

高野岩三郎の統計学講義録(3) :
大正8年度東大講義録の復刻とその考証

メタデータ	言語: ja 出版者: 静岡大学人文社会科学部 公開日: 2013-09-06 キーワード (Ja): キーワード (En): 作成者: 上藤, 一郎 メールアドレス: 所属:
URL	https://doi.org/10.14945/00007436

資料

高野岩三郎の統計学講義録 (3) —大正8年度東大講義録の復刻とその考証—

上 藤 一 郎

1. 資料の復刻 (承前)

第四章 統計的研究

独逸ノ学者ナル Rümelin ハ曰ク、数字ハ黙シテ語ラス。吾人ハ之レヲシテ口ヲ開カシメ得ルニスキスト (Die Zahlen sind stumm, man kann ihnen der Mund Öffnen)。

実ニ此ノ統計的研究ノ目的ハ、此ノ黙シテ語ラサル数字ヲシテ社会現象ニ関スル事実ヲ説明セシムルニアリ。換言スレハ、統計ヲ集メ之ヲ比較シ、ソノ中ノ道理通則ヲ発見シ、更ニソノ因果関係ヲ明カナラシムルコトヲ以テソノ目的トス。最善ノ技術ヲ尽シテ最良ノ統計ヲ作ルコトハ統計家ノ任務ナリ。之ニ対シテ統計学者ノ職務ハ、統計ヲ材料トシテ之レニ加工シ、社会的ノ学理ヲ発見スルニアリ。統計ハ統計的研究ニヨリ初メテ真ニ価値ヲ發揮スル。又ソノ改良進歩ヲ遂クルモノナリ。

統計的研究ハ又統計ノ完備セルヤ否ヤ、ソノ善悪ニヨリ大ニソノ発達ヲ左右サラル。統計家ト統計学者、統計ト統計的研究、コノ両者ハ相並ヒテ進マサルヘカラサルナリ。

統計的研究ニ関シ一般ニ採ルヘキ順序又守ルヘキ注意事項ヲ簡單ニ述ヘン。

第一ニハ統計材料ノ統計トシテノ価値判断ナリ

スヘテ一定ノ社会現象ニ関シ統計的研究ヲ為サシメントスレハ、如何ナル材料カ必要ナルカヲ夫々取扱フ現象ニ関スル専門ノ知識ニヨリテ考ヘ、出来ルタケ完全ニ集ムルコトヲ要スルハ勿論ナリ。之レヲ例ヘハ労働者階級ノ収入ノ状況ヲ調ヘントスルレハ、労働者ノ得ル經常的収入ニ、ナオ臨時的収入、ソノ家族ノ収入並ニ実物ノ収入、更ニ進テハ名義的収入ノ購買力、即チ所謂實際上ノ賃銭ヲハカルニ足ル材料等ノモノヲ必要トスルカ、現今ノ實際ニ於テハ何レノ方面ニモ統計材料充分ニ具ハラス。為メニ研究ヲ全然断念スルカ、不充分ナル程度ニ止メサルヲ得サル場合ノ多キハ、常ニ吾人ノ遺憾トスル所ナリ。サレトモ兎ニ角出来タケ材料ヲ集メシメ之ニ忘ルヘカラサルハ、集メシ材料ノ個々ニツキ詳細ニソノ性質ヲ調ヘ且ツ価値ヲ判断スルコトナリ。

蓋シ全〔一〕種類ノ事項ニ関スル統計ニテモ、之ヲ用ヒラレタル調査方法カ比較的正確ナリ

ヤ否ヤ、又調査方法カ予期通り正確ニ行ハレタルカ否カニヨリ、統計ソノモノノ価値ニモ自ラ大小ノ差ヲ生ス。從テ吾人ハ、之ヲ利用スルニ当リテハソノ出所ヲ定メ価値ヲ判断シ、必要ナル斟酌ヲ加フルコトハ真面目ナル研究ニ怠ルヘカラサルコトナリトス。

