圖　(回歸線)

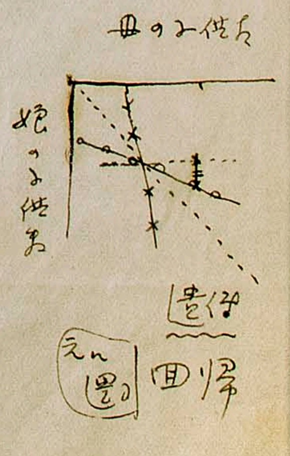


昭和四年内地の結婚

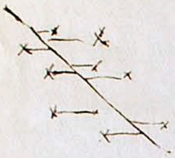


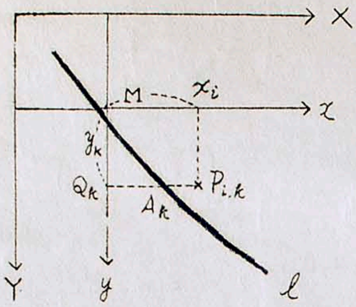






X, Y の標準測定値 x, y を用ひれば　$M_{(x)}=0$, $M_{(Y)}=0$





最小偏差線　lines of minimum deviation

$$l:\quad x=ay+b$$

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

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

$$\equiv S\quad \text{極小}$$

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

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

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

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

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

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

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

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

$$\frac{X-M_{(x)}}{\sigma_{(x)}}=r.\frac{Y-M_{(Y)}}{\sigma_{(Y)}}.\qquad l:\quad X-M_{(x)}=r\frac{\sigma_{(x)}}{\sigma_{(Y)}}\cdot(Y-M_{(Y)}).$$

もう一つの最小偏差線 l' は　　$l':\quad Y-M_{(Y)}=r\dfrac{\sigma_{(Y)}}{\sigma_{(x)}}\cdot(X-M_{(x)})$

体重と身長の場合　　$N=817.$　　$M_{(x)}=52.2^{kg}$,　　$M_{(Y)}=160.0^{cm}$

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

$r=+0.534$

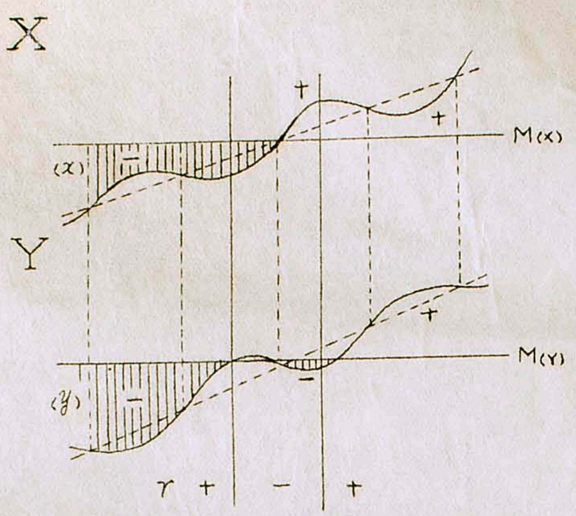
$l:\quad X-52.2^{kg}=+0.536\ (Y-160.0^{cm})$

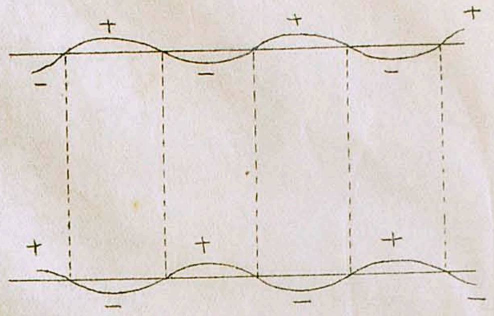
$l':\quad Y-160.0^{cm}=+0.532\ (X-52.2^{kg})$

