

中大農學院圖書館惠存

統計學教材

羅大元贈

廿八七十五

統計學教材

目次

第一章

第一節

第二節

第三節

第四節

第五節

第六節

第七節

第八節

第九節

第十節

第十一節

第十二節

第十三節

第十四節

緒論

統計學簡史

統計之定義及效用

大量觀察法

抽樣方法

統計法則

統計單位

統計數列

統計之謬誤

小數之精度

有效數字

數之完整

計算誤差

誤差之傳播

統計方法之程序

第二章

第一節

統計表

表之功用

製表規則

第三節
第四節

表之種類
分組法

第三章

統計圖

第一節
第二節
第三節
第四節

圖之功用
製圖之基本法則
統計圖種類
次數分布圖

第四章

統計常數

第一節
第二節
第三節

緒言
總和法
集勢常數

- (一) 算術平均數 (二) 倒數平均數
(三) 幾何平均數 (四) 中位數
(五) 眾數 (六) 分制數

第四節

離勢常數

- (一) 兩極差 (二) 四位分差
(三) 平均差 (四) 標準差
(五) 極差 (六) 各離勢常數之關係
(七) 各離勢常數與次數之關係

第五節
第六節

離勢係數
偏度

其標準誤，并繪製一圖。

求校正之相關比 $K=6$

$$\bar{r}_y^2 = \frac{r_y^2 - \frac{K-3}{N}}{1 - \frac{K-3}{N}} = \frac{(0.832)^2 - \frac{6-3}{24}}{1 - \frac{6-3}{24}} = \frac{0.69193 - \frac{3}{24}}{1 - \frac{3}{24}} = 0.64792$$

$$\bar{r}_y = \sqrt{0.64792} = 0.805$$

求校正後所求出之 r_x 。

習題

求下列各表之相關。

美國俄倫俄 (Ohio) 州商場所售之小麥
之澱粉含量與其麩包体积之相關表

	澱粉含量百分比									合計
	9.0	10.0	11.0	12.0	13.0	14.0	15.0	16.0	17.0	
1600-	2	—	1	—	1	—	—	—	—	4
1700-	—	2	1	—	2	—	—	—	—	5
1800-	6	5	5	2	1	—	—	—	—	19
1900-	5	4	6	7	—	—	—	—	—	22
2000-	2	3	9	10	2	1	—	—	—	27
2100-	1	—	3	3	5	2	1	—	—	15
2200-	—	—	1	2	—	2	1	—	—	6
2300-	—	—	—	—	—	—	—	1	1	2
合計	16	14	26	24	11	5	2	1	1	100

麩包体积 (c.c.)

或用公式(30)

$$b_y = r \frac{\sigma_y}{\sigma_x} = .8 \times \frac{7.40}{6.85} = 0.86$$

其 Y 對於 X 之迴歸方程式為

$$y = b_y x = .86x$$

$$\text{或 } Y - 28.125 = .86(X - 29.167)$$

$$\text{即 } Y = 3.04 + .86X$$

再直接由迴歸方程式亦可求標準誤差

X	$\bar{Y}_x$	f_x	$f_x \bar{Y}_x$	$f_x \bar{Y}_x^2$	$f_x / \bar{Y}_x$
17.5	17.5	2	35.00	612.50	612.50
22.5	21.5	5	107.50	2311.25	2418.75
27.5	28.93	7	202.51	5858.61	5569.03
32.5	28.5	5	142.50	4061.25	4631.25
37.5	35.83	3	107.49	3851.37	4030.88
42.5	40	2	80.00	3200.00	3400.00
合計		24	675.00	19894.98	20662.41

$$a = 3.04, b_y = .86, \sum f_x \bar{Y}_x = 675, \sum f_x \bar{Y}_x^2 = 19894.98$$

$$\sum f_x X \bar{Y}_x = 20662.41, \text{ 代入公式(51), 得}$$

$$S_y^2 = \frac{\sum f_x \bar{Y}_x^2 - a \sum f_x \bar{Y}_x - b_y \sum f_x X \bar{Y}_x}{N}$$