第二ノ順序ハ統計材料ニ均一ナル比較ヲ行フコトナリ

換言スレハ、出来タケ同一ノ場所、期間ニ関スル事項又ハ性質ノ同一ナル事項ヲトリテ適切ニソノ比較ヲナスコトヲ努メサルヘカラス。例ヘハ死産ヲ算入セル出産統計ト之レヲ算入セサルモノトヲ比較シ、或ハ又通過貨物ヲ含メルト含マサル外国貿易ノ高トノ比、又ハ予算ト決算トヲ直ニ比スルコト等、何レモ適當ナラス。又地域ノ變動セル都会ノ人口ヲトリ、變動ナキ如ク看做シテ対照シ、春行ヘル營業調査ト冬行ヘルモノトヲ比シ、又繁殖期ニ行ヘル家畜調査トソノ以外ノモノトヲ比スルハ妥當ナラス。大国ノ数字ヲ小国ノ数字ト何等ノ斟酌ナク比較スル如キモ宜シキヲ得サルヤ明カナリ。

第三ノ順序ハ数列ヲ批判シテ其ノ通性ヲ求ムルコトナリ

一体單一ノ数字ハ、仮令ソレカ平均数又ハ比例数ナル場合ニテモ〔一ツ〕ノ單純ナル事實ヲ示スニ止マリ、殆ントソノ他ニ何事ヲモ説明セス。從テ吾人ハ徐々ニ〔ツ〕以上ノ数字ヲ含ム数列ヲトリ比較ヲ行ハサルヘカラス。

但シ、コノ数列ノ中ニ於テ平常ナル又ハ急激ナル變動ニ失スト認メラルル数字アル時ハ、之ニ修正ヲ施シ、又数例中ニ欠陥アルトキハ之ヲ補充シテ、所謂問置法（Interpolation）ヲ行フコトアリ。ソノ方法ニ至リテハ社会統計ノ場合ニハ簡易ナル数字ノ方法ニヨリテ以テ足レリトス。而シテ又絶対数ハ、上述ノ如ク基本ノ数字トシテ重要ニシテ且ツ比例数及平均数ノ正当ナル判断上、常ニヨルヘキ最後ノ標準トナレトモ、到底比例数、殊ニ平均数ノ便利ナルニハ如カス。即チ吾人ハ、統計的研究上主トシテ利用スル所ハ比例数、及ヒ殊ニ平均数ノ数列ナリ。

（註）

初ノ三項ヲ平均シ下段ノ2ヲ得。次ニ3, 2, 7ノ三項ヲ平均シテ4ヲ得。カク両側ノ数字ヲトリ、算術的平均ヲ行ヒタルナリ。

1, 3, 2, 7, 11, 12, 14

1, 2, 4, 7, 10, 12, 14

五年毎ノ人口調査ニテ断続的ニ行ヒタル統計以上ニ各同シ数字ヲ得タキ時ハ、各年平等ノ増加率ト見ナシ複利率ノ計算ニヨリ之ヲ求メテ知ルコトヲ得。但シ極メテ正確ト云フヲ得ス。機械的ナリ。

$$\text{複利率} = \sqrt[5]{\frac{\text{元利計算}}{\text{元金}}} - 1$$

推定ニ〔基〕ク問題ナリ。underlineアハ数字ノ欠クル場合、之ヲ補ハントス。A地方ノモノ欠クル時ハ両側及ヒ平均数ヲ見テ15ト推定スルナリ。

地方 \ 年	1860	_62	_64	_66	_70
A	12.6	15.0	15.0	15.0	15.0
B	18.0	19.0	19.0	20.0	20.0
C	10.0	11.0	11.0	12.0	12.0
平均	13.6	15.0	15.0	15.7	15.7

爰ニ吾人ハ、一定ノ数列ヲシラベテ時ト処トニヨル相違及変動ニ関シテソノ通性ヲ求メサルヘカラス。殊ニソノ数列カ果シテ安定的ナル趣ヲ呈スルカ、或ハ増加ノ傾向ナキカ、又ハ減少ノ傾キナキカ、又ハ回帰の変動ノ趣ナキカ、或ハ全ク不規則的ノ変動ヲ示スカ、之ラノ特性ヲ検討セサルヘカラス。之ヲ検討スルニ当リテ利用セラルル方法種々アレトモ、社会現象ノ取扱上常ニ用ヒラルル簡単ナル方法ヲ挙クレハ次ノ如シ。

其一ハ、全数列ノ平均ニ対シ各数ノ変動度ヲ算スル方法ナリ。即チ、所謂平均偏差、又ハ標準偏差ヲ算出シテ行ク方法ナリ。

其二ハ、数列中ノ最大数、最小数又ハ任意ニ一ツノ数ヲトリテ之ヲ基準数トシ、又ハ全数列ノ平均ヲ基準数トシテ、之ニ対スル各数ノ比例ヲ計算スル方法ナリ。

其三ハ、全数列ヲ分テ数個ノ分階ヲ設ケ、ソノ各分階ニ対シテ平均数ヲ算出スル方法ナリ。

(註)