直線回帰の場合には，二つの最小偏差線は夫々二つの回帰線と一致する。

（證明は自ら試みよ）．

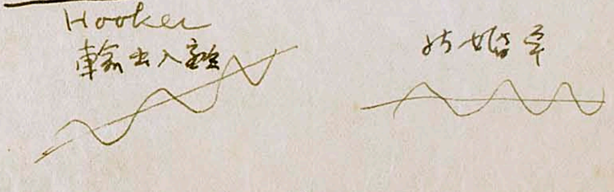
標準測定値を用ひれば





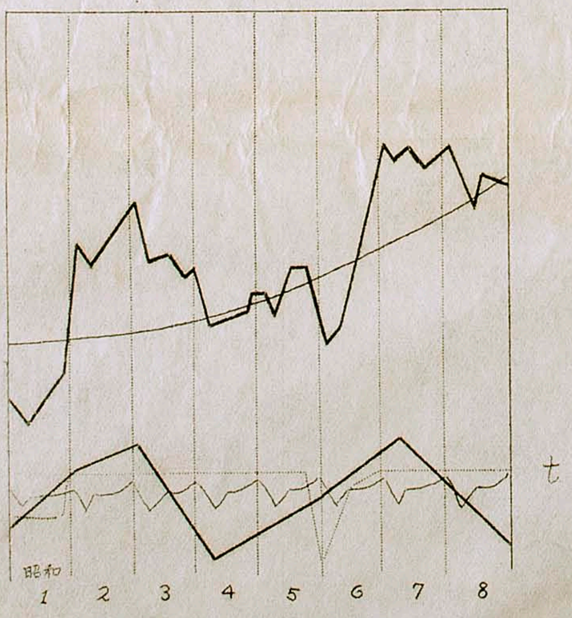
XとYとを其の儘で比較すれば，短期間では負の相關係數を得ることもあれど，長期間では，共に趨勢的に增して居る。結局 $r>0$.

一般的趨勢を除いて考へれば，明かに $r<0$.



時系列の分析

Ⅰ. 趨勢
Ⅱ. 季節的變化
Ⅲ. 循環的變化
Ⅳ. 不規則變化



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.

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

II　幾何平均

$$G = \sqrt[N]{X_1^{f_1} \cdot X_2^{f_2} \cdots\cdots X_n^{f_n}}$$

$$\log G = \frac{f_1 \log X_1 + \cdots\cdots + f_n \log X_n}{N} = \frac{1}{N} \cdot \Sigma (f_k \log X_k)$$

X_k	$\log X_k$	f_k	$f_k \log X_k$
10銭	1.00000	6	6.00000
30	1.47712	8	11.81696
50	1.69897	102	171.59597
70	1.84510	260	479.72600
90	1.95424	162	316.58688
110	2.04139	73	149.02147
130	2.11394	36	76.10184
150	2.17609	30	65.28270
170	2.23045	18	40.14810
190	2.27875	8	18.23000
210	2.32222	1	2.32222
合計 —	———	704	1336.83214

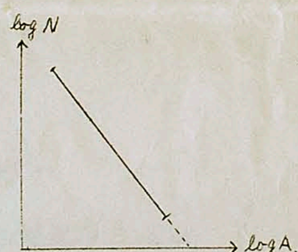
$$\log G = \frac{\Sigma (f_k \log X_k)}{N} = \frac{1336.83214}{704} = 1.90274$$

$$G = 79.936 \text{銭}$$

幾何平均は半對數方眼紙上に畫ける縱線の長さの算術平均である.

絶對的變化と相對的變化。(差と比)。

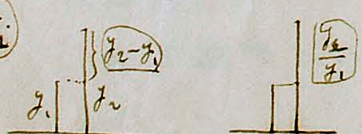
図　図



$y = a \cdot b^x$

$\log y = \log a + x \cdot \log b$

$\log y_2 - \log y_1 = \log \frac{y_2}{y_1}$



對數方眼紙

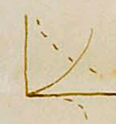
A円以上の所得を得る人數をNとする。

パレト線

$\log N = a + b \cdot \log A.$

汐見三郎外三氏著

國民所得の分配。



$y = a x^b$

$\log y = \log a + b \cdot \log x$

$\eta = \log a + b \cdot \xi$

$A = \pi r^2$