$$= \frac{19894.98 - 3.04 \times 675 - .86 \times 20662.41}{24}$$

$$= 3.054$$

$$S_y = \sqrt{3.054} = 1.74$$

學者欲估計 Y 對於 X 之迴歸係數，迴歸方程式及

$$\therefore \sigma_{ay} = \sqrt{16.87} = 4.1$$

再用序列标准差计算相关比如下：

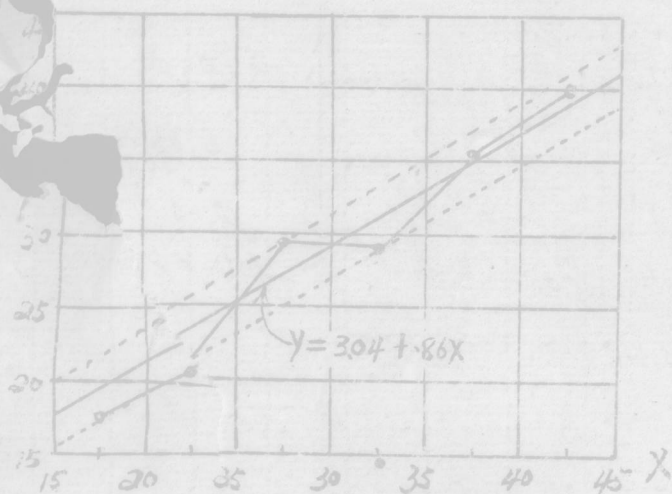
$$\eta_y^2 = 1 - \sigma_{ay}^2 / \sigma_y^2 = 1 - 16.87 / (7.4)^2 = 0.69193$$

$$\therefore \eta_y = \sqrt{0.69193} = 0.832$$

若有试计标 X 之序列标准差 (标) 平均值及其标准差并计标 X 之序列标准差及其标准差，再用序列标准差计标 η_x 。

计标回归系数及回归方程式为次：

x	17.5	22.5	27.5	32.5	37.5	42.5
$\bar{y}_x$	17.5	21.5	28.93	28.5	35.83	40
f_x	2	5	7	5	3	2



将 $x, \bar{y}_x$ 之各对变值绘成散布图，为直线趋势，故可配成合一直线，用简标法公式 (38) 计标 Y 对 X 之回归系数为

$$b_y = \frac{\sum y f_x - N \bar{x} \bar{y}}{\sum x^2 f_x - N \bar{x}^2}$$

$$= \frac{40 - 24 \left(\frac{1}{3}\right) \left(\frac{8}{5}\right)}{48 - 24 \left(\frac{1}{3}\right)^2} \frac{5}{5} = .86$$

$$\therefore \eta_y = \frac{\sigma_{\bar{y}_x}}{\sigma_y} = \frac{6.159}{7.4} = 0.832$$

计算序列标准差及其总标准差

$$\sigma_1^2 = \frac{\sum f_1 (x_1 - \bar{x}_1)^2}{f_1} = \frac{0}{2} = 0, \quad \sigma_2^2 = \frac{\sum f_2 (x_2 - \bar{x}_2)^2}{f_2} = \frac{70}{5} = 14,$$

同理 $\sigma_3^2 = \frac{1}{7}(35.7143) = 5.12, \quad \sigma_4^2 = \frac{270}{5} = 54.$

$$\sigma_5^2 = \frac{16.6667}{3} = 5.5556, \quad \sigma_6^2 = \frac{12.5}{2} = 6.25.$$

组别	$x - \bar{x}$	$(x - \bar{x})^2$	f	$f(x - \bar{x})^2$	$\sum f(x - \bar{x})^2$
1	0	0	2	0	0
2	-4	16	2	32	
	1	1	2	2	
	6	36	1	36	70
3	-1.43	2.0449	5	10.2245	
	3.57	12.7449	2	25.4898	35.7143
4	-11	121	1	121	
		36	1	36	
	4	16	2	32	
	9	81	1	81	270
	-1.33	1.0889	1	1.0889	
	1.67	2.1889	2	5.5778	16.6667
6	-2.5	6.25	1	6.25	
	2.5	6.25	1	6.25	12.5
合 计			24	404.8810	404.8810