人口十万ニ付キ犯罪

A	B
5	8
10	10
12	12
4	9
19	11
10	10
平均値	

平均偏差 average deviation

$$A) \frac{5 + 0 + 2 + 6 + 9}{5} = \frac{22}{5} = 4.4 \quad \text{平均値ニ対スル割合} \quad 44\%$$

変動係数 coefficient of variation

$$B) \frac{2 + 0 + 2 + 1 + 1}{5} = \frac{6}{5} = 1.2 \quad \text{変動係数} \quad 12\%^{(4)}$$

⁽⁴⁾ 計算式から明らかなように、ここで述べられている変動係数というのは、今日よく知られた変動係数のことではない。これは上段と同じく平均偏差の平均値に対する比率を指している。

$$\text{平均偏差} = \frac{d+d+\cdots d}{n} = \frac{\sum d}{n}$$

d = 平均値ノ偏差, Σ = 和ノ符号, n ハ項数

$$\sqrt{\frac{(\pm d_1)^2 + (\pm d_2)^2 + \cdots + (\pm d_n)^2}{n}} = \sqrt{\frac{\sum \pm d^2}{n}} = \sqrt{\text{average of squares of deviation}}$$

A. $\sqrt{\frac{(-5)^2 + 0^2 + 2^2 + (-6)^2 + 9^2}{5}} = \sqrt{\frac{146}{5}} = 5.4$ 変動係数 54%

B. $\sqrt{\frac{(-2)^2 + 0^2 + 2^2 + (-1)^2 + 1^2}{5}} = \sqrt{\frac{10}{5}} = 1.4$ 変動係数 14%

	最大数ヲ100トス	最小数ヲ100トス	$\frac{a_1 + a_2 + \cdots + a_n}{n} = 100$
a_1	100	102	96
a_2	99	103	95
a_3	98	100	101
a_4	96	101	99
a_5	95	105	102
⋮	⋮	⋮	⋮
a_n	97	106	99

終ニ此ノ点ニ関連シテ注意ヲ促シタキハ、仮令ニ以上ノ数列カ全様ノ安定度又ハ増減ノ傾向ヲ示セリト云ヒテ、直チニ之レヲ同一視スヘカラサルコトナリ。例ヘハ甲乙二国ノ人口カ共ニ同様ノ増加率ヲ示スモ、甲国ハ主トシテ出生死亡ノ増減ニ原因シ、乙国ハ来住超過ニ原因ストセハ、決シテ二者ノ増加ハ同シカラサルカ如シ。尚ホ我々ノ陥リ易キ誤リトシテ指摘スヘキハ、比例数ノ増減ヲ以テ直チニ絶対数ノ増減ト判断スルコト之ナリ。云フ迄モナク比例数ノ増減ハ絶対数ノ増減ト常ニ相伴フモノニアラス。従テ例ヘハ比例数ノ減少ナルコトハ、絶対数ノ減少若クハ増加ヨリ生シ得ヘク、此ノ兩値ノ場合ハ、決シテ同一ニ論スルヲ得ス。要スルニ、吾人ハ数字ノ内容ソノ組立ヲ精査シテ通性ノ判断ヲ誤マラサル様セサルヘカラス。

(註)

	A年	全人口 百中%	B年	%	C年	%
都会人口	480	60	630	70	770	70
田舎人口	320	40	270	30	330	30
	800	100	900		1,100	

人口千人ニ付出生

22	}	27	}	30	}	34	}	32
30	}	25	}	27	}	30	}	
23	}	28	}	35	}	32	}	
23	}	28	}	26	}	25	}	25
22	}	25	}	26	}	22	}	
30	}	24	}	28	}	21	}	

第四ノ順序ハ、数列中ニ見出シタル通性ノ原因ニサカノホリテ因果関係ノ説明ニ資スルコトナリ。今、社会現象ニ関シテ、コノ目的ニ向テ用ヒラルヘキ統計の方法ヲ略説スレハ、ソノ方法ハ大別シテ二種トナスヲ得。一ハ之ヲ数列分類法ト名クルヲ得ヘク。他ノ一ハ之ヲ数列対比法ト云フヲ得ヘシ。

