

大正四年六月二十一日第三種郵便物認可 (毎月一回一日発行)

京都市帝國大學經濟學會

經濟論叢

第三卷 第四號

昭和五年四月一日發行

論叢

家屋税の課税標準 法學博士 神戸正雄

貨幣數量説について 文學博士 高田保馬

經營學と經濟學 經濟學博士 小島昌太郎

時論

配給組織の合理化と中央市場の單複制 經濟學士 谷口吉彦

說苑

統計學に於ける二つの傾向に就いて 經濟學士 蜷川虎三

ボーレの恐慌理論 經濟學士 靜田均

雜錄

英蘭銀行の職能 經濟學士 有井治

月賦信用の特質 經濟學士 今津正二

カッセの價值論廢止と價格問題の取扱 經濟學士 高森晋

相關係數の意義 經濟學士 益田熊雄

酒税の立替 經濟學博士 沙見三郎

近着外國經濟雜誌主要論題

(禁 轉 載)

相関係数の意義

益田熊雄

一

相関係数は統計學上甚だ重要な地位を占めて居るに拘らず、從來吾國に於ては餘り研究せられなかつた。否寧ろ敬遠せられて居たかの様に思はれる。然し¹⁾二個の變數を取扱ふ場合に、その間の關係を究明せんとするには是非此の相関係數に依らねばならぬのであるから、相関係數を普遍化する目的を以て Mordecai Ezekiel 氏の論文を紹介して見やう。

氏の論文は私と同じ目的よりして、相関係數の意義を平易に説明せんとする所に重點があるから、従つてその用ふる材料は普通相関係數を取扱ふ場合に於けるよりは遙かに簡單なるものである。即ち一九二二—二八年の聯邦の調査に基く屠殺豚の總重量と豚肉の市價との相關係を例として、相関係數の意義を説明して

- 1) Yule: An Introduction to the Theory of Statistics p. 157
- 2) Mordecai Ezekiel: Meaning and Significance of Correlation Coefficients (The Amer. Eco. Review vol. XIX. No.2. Jun. 1929)

居るのである。

二

相關的分析の基本となる數は、回歸係數・測定の標準誤差及相關係數の三者である。此の三者を第一表より計算すれば次の如くである。

相關係數 $r_{xy} = 0.99$

回歸係數 $b_{yx} = 2.087$

測定の標準誤差 $S_e = 0.22 \text{ dollars}$

之等は普通の形によつて計算せられた。即ち

$M_x = X$ の平均 $M_y = Y$ の平均…とすれば

$X - M_x = x$ $Y - M_y = y$

$$\sigma_x^2 = \frac{\sum x^2}{n} = \frac{\sum X^2 - nM_x^2}{n}$$

同様に $\sigma_y^2 = \frac{\sum y^2}{n}$

$$\therefore b_{yx} = \frac{\sum xy}{n\sigma_x^2} \quad 1) \quad r_{xy} = \frac{\sum xy}{n\sigma_x\sigma_y} \quad 2)$$

$$S_e = \sigma_y \sqrt{1 - r_{xy}^2} \quad 3)$$

第一表

十一月一日に 始まる年	屠殺豚の重量 (X)	豚肉の市價 (Y)
1922—23	11.7	7.62
23—24	11.8	7.61
24—25	10.8	10.71
25—26	9.7	12.16
26—27	10.0	10.84
27—28	11.0	9.10

回歸係數が—2.089と云ふ事は、係數が負であるから豚の供給が一ポンド増加すれば豚肉の價格は二・〇九弗下落する事を示して居るのである。之によつて回歸直線の方程式を作れば次の式を得る。

$$Y = a + b_{yx}X$$

$$Y = 32.11 - 2.089X \quad 4)$$

此の式のXに夫々第一表のXの値を代入すれば第二表を得る。

1) 小林新：經濟統計學 P393
 2) 同 P 373
 3) 同 P 395
 4) $a = M_y - b_{yx} M_x$ より計算さる

第二表

年	現實價格 (Y)	供給より測定せられたる價格 (Y')	現實價格の測定價格超過額 (Y-Y')
1922—23	7.62	7.69	- 0.07
23—24	7.61	7.48	+ 0.13
24—25	10.71	10.61	+ 0.10
25—26	12.16	11.87	+ 0.29
26—27	10.84	11.24	- 0.40
27—28	9.10	9.15	- 0.05