$$\begin{aligned} \text{但 } N\sigma_{ay}^2 &= f_1\sigma_1^2 + f_2\sigma_2^2 + \dots + f_6\sigma_6^2 \\ &= \sum f_1(x_1 - \bar{x}_1)^2 + \sum f_2(x_2 - \bar{x}_2)^2 + \dots + \sum f_6(x_6 - \bar{x}_6)^2 \\ &= \sum \sum f(x - \bar{x})^2 = 404.8810. \end{aligned}$$

$$\therefore \sigma_{ay}^2 = \frac{\sum \sum f(x - \bar{x})^2}{N} = \frac{404.881}{24} = 16.87$$

$$= \sqrt{\frac{36.80}{24} - \left(\frac{3}{24}\right)^2} / 1.48$$

$$= \sqrt{1.5177} / 1.48 = 0.832.$$

學者試求其X對Y之相關比。

計標序列平均數及序列平均數之加權標準差。

$\bar{y}_x$	f_x	$\bar{y}_x - \bar{y}$	$(\bar{y}_x - \bar{y})^2$	$f_x(\bar{y}_x - \bar{y})^2$
17.5	2	-10.63	112.9969	225.9938
21.5	5	-6.63	43.9569	219.7845
28.93	7	0.80	0.6400	4.2800
28.5	5	0.37	0.1369	0.6845
35.83	3	7.70	59.2900	177.8700
40	2	11.87	140.8969	281.7938
合計	24			910.6066

求得各行平均數列於第一欄直接以冗長法求得

$$\sigma_{\bar{y}_x} = \sqrt{\frac{\sum f_x(\bar{y}_x - \bar{y})^2}{N}} = \sqrt{\frac{910.6066}{24}} = 6.159$$

在組距相等之相關表，可以簡標法求之如下：

$$\sigma_{\bar{y}_x} = \sigma \sqrt{\frac{\sum (S_x^2 / f_x)}{N} - \left(\frac{\sum S_x}{N}\right)^2} = 5 \sqrt{\frac{26.80}{24} - \left(\frac{3}{24}\right)^2} = 5 \sqrt{1.5177} = 6.159.$$

例 3. 求前 24 对夫妇年龄相关表之各种相关数值

夫 年

	X	17.5	22.5	27.5	32.5	37.5	42.5	f	yf _y	y ² f _y	S _y	y'S _y	S _y ²	S _y ² /f _y	各行平均数	
妻 年	Y	-2	-1	0	1	2	3	f _x	yf _y	y ² f _y	S _y	y'S _y	S _y ²	S _y ² /f _y	$\bar{y}$	$\bar{x}$
17.5	-2	2	2		1			5	-10	20	-5	10	25	5	-1.00	22.5
22.5	-1		2		1			3	-3	3	-1	1	1	33	-1.33	25.5
27.5	0			1	5			6	0	0	-1	0	1	17	-1.17	26.7
32.5	1			2	2	1		5	5	5	4	4	16	3.2	1.80	31.5
37.5	2				1	2	1	4	8	16	8	16	64	16	2.00	37.5
42.5	3						1	1	3	9	3	9	9	9	3.00	42.5
	Σf_x	2	5	7	5	3	2	24	3	53	8	20		33.70		
	$\Sigma x'f_x$	-4	-5	0	5	6	6	8								
	$\Sigma x'^2f_x$	8	5	0	5	12	18	48								
	ΣS_x	-4	-6	2	1	5	5	3								
	$\Sigma x'S_x$	8	6	0	1	10	15	40								
	ΣS_x^2	16	36	4	1	25	25									
	$\frac{\Sigma S_x^2}{\Sigma f_x}$	8	7.2	5.7	2	8.33	12.5	36.70								
各行平均数	$\bar{y}_x$	-2	-1.2	2.9	2	1.67	2.5									
均数	$\bar{y}$	17.5	21.5	26.73	28.5	35.83	42.0									