数列分類法トハ、一定ノ性質的又ハ分量的ノ標準、又若クハ性質的分量的ノ兩者ヲ混合シタル標準ニヨリテ統計材料ヲ分類シ、以テ因果関係ヲ立タラントスル方法ナリ。

此ノ方法ニ於テ一定ノ標準事項ヲ選択スルハ、畢竟コノ事項ト研究对象タル社会現象トノ間ニ、経験上又ハ推理上何等カノ関係アリトノ想像ニ基クモノニシテ、從テ此ノ方法ニヨル研究ノ結果ハ予期セラレタル関係ヲ肯定又ハ否定スルコトナヲ得ルナリ。更ニ又ソノ関係ノ状況程度ヲ知ルコトヲ得ルナリ。但シ此方法ヲ採用スルカ為メニハ統計材料カ既ニ分類ヲ可能ナラシムル如ク編成セラレ居ルコトヲ必要トス。若シ統計材料ノ編集カ不十分ナル程度ニ於テナサレタルニスキサ場合ハ、此方法ノ応用モ亦勢ヒ不十分ナルヲ免カレス。

(註)

(1) 質的分類ニ依ル方法

全人口千ニ付死亡	32
各職業人口千ニ付	各職業死亡
農業主	27

大工	28
工場職工	33
飲食店給仕	35
日雇人足	38

(2) 量的分類ニ依ル法

平均一戸家賃額	人口千ニ付出生
5 円以下ノ地方	45
5 - 10 円ノ地方	40
10 - 20 円ノ地方	35
25 円以上	32

(3) 質量的分類ニ依ル方法

各年齢級及配偶関係	人口一万ニ付犯罪		
	20 - 30 歳	30 - 40 歳	40 - 50 歳
未婚者	70	65	58
既婚者	61	58	52

災害百中

日曜日	16
月曜日	18
火曜日	16
水曜日	12
木曜日	13
金曜日	14
土曜日	13

数列対比法トハ、二以上ノ数列ヲ対比シテソノ間ノ因果關係ヲ求メントスル方法ナリ。前述ノ数列分類法トハ、一定ノ社会現象ニ関スル統計材料ノ分類ニヨリテ成〔ル〕標準事項ト該社会現象トノ間ニ存スル直接又ハ間接ノ因果關係ヲ明カナラシメント欲スルモノニシテ、從テ一種類ノ統計材料アルヲ以テ足レリトス。之ニ反シテ数列対比法ハ、常ニ二種類以上ノ数列アルコトヲ必要トシ、而モコノ二種以上ノ数列ノ対照ニヨリテ、〔初メテ〕ソノ間ニ存スル關係ヲ発見セントス

ルモノナリ。但シ此ノ場合ニモ数列ノ対比ト云フコトハ、畢竟吾人ノ経験又ハ推理ノ上ヨリシテソノ間ニ何等カノ因果関係アルヘキ予想ニ基キテ為サルモノナリ。又各数列ハ、年次別ハ場合所別等ノ標準ニヨリテ区分セラレ、且ツ数列ノ比較上ニ此ノ区分標準ナルモノ用ヒラルルモ、此ノ方法ノ目的トスル処ハ此区分標準ト数列トノ関係ヲ見ルニアラス。数列ト数列トノ関係ヲ求めて行クニ在ルナリ。而シテ、対比セラルル数列間ニ於テ一方ノ変化ニ從ヒテ他方モ亦変化シ行クトキハ、之ヲ称シテ二者間ニ相関関係 (correlation) アリト称ス。且ツ之ノ関係カ平行的ニシテ、即チ一方ノ増減ニ伴ヒテ他方モ亦増減スル時ハ、之ヲ平行関係又ハ正ノ相関関係 (parallelism, positive correlation) ト云ヒ、之ニ反シテ兩者ノ関係カ反抗的、即チ一方ノ増減カ他方ノ増減ト相反スルトキニハ、之ヲ逆行関係又ハ負ノ相関関係 (antagonism, negative correlation) ト云フ。

(註)

	一年平均 米価一石 (円)	人口一万ニ付 財産罪	人口一万ニ付 身体罪	職工組合員百ニ付 失業者
1906	18	15	13	3
1907	22	19	10	2
1908	17	13	15	4
1909	20	18	12	1.5
1910	16	16	14	4
		(+)	(-)	(-)