第二表の示すが如くに、「現實價格」と「供給より測定せられたる價格」との間には誤差があるが、此の誤差の標準偏差が即ち測定の標準誤差であつて、之が先に述べた如く⁵⁾〇・二弗である。

即ち此の事は豚肉の價格に影響する事情にして一九二二年より二八年まで變化なき限り、年々の豚肉の價格は屠殺豚の重量より〇・二弗の標準誤差を持つ方程式によつて測定せらるる事を示して居るのである。

上に述べたる供給より測定せる價格(Y')の標準偏差($\sigma_{Y'}=1.69$)を現實價格の標準偏差($\sigma_Y=1.70$)にて除す

れば相関係數(0.99)を得る。故に相関係數は測定せる價格の偏差と現實價格の偏差との比を示す事となる。

III

次に氏の力説せんとする所は、相關の標準として相關係數(r)を用ふるよりも、相關係數の平方(r^2)を用ひた方が係數の意味を示すには適して居ると云ふ事である。

先づ之を代數的に説明して見やう。今

$$Z = Y - Y'$$

と置き之等三つの數列の標準偏差を求むれば、次の關係が得られるであらう。

$$\sigma_z^2 = \sigma_y^2 - \sigma_{y'}^2$$

$$\text{or } \sigma_{y'}^2 = \sigma_y^2 + \sigma_z^2$$

$$\text{or } 1 = \frac{\sigma_{y'}^2}{\sigma_y^2} + \frac{\sigma_z^2}{\sigma_y^2}$$

斯くして求められたる $\sigma_{y'}^2/\sigma_y^2$ は相關係數の平方であり、 σ_z^2/σ_y^2 は測定の標準誤差をYの標準偏差で除して

5) $\sigma_{y'}^2 = \frac{\sum_{1/2}^2}{n}$ により $\sigma_{y'}$ を第二表のY-Y'より計算すれば約0.22を得る。

平方したるものである。兩項の和が1である事は此の兩項によつて變化の全部が説明せられる事を意味して居る。此の兩項は共に平方であるから正であり、従つて各項は1より小なる事を示して居るのであるから、その平方根は平方値よりは大である事は明かである。

故に普通なさるるが如くに相關係數 (r_{xy}) の平方根即ち $\sqrt{r_{xy}}$ を以て相關の標準となすならば、兩項の和が1より大となるからして、相關を過大に見積る、(over-estimate) 事となるであらう。故に標準として相關係數 (r) を用ふるよりも、相關係數の平方 (r^2) を用ひた方が係數の意味によりよく適するのである。

次に之を算術的に説明して見やう。今A、B二つの數列は互に零の相關を持ち、CはA、Bの和を示す數列なりとする。

	A	B	C
1	1	4	5
2	2	6	8
3	3	4	7

相關を計算する爲に平均よりの偏差を求めて平方及積を作つて見る。

a	b	c	a ²	b ²	c ²	ab	ac	bc
1	0	1	1	0	1	0	1	0
0	2	2	0	4	4	0	0	4
0	2	2	0	4	4	0	0	4
1	0	1	1	0	1	0	1	0
差			2	8	10	0	2	8
總和			0.71	1.41	1.58			

此の表に於てabの和が0である事はA、Bの間に相關なき事を示して居る。他の相關係數を求めると次の値を得る。

$$1) \quad r_{ac} = 0.45 \quad r_{bc} = 0.89$$

$$\therefore \quad r_{ac}^2 = 0.20 \quad r_{bc}^2 = 0.80$$

故に r_{bc}^2 を相關の標準として採用するならばCの二〇%の變化はAによつて説明せられ、Cの八〇%の變化はBによつて説明せられる。合計一〇〇%であるから、従つてCの變化の全部は兩者によつて説明し得らるる事となる。若し r を探つて見るならば合計

1) 原文には $r_{ab} = 0.45$ $r_{ac} = 0.89$ とあるも誤植なるべし

$$0.45 + 0.89 = 1.34$$

となり過大となる。 $\sqrt{1.34}$ を採れば同様に1.34となり、 $\sqrt{1.34}$ を採れば0.96となりて、過大或は過小となり、従つて之等はCの變化の全部を説明し得ないから、相關の標準としては不適當である以上によつて明かなる如く、相關の標準としては相關係數の平方を用ふる方がより適切であるのである。

(五、三、一三)