前已求得 $\bar{x} = \frac{1}{3}$, $\bar{y} = \frac{1}{8}$

$\bar{x} = 29.167$,

$\bar{y} = 28.125$,

$\sigma_x = 1.37$, $\sigma_y = 1.48$,

$\sigma_x = 6.85$,

$\sigma_y = 7.40$,

$r = .80$

计算相关系数次:

$$r_y = \sqrt{\frac{\Sigma(S_x^2/f_x)}{N} - \left(\frac{\Sigma S_x}{N}\right)^2} / \sigma_y$$

$$\therefore r_y = \sqrt{0.9011} = 0.949$$

又直接由回归方程式可求得如下：

$$a = 1.56, b = -0.088, c = 0.00144, \sum y = 68.5, \sum xy = 5985.6,$$

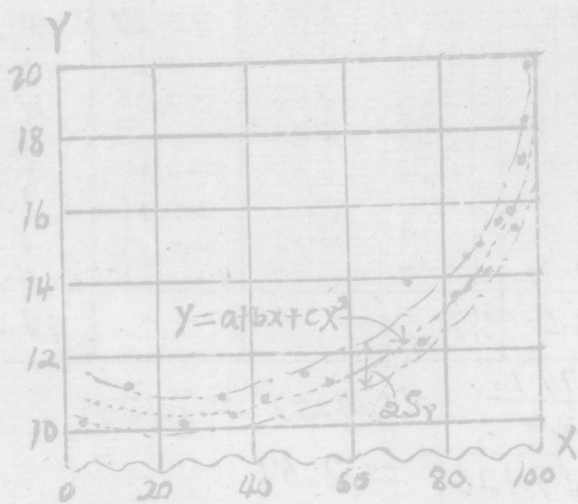
$$\sum x^2 y = 542584.6, N = 20, \bar{x} = 3.425, \sum y^2 = 377.15 \quad \text{代入式(5)}$$

$$\begin{aligned} r_y^2 &= (a \sum y + b \sum xy + c \sum x^2 y - N \bar{y}^2) / (\sum y^2 - N \bar{y}^2) \\ &= \frac{1.56 \times 68.5 - 0.088 \times 5985.6 + 0.00144 \times 542584.6 - 20(3.425)^2}{377.15 - 20 \times (3.425)^2} \\ &= \frac{126.82}{142.54} = 0.8897 \end{aligned}$$

$$\therefore r_y = \sqrt{0.8897} = 0.943$$

$$\begin{aligned} r_y^2 &= 1 - (1 - r_y^2) \frac{N-1}{N-m} \quad m=3 \\ &= 1 - (1 - 0.943^2) \times \frac{23}{21} = 0.8792 \end{aligned}$$

$$\therefore r_y = 0.937$$



觀上圖各觀測點皆密集於迴歸線，落於其兩旁者一個標準誤之範圍以內者，超過 68% 以上，可知其有顯著之相關，故所求之相關指數在 0.9 以上也。

$$X=75, Y=11.56-0.088(75)+0.00144(75^2)=13.06,$$

照樣求出各估計值 $\hat{y}$ ，列於下表，求剩餘標準誤及相關指數，並劃出迴歸線及標準誤之範圍為次：

X	Y	y^2	$\hat{y}$	$y-\hat{y}$	Z^2
6	0.3	0.09	1.08	-0.78	0.6084
75	2.2	4.84	3.06	-0.86	0.7396
87	4.5	20.25	4.80	-0.30	0.0900
55	1.1	1.21	1.08	+0.02	0.0004
34	0.9	0.81	0.23	+0.67	0.4489
98	8.1	65.61	6.77	+1.33	1.7689
91	4.0	16.00	5.48	-1.48	2.1904
45	0.8	0.64	0.52	+0.28	0.0784
51	1.2	1.44	0.82	+0.38	0.1444
17	1.0	1.00	0.48	+0.52	0.2704
36	0.2	0.04	0.26	-0.06	0.0036
97	7.0	49.00	6.57	+0.43	0.1849
74	3.8	14.44	2.93	+0.87	0.7569
24	0.1	0.01	0.28	-0.18	0.0324
85	4.4	19.36	4.48	-0.08	0.0064
96	5.8	33.64	6.38	-0.58	0.3364
82	5.6	31.36	5.65	-0.05	0.0025
94	5.0	25.00	6.01	-1.01	1.0201
84	3.3	10.89	4.33	-1.03	1.0609
99	9.0	81.00	6.96	+2.04	4.1616
1340	68.5	377.15			14.0975