	人口千ニ付 結婚	破産宣告数
1906	8	1500
1907	9	1200
1908	10	900
1909	7	1600
1910	6	1700
		(-)

数列ノ間ニ相関関係認メラルル場合ニハ、概シテ吾人ハ其数列ニ依リテ表ハルル社会現象ノ間ニ何等カ因果関係ノ伏在スルコトヲ明ニシタリト云フヲ得ヘシ。但シ其ノ何レカ原因的現象ニシテ、何レカ結果的現象ナリヤト云フコトハ各現象ノ性質ニ顧ミテ決スルノ外ナシ。然レトモ、該社会現象ヲ以テ常ニ直ニ原因其ノモノト速断スヘカラス。ソハ蓋シ現象其ノモノハ、単ニ真原因ノ徴候ヲ示スニ止マル場合少カラサレハ也。況ンヤ例ヘハ、前述ノ結婚ト破産宣告数トノ間ノ相関関係ニ於テ一見関係アル如キ二個ノ現象ハ、共ニ原因タラス又結果タラス。原因ト認ムヘキ事

項ハ、其ノ外ニ存在シ、共ニ二者ノ上ニ働キテ、一定ノ相関關係ヲ惹起スルコトアレハ也。又況ンヤ單純ナル相関關係ハ、必スシモ因果の關係ヲ指示スルモノニ非スト云フ場合アレハ也。

宗教ニ依リテ盲者ノ数ヲ調査スルニ、旧教ヲ信スル者ニハ盲者多シ。然ルニ旧教ハ多ク貧者之ヲ信スルカ故ニ、貧民ニ盲目多シト云フコトニナル。

	人口一ニ付 外国貿易額	人口十万中 自殺	外国 貿易額	自殺
1911	a_1	x_1	20円	12
1912	a_2	x_2		
1913	a_3	x_3		
1914	a_4	x_4		
1915	a_5	x_5		
1916	a_6	x_6	25円	8
1917	a_7	x_7		
1918	a_8	x_8		
1919	a_9	x_9		
1920	a_{10}	x_{10}		

	一年平均 米価一石	出生千ニ付 乳児死亡
1911	13	140
1912	13	137
1913	11	116
1914	15	133
1915	12	124
1916	18	165
1917	14	130
1918	17	153
1919	19	162

		米価	死亡
高 価 年	1906	18円	160
	1908		
	1909		
中 価 年	1901	13.75	135
	1902		
	1904 1907		
安 価 年	1903	11.50	120
	1905		

要之種々ノ社会現象ニ関スル数列間ノ相関關係基キテ因果關係ヲ判断スルニ当リテハ、広ク各方面ヨリ事實ヲ研究シ、真因ノ発見ニ注意セサルヘカラス。然レトモ数列間ニ存在スル相関關係ハ、仮令其レカ正ナルモ負ナルモ、一見明瞭ニシテ容易ニ之ヲ発見シ得ルコトハ實際上寧ロ稀ナリトス。

数列中ニ含マルル数字甚タ多キ場合ニハ益々然リ。乃チ本来存在スル数列ニ対シテ直ニ数列対比法ヲ其ノ僣適用スル代リニ、数列中ノ数字ヲ適宜ニ分類シテ対照ヲ試ミ、以テ相関關係ヲ見出スコト行ハル。此ノ方法ノ主ナルモノ三、四ヲ挙クレハ〔次ノ通り〕。

(1)ハ、例へハ五年毎ニ数字ヲ集計スルト云フ如ク、単ニ機械的標準ニヨリテ分類ヲ試ムル方法也。

(2)ハ、例へハ穀物ノ価ト乳児ノ死亡率トノ二ツノ数列ヲ対比スル場合ニ、米価ノ高キ時ト低キ時ト云フカ如キ、一種ノ合理的標準ニヨリテ該数列ヲ分類シ、各分類ノ平均ヲ算出シ之ヲ以テ対比ヲ行フト云フ方法也。

(3)ハ、前述ノ例ニ付テ云へハ、米価ノ高低、乳児死亡率ノ多少アリタル年ト云フカ如キ標準ニ依リテ各自該数列ヲ分類シ、然ル後、各々之ヲ数ノ順位ニ従ヒテ配列シ分配ノ状況ヲ調べ、以テ相関関係ヲ推量スル方法也。

