Introduction to Correspondence Analysis

Jan de Leeuw

DEPARTMENT OF STATISTICS, UNIVERSITY OF CALIFORNIA, LOS ANGELES, CA 90095-1554

E-mail address, Jan de Leeuw: deleeuw@stat.ucla.edu

URL, Jan de Leeuw: http://gifi.stat.ucla.edu

2000 *Mathematics Subject Classification.* 62H25

Key words and phrases. Multivariate Analysis, Correspondence Analysis

ABSTRACT. We introduce Correspondence Analysis of non-negative matrices from the point of view of maximixing the correlation, linearizing the regressions, approximating the Benzécri distances, and drawing the graph corresponding with the matrix.

Contents

Chapter 1. Frequency and Proportion Data 5
1. Notation 5
2. Examples 5

Chapter 2. Maximum Correlation 9
1. Introduction 9
2. Stationary Equations 10
3. Linearizing the regressions 12
4. Maximizing Correlation Ratio's 12
5. Reciprocal Averaging and the Centroid Principle 13
6. Computation 13
7. Monotonicity Restrictions 13
8. Examples 13

Chapter 3. Chi-square Decomposition of Bivariate Tables 15
1. Different solutions 15
2. More on Existence 16
3. The Burt Table, Indicator Matrices, and Canonical Correlation 17
4. Approximating the Benzécri Distances 17
5. Decomposing Chi-square 18
6. The Bivariate Normal, Polynomiality 18
7. Horseshoes 18
8. Joint Plots 18

Chapter 4. Graph Drawing 19
1. Graph Drawing 19
2. Binary Data 20
3. Dédoublement 20

4. Bounded Data 20
5. Paired Comparisons and Rank Orders 20

Chapter 5. Data Sets 21
1. Galton Data 22
2. Pearson Data 23
3. Senate Data 23
4. Sleeping Bag Data 23
5. Mammals Data 23
6. GALO Data 23

Bibliography 25

CHAPTER 1

Frequency and Proportion Data

1. Notation

Suppose F is an $n \times m$ table with *frequencies*. Without loss of generality we assume $m \leq n$ (if $m > n$ just transpose the table).

The row sums are collected in a diagonal matrix F_r, the column sums in a diagonal matrix F_c. We use τ for the sum of all elements of F. Again without loss of generality we assume the marginals are non-zero. If they happen to be zero, we just delete the corresponding row or column.

Also define the matrix P of *proportions* by $P = \frac{1}{k}F$. The diagonal matrices of marginals of P are P_r and P_c. Clearly the elements of P (and of P_r and P_c) add up to one.

Finally, we define *row-profiles* as $R = F_r^{-1}F = P_r^{-1}P$ and the *column-profiles* as $C = FF_c^{-1} = PP_c^{-1}$. Rows of R all add up to one, so do columns of C.

2. Examples

In this section we give some classical examples of frequency tables. We shall use these examples in other sections to illustrate our techniques.

2.1. Intelligence and Clothing. Around 1910 Karl Pearson, among many others, was very interested in general intelligence, and in particular in the role of nature and nurture in transmission of intelligence. One scale used at the time was an estimate by the teacher of the general intelligence of the child. Pearson defined seven ordered categories. We take the following descriptions literally from Waite [1911, pag 93].

A – Mentally Defective: Capable of holding in the mind only the simplest facts, and incapable of perceiving or reasoning about the relationship between facts.

B – Slow Dull: Capable of perceiving relationship between facts in some few fields with long and continuous effort; but not generally nor without much assistance.

C – Slow: Very slow in thought generally, but with time understanding is reached.

D – Slow Intelligent: Slow generally, although possibly more rapid in certain field; quite sure of knowledge when once acquired.

E – Fairly Intelligent: Ready to grasp, and capable of perceiving facts in most fields; capable of understanding without much effort.

F – Distinctly Capable: A mind quick in perception and in reasoning rightly about the perceived.

G – Very Able: Quite exceptionally able intellectually, as evidenced either by the person's career or by consensus of opinion of acquaintances, or by school record in case of children.

Pearson and his co-workers were interested is disentangling the effects of nature and nurture, and for that reason they studied the cross-tabulation of teacher's rating of intelligence and teacher's rating of the clothing of the student.

The clothing scale, taken from Gilby and Pearson [1911, page 96], is

I: Very well clad.

II: Well clad, stuff suit, good boots; sufficient, even if poor.

III: Clothing poor, but passable; and old and, perhaps, ragged suit with some attempt at proper underclothing.

IV: Clothing insufficient; boots bad and leaking.

V: Clothing the worst; no boots or makeshift substitutes for them.

The data are [Gilby and Pearson, 1911, page 103].

Intelligence

Clothing	B	C	D	E	F	G	
I	33	48	113	209	194	39	636
II	41	100	202	255	138	15	751
III	39	58	70	61	33	4	265
IV,V	17	13	22	10	10	1	73
	130	219	407	535	375	59	1725

Of course Pearson was well aware of the difficulty of solving the nature and nurture problem in this way.

> Is the lower intelligence of the children due to the poorer home environment evidenced by the worse clothing, or is the worse clothing only a mark of the lower intelligence of the parents, which is naturally reproduced in their children ? [Gilby and Pearson, 1911, page 97]

2.2. Eye Color and Hair Color. The data are from a large scale anthropometric study by Tocher [1908, 1909]. The actual table we use was compiled from Tocher's data for Caithness by Maung [1941a,b]. Fisher [1940] used it for what is possibly the first practical application of CA.

Hair Color

Eye Color	Fair	Red	Medium	Dark	Black	
Blue	326	38	241	110	3	718
Light	688	116	584	188	4	1580
Medium	343	84	909	412	26	1774
Dark	98	48	403	681	85	1315
	1455	286	2137	1391	118	5387

2.3. Occupational Mobility in the Fifties. The data collected by Glass [1954] give the occupational status of $T = 3497$ British father-son pairs. Occupational status in this study is a variable with seven discrete values (or categories). Thus $m = n = 7$ and the table F is square.

The labels are abbreviations for

PROF: professional and high administrative
EXEC: managerial and executive
HSUP: higher supervisory
LSUP: lower supervisory
SKIL: skilled manual and routine non-manual
SEMI: semi-skilled manual
UNSK: unskilled manual

The data are

		Occupation of Son							
		PROF	EXEC	HSUP	LSUP	SKIL	SEMI	UNSK	
Occupation of Father	PROF	50	19	26	8	18	6	2	129
	EXEC	16	40	34	18	31	8	3	150
	HSUP	12	35	65	66	123	23	21	345
	LSUP	11	20	58	110	223	64	32	518
	SKIL	14	36	114	185	714	258	189	1510
	SEMI	0	6	19	40	179	143	71	458
	UNSK	0	3	14	32	141	91	106	387
		103	159	330	459	1429	593	424	3497

CHAPTER 2

Maximum Correlation

1. Introduction

Computing the Pearson product moment correlation coefficient for some of our frequency tables F seems impossible. In most examples the variables in the table are not numerical, which means we have no numbers to work with to compute our means, variances, and covariance.

The value of the correlation coefficient $\rho(x, y)$ depends on values (scores) x and y assigned to the rows and the columns. In most of our example so far these scores are not available. And even if they are, we may wonder how the correlation coefficient changes if we changed the scores (i.e. if we transformed the variables).

Over the last 100 years there have been many statisticians who have studied ρ as a function of x and y [Pearson, 1906; Hirschfeld, 1935; Fisher, 1940; Maung, 1941b; Gebelein, 1941; Lancaster