$$\sigma_y^2 = \frac{377.15}{20} - (3.425)^2 = 7.1269$$

$$\sigma_y = \sqrt{7.1269} = 2.67$$

$$S_y^2 = \frac{14.0975}{20} = .704875$$

$$S_y = \sqrt{.704875} = .84$$

$$R^2 = 1 - S_y^2 / \sigma_y^2 = 1 - \frac{.704875}{7.1269} = 0.9011$$

$$\sum x^2 = \sum X^2 - n\bar{X}^2 = 107566 - 20(67)^2 = 17786.$$

$$\begin{aligned}\sum Xu &= \sum XU - n\bar{X}\bar{U} = 9259238 - 20(67)(5378.3) \\ &= 2052316.\end{aligned}$$

$$\begin{aligned}\sum u^2 &= \sum U^2 - n\bar{U}^2 = 824403226 - 20(5378.3)^2 \\ &= 245881008.\end{aligned}$$

$$\begin{aligned}\sum xy &= \sum XY - n\bar{X}\bar{Y} = 5985.6 - 20(67)(3.425) \\ &= 1696.1.\end{aligned}$$

$$\begin{aligned}\sum uy &= \sum UY - n\bar{U}\bar{Y} = 542584.6 - 20(5378.3)(3.425) \\ &= 174171.05.\end{aligned}$$

将上各值代入(60)式得

$$17786b + 2052316c = 1396.1$$

$$2052316b + 245881008c = 174171.05$$

解之得

$$b = -0.0879, \quad c = 0.001442.$$

$$\begin{aligned}a &= \bar{Y} - b\bar{X} - c\bar{U} = 3.425 - (-0.0879)(67) \\ &\quad - (0.001442)(5378.3) = +1.56.\end{aligned}$$

代入(57)式，但因计算时，Y已减去10，故得翅力
缘方程式为

$$Y - 10 = 1.56 - 0.088X + 0.00144X^2.$$

$$\text{即 } Y = 11.56 - 0.088X + 0.00144X^2$$

$$\text{当 } X=6, \quad Y = 11.56 - 0.088(6) + 0.00144(6^2) = 11.08.$$

校正各數 $N=7$ $m=2$

$$S_y = 3.5, \quad r = -0.5, \quad p_y = 5, \quad b_y = -0.758.$$

$$\bar{S}_y = S_y \sqrt{\frac{N}{N-m}} = 3.5 \sqrt{\frac{7}{5}} = 4.13$$

$$\bar{r} = -\sqrt{1 - (1-r^2) \frac{N-1}{N-2}} = -\sqrt{1 - [1 - (-0.5)^2] \frac{6}{5}}$$

$$= -\sqrt{1 - 0.75 \times 1.2} = -\sqrt{0.1} = -0.32.$$

$$\bar{p}_y = \sqrt{1 - (1-p_y^2) \frac{N-1}{N-2}} = \sqrt{1 - (1-5^2) \frac{6}{5}} = 0.32$$

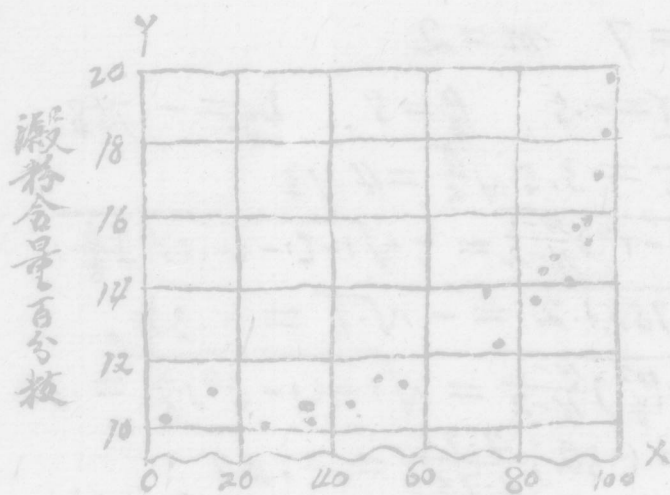
$$\bar{b}_y = \bar{r} \frac{\sigma_y}{\sigma_x} = (-0.32) \frac{4.11}{2.73} = -0.54.$$