(4)ハ、相関表 (correlation table) ト称スル一種ノ表式ニヨリテ二個以上ノ数列ノ相関関係ヲ測知スル方法ニシテ、之ニ付テ近来独逸ノ Tönnies ノ唱へタルモノ、簡便ナル良法タルヲ失ハサルカ故ニ之ヲ紹介スヘシ。

之ニヨレハ、先ツ比較スヘキ二種ノ数列ヲ取り、各々之ヲ一定ノ標準ニ従ヒ若干ノ階級ニ区分ス。但シ各数列ニ対シテ設ケラルル階級ノ数ハ同一ナルヲ要ス。斯クシテ二種ノ数列ヲ其ノ階級ニヨリテ比較シ、一方ノ数列ノ各級ノ各数カ如何ニ他ノ数列ノ階級ニ分配セラルルカヲ見ルニ当リ、之ヲ碁盤目配列シ図式ヲ以テ表ハスモノトス。

而シテ若シ二種ノ数列ニ依リテ現ハサルル二種ノ現象カ完全ニ平行又ハ逆行ノ関係ニ立ツ時ハ、二数列ノ各階ノ各数ハ碁盤目ニ於テ全然正又ハ負ノ対角線上ノ区画ニ集マル也。即チ、所謂吻合位置之也。又斯クノ如ク吻合セラルルモノ兩者ノ関係カ可ナリ密着ナル時ハ、対角線上並ニ其ノ両側ノ区画、即チ傍触位置ニ集マルヘキ也。吾人ハ之ニ依リテ両現象間ニ存スル正又ハ負ノ相関関係ヲ明ニスルコトヲ得ル也。Tönniesハ更ニ進ミテ此ノ相関関係ノ程度ヲ計ル方法ヲ述ヘタリ⁽⁵⁾。

〔図表1〕

年	穀価	財産罪	身体罪
a_1	I	I	V
a_2	II	II	IV
a_3	I	I	V
a_4	III	III	III
a_5	IV	IV	II
a_6	II	II	IV
a_7	V	V	I
a_8	IV	IV	II

⁽⁵⁾ 明らかに、ここで述べられている相関係数とは、今日よく知られたK. Pearsonによる積率相関係数のことではない。しかし以下の本文では、関係する図表や式は示されているものの、Tönniesの方法については詳しく論じられていない。このため本稿末に〔筆者補注〕を加えた。なお〔筆者補注〕の解説を明確にするため、本来、関係する図表には図表番号は付されていないが、ここでは〔図表1〕～〔図表7〕の通し番号を付した。

〔図表 2〕

		財産罪				
		I	II	III	IV	V
穀 備	I	a_1				
	II	a_3				
	III		a_2			
	IV			a_6		
	V				a_4	

〔図表 3〕

		身体罪				
		I	II	III	IV	V
穀 備	I					a_3
	II				a_6	a_1
	III			a_4		
	IV				a_2	
	V		a_5			

〔図表 4〕

府県	人口千ニ付出生		人口千ニ付死亡	
a_1	40	I	32	I
a_2	42	I	31	I
a_3	28	IV	19	IV
a_4	36	II	27	II
a_5	23	V	12	IV
a_6	33	III	23	III
a_7	43	I	13	V
:	:	:	:	:
a_{16}	29	IV	22	II
a_{17}	32	III	28	II
a_{18}	37	II	23	III
a_{19}	26	IV	14	V
a_{20}	29	IV	17	IV

〔図表 5〕

出生年階級			死亡年階級		
人口千人 ニ付出生	40以上	I	人口千人 ニ付死亡	30以上	I
	35-40	II		25-30	II
	30-35	III		20-25	III
	25-30	IV		15-20	IV
	25以下	V		15以下	V

〔図表6〕

		死亡年				
		I	II	III	IV	V
出生年	I	a_1 a_7				
	II		a_4	a_2		
	III		a_{17}	a_6		
	IV			a_{16}	a_3	a_{19}
	V					a_5

〔図表7〕

	A	B	C	D	E
A	2		1		1
B		1	1	1	1
C		3		1	
D	1		1	2	
E	1		1		2

(一) 正負関係ノ計算

吻合位置 + 7 - 3

傍触位置 + 6 - 8

(二) 吻合位置ト傍触位置トノ価値

(吻合区：全区画) = (5 : 25) = (1 : 5)