學者就就本題求 X 對於 Y 之迴歸線，並求 S_x ，

b_x ， p_x 各值，且校正之。

例2 試求下列各樣本中玻璃狀麥粒數目之比例及其與
含水量之相關。

樣本數	麥粒含水量	玻璃狀麥粒 比例百分數
1	10.3	6
2	12.2	75
3	14.5	87
4	11.1	55
5	10.9	34
6	18.1	98
7	14.0	91
8	10.8	45
9	11.4	51
10	11.0	17
11	10.2	36
12	17.0	97
13	13.8	74
14	10.1	24
15	14.4	85
16	15.8	96
17	15.6	92
18	15.0	94
19	13.3	84
20	19.0	99



玻璃狀麥粒百分數

將上列資料繪圖，其 Y 對於 X 之關係顯然有為二級拋物線之趨勢，今設 Y 對於 X 之關係式為

$$Y = a + bX + cX^2 \quad \text{----- (57)}$$

以 U 表 X^2 ，可變為 $Y = a + bX + cU$ ----- (58)

則用最小二乘法配合點之區帶方程式，為

$$\left. \begin{aligned} na + b\sum X + c\sum U &= \sum Y \\ a\sum X + b\sum X^2 + c\sum UX &= \sum XY \\ a\sum U + b\sum UX + c\sum U^2 &= \sum UY \end{aligned} \right\} \text{----- (59)}$$

若改以平均數為原点計標，則上式變為

$$\left. \begin{aligned} b\sum x^2 + c\sum xu &= \sum xy \\ b\sum xu + c\sum u^2 &= \sum uy \end{aligned} \right\} \text{----- (60)}$$

而 $a = \bar{Y} - b\bar{X} - c\bar{U}$

為簡便計上之各係數可用下各式求之。

$$\sigma_x = \sqrt{\frac{\sum X^2}{N} - \left(\frac{\sum X}{N}\right)^2} = \sqrt{\frac{23590}{7} - \left(\frac{406}{7}\right)^2} = \sqrt{7.43} = 2.73$$

$$\sigma_y = \sqrt{\frac{\sum Y^2}{N} - \left(\frac{\sum Y}{N}\right)^2} = \sqrt{\frac{7544}{7} - \left(\frac{228}{7}\right)^2} = \sqrt{16.9} = 4.11$$

$$p = \frac{1}{N} \sum (X - \bar{X})(Y - \bar{Y}) = \frac{\sum XY}{N} - \bar{X} \cdot \bar{Y} = \frac{13184}{7} - 58 \times 32.57$$

$$= 1883.43 - 1889.06 = -5.63$$

$$r = \frac{p}{\sigma_x \sigma_y} = \frac{-5.63}{2.73 \times 4.11} = -0.52$$

$$b_y = \frac{N \sum XY - \sum X \sum Y}{N \sum X^2 - (\sum X)^2} = \frac{\frac{1}{N} \sum XY - \bar{X} \cdot \bar{Y}}{\frac{\sum X^2}{N} - \bar{X}^2} = \frac{p}{\sigma_x^2} = \frac{-5.63}{(2.73)^2}$$

$$= \frac{-5.63}{7.43} = -0.758$$

用公式 $b_y = r \frac{\sigma_y}{\sigma_x}$ ，亦为计算此次：

$$b_y = -0.52 \times \frac{4.11}{2.73} = -0.782 \text{ (相差 } 0.0248 \text{ 由四舍五入所致)}$$

$$Y = \bar{Y} + b_y(X - \bar{X}) = 32.57 - 0.758(X - 58)$$

即回归方程式为 $Y = 76.534 - 0.758X$ 。

当 $X = 55$ ，其估计之 $\tilde{Y} = 76.534 - 0.758 \times 55 = 34.84$ 。同样求得在 Y 之估计值，列于上表第六栏。

求剩餘标准误差

$$S_y^2 = \frac{\sum (Y - \tilde{Y})^2}{N} = \frac{\sum (2)^2}{7} = \frac{86.9923}{7} = 12.42747$$

$$S_y = \sqrt{12.42747} = 3.5$$

又用公式(51) 直接由回归方程式求得此次：

$$S_y^2 = \frac{1}{N} [\sum Y^2 - a \sum Y - b \sum XY]$$

$$(a = 76.534, b = -7.58)$$

$$= \frac{1}{7} [7544 - 76.534 \times 228 - (-7.58) \times 13184]$$

$$= 12.53143,$$

$$S_y = \sqrt{12.53143} = 3.5.$$

求相关指数:

$$r_y^2 = 1 - S_y^2 / \sigma_y^2 = 1 - \frac{12.53033}{16.9} = 0.2586,$$

$$r_y = \sqrt{0.2586} = 0.5.$$

又可直接由回归方程式用公式(53)求得之。

$$r_y^2 = [a \sum y + b \sum xy - N \bar{y}^2] / [\sum y^2 - N \bar{y}^2]$$

$$= \frac{76.534 \times 228 + (-7.58) \times 13184 - 7 \times (32.57)^2}{7544 - 7 \times (32.57)^2}$$

$$= \frac{30.6457}{118.3657} = 0.2589,$$

$$r_y = \sqrt{0.2589} = 0.5.$$

此时因係直线相关，故 r 与 r_y 完全相等，惟 r 无负号耳。

再从估计值标准差，用公式(48)可求得相关系数 r 为次：

$$\sigma_{\hat{y}} = \sqrt{\frac{\sum (\hat{y} - \bar{y})^2}{N}} = \sqrt{\frac{\sum \hat{y}^2}{N}} = \sqrt{\frac{30.9612}{7}} = 2.1$$

$$\sigma_y = 4.11 \quad \therefore r = \sigma_{\hat{y}} / \sigma_y = 2.1 / 4.11 = 0.5$$

由此法求得之相关系数，亦为相关指数仅得绝对值，而无性质符号之表现。

$$\begin{aligned}
 \therefore \bar{r}_y^2 &= 1 - \frac{S_{y \cdot f(x)}^2}{\bar{\sigma}_y^2} = 1 - \frac{S_{y \cdot f(x)}^2 \frac{N}{N-1}}{\sigma_y^2 \frac{N}{N-1}} \\
 &= 1 - \frac{S_{y \cdot f(x)}^2}{\sigma_y^2} \frac{N-1}{N-m} \\
 &= 1 - [1 - r_y^2] \frac{N-1}{N-m} \dots \dots \dots (60)
 \end{aligned}$$

(3)、相关比之校正：

前已证明任意增加相关表之行数或列数，可使使相关比增至1。且相关表之廻为线属曲线，若行数或列数至少在3以上，故曲线由行数或列数所虚拟的相关比之值，可用下式减低而校正之。

$$\bar{r}^2 = \frac{r^2 - \frac{K-3}{N}}{1 - \frac{K-3}{N}} \dots \dots \dots (61)$$

K表行数或列数。

(4)、廻为直线之校正(直线)：

$$\begin{aligned}
 \bar{r}_y^2 &= \bar{r}^2 \frac{\bar{\sigma}_y^2}{\bar{\sigma}_x^2} = \left[1 - (1 - r^2) \frac{N-1}{N-2} \right] \frac{\sigma_y^2 \frac{N}{N-1}}{\sigma_x^2 \frac{N}{N-1}} \\
 &= \left[1 - (1 - r^2) \frac{N-1}{N-2} \right] \frac{\sigma_y^2}{\sigma_x^2} \dots \dots \dots (62)
 \end{aligned}$$

有人用苗米产量与米价表求其相关，