(傍触区：吻合区ヲ除ケル残区) = (8 : 20) = (1 : 2.5)

(吻合区：傍触区) = (5 : 2.5) = (2 : 1)

$(7 \times 2) - (3 \times 2) = 14 - 6 = + 8$

$(6 \times 1) - (8 \times 1) = 6 - 8 = - 2$

$+ 8 - 2 = + 6$

(三) 正負関係ノ限度

限度 $\pm (20 - 4) \times 2 = \pm 32$

$+ 32 > 0 > - 32$

$6 : 36 = 1 : 6$

$10 : 30 = 1 : 3 \quad 2 : 1$

限度 $+ 2x > 0 > - 2x$

Coefficient of Correlation

相互關係

a_1	b_1	$\frac{+n}{n} = +1$
a_2	b_2	
a_3	b_3	$\frac{-n}{n} = -1$
a_4	b_4	
a_5	b_5	$\frac{+0(-0)}{n} = 0$
\vdots	\vdots	
a_n	b_n	$+1 > 0 > -1$

要之吾人ハ、上述ノ如キ順序及ヒ方法ニ依リテ統計材料ヲ新ニ取扱ヒ、吾人ノ研究目的ヲ達スルニ努メサルヘカラス。殊ニ又此ノ際吾人ノ注意スヘキハ、公平冷静態度ヲ維持スルコト也。之蓋シ統計ニ基ク研究ハ、人ニ正確ナリトノ感ヲ与フルコト正シキ傾向アリ。而カモ研究ノ結果ハ、屢々著シキ利害關係ヲ左右スルコトアルヲ以也。兎ニ角吾人ハ、各方面ニ充分ノ注意ヲ払ヒテ通性ノ発見、因果關係ノ説明ニ思ヒヲ凝ラササルヘカラス。不充分、不正確ナル統計ニヨリテ、輕々シク大胆ナル結論ヲ試ミルコトハ、云ハハ統計的研究者ノ自殺行為ト云フヲ得ヘシ。假令研究ノ結果カ或ル予想セラレタル因果關係ヲ否定スルニスキサル結論ニ達シタル場合ニ於テモ、尚一種ノ確實ナル証明ニシテ却テ彼ノ大胆ニシテ積極的ナル結論カ實際上何事ヲモ証明セサルニ比シ、遙ニ優レリトス。但シ之ト同時ニ、吾人ハ、單純ナル統計蒐集家ニ終ルコトナキ様注意セサルヘカラス。統計ノ上ニ假定ヲ作り、假定ノ下ニ更ニ統計ヲ集ムルト云フ如ク、常ニ宜シク材料ヲ制整シテ其ノ活用ニ心ヲ費ササルヘカラス。單純ナル材料蒐集家タルコトト独断家トノ間ニ於テ、統計的研究者ノ採ルヘキ針路ハ自ラ横ハレルモノト考フ。

(以下次号)

〔筆者補注〕

高野がここで紹介しているTönniesとは、よく知られたドイツの社会学者Ferdinand Tönnies (1855–1936)が提案した方法である。長屋政勝によれば、Tönniesは高野の師であったMayrの統計学に対する先鋭的な批判者の一人であった⁶⁾。しかしこのような事情にも拘わらず、高野がTönniesの方法を紹介しているのは、これが相関関係の把握を目的とする方法として当時よく知られていたからであろう。例えば森数樹は「此ノ方法ハ交聯性〔相関関係〕ノ有無ヲ判断スルモノトシテハ極メテ幼稚デアアルガ、前段述ベシ交聯係数〔積率相関係数〕ヲ求ムル方法ニ比シテハ頗ル容易デアアルガ為、現今尚盛ニ用ヒラルルノデアアル」と述べている⁷⁾。そこで以下では、森の解説と高野の図表を使いながらTönniesの方法について説明する⁸⁾。

この方法は、2変数のデータをまず奇数の階級に分類することから始まる。その際、階級数は両変数共に同数でなければならない。例えば高野の講義録の事例では、〔図表5〕のように各道府県の年間出生数(千分率)と年間死亡数(千分率)をI~Vの階級に分け、〔図表4〕のようなデータを作成している。このデータを基に、出生年(年間出生数)と死亡年(年間死亡者数)からなる5×5のクロス集計表を作成している。このようなクロス集計表において、対角線上における各セルに等しく度数が集中している〔図表2〕のような場合(吻合位置)は完全な正の相関、〔図表3〕のような場合は完全な負の相関があると看做される。〔図表4〕の場合では、データ数が20であるから、例えば以下の〔図表8〕のように対角線上にあるグレーのセルの各度数が4として分布している場合、完全な正の相関があると認められる。しかしながらこれは完全な相関の場合であり、この状態を基準に相関の強さが測定されなければならない。具体的には、吻合位置にあるセル度数の相違や、〔図表9〕のように対角線上の近傍に布置するセル(傍触位置)の数並びにその度数の影響も斟酌した上で相関の強さが評価される。

高野の講義録における事例に沿って具体的に見ていこう。〔図表7〕は、〔図表4〕におけるセル度数の分布を示したものである。講義録の「(一) 正負関係ノ計算」でも示されているように、〔図表7〕に示されている、正の吻合位置にあるセル度数(吻合セル度数)の合計は7で+7、負の吻合セル度数の合計は3で-3となる。また正の傍触位置にあるセル度数(傍触セル度数)は6で+6、負の傍触セル度数は8で-8となる。この数字に基にして、「(二) 吻合位置ト傍触位置トノ価値」では次のように計算を展開している。

⁶⁾ この点については、長屋政勝『ドイツ社会統計方法論史』梓出版、1992年、13~70頁を参照のこと。

⁷⁾ 森数樹『一般統計論』丸善、1920年、183頁。なおこの著書は、その序文でも述べられているように、G. U. Yuleのよく知られた次のテキストを下敷きとして書かれたものであるが、Tönniesの方法については、Yuleのテキストには取り上げられていないようなので、この部分の記述は森が補足したものと考えられる。Yule, G. U., *An Introduction to the Theory of Statistics*, 2nd ed., Griffin, 1912.

⁸⁾ 森の前掲書と併せて、同著者による次の文献も参照した。森数樹『統計学概論』巖松堂、1924年。

〔図表8〕

		死亡年（年間死亡者数）				
		I	II	III	IV	V
出生年 （年間出生数）	I					
	II					
	III					
	IV					
	V					

〔図表9〕

		死亡年（年間死亡者数）				
		I	II	III	IV	V
出生年 （年間出生数）	I					
	II					
	III					
	IV					
	V					

先ず〔図表7〕の事例では、5×5のクロス集計表であるため、セルの総数は25となり、この内、正もしくは負の吻合位置にあるセルの数（吻合セル数）は5であるから、吻合セル数と総セル数の比は5：25=1：5となる。一方、正もしくは負の傍触位置にあるセルの数（傍触セル数）と吻合セル数を除いた総セル数の比は、8：20=1：2.5となる。故に吻合セル数と傍触セル数の比は5：2.5=2：1となるため、吻合セル度数に2、傍触セル度数に1のウェイトをかけてセル度数の合計を求める。計算結果は本文中にも示されているが、再掲すると、ウェイトをかけた吻合セル度数は、正のセル度数が7×2=14で+14、負のセル度数が3×2で-6、従って+8となる。また傍触セル度数については、正のセル度数が6×1=6で+6、負のセル度数が8×1で-8、従って-2となり、最終的に吻合セル度数と傍触セル度数の合計は+8-2=+6となる。

〔図表10〕

		死亡年（年間死亡者数）				
		I	II	III	IV	V
I	4					
II		4				
III			4			
IV				4		
V					4	

次にこの数値例の最大値を見てみよう。〔図表10〕は、〔図表4〕のデータにおいて完全な正の相関が与えられる場合のセル度数を示している。前述の計算規則に従い、吻合セル度数を2倍し、且つ行カテゴリーⅢと列カテゴリーⅢに布置する吻合セル度数を、正の吻合セル度数と負の吻合セル度数が交差するため差引0とすると、合計は $8 \times 4 = 32$ となる。「(三) 正負関係ノ限度」で、限度が $+32 > 0 > -32$ とされているのはこのことを意味する。従って、係数でこの事例の相関の強さを評価すると、 $6 / 32 = 0.187$ となり、この事例ではほとんど相関が認められないということになる。なお同節で限度が、「6 : 36…」の記述があるが、これは講義録の説明からは判読不可能で、恐らく講義中に 6×6 のクロス集計表に拡張した場合の計算方法とその一般式を示したものと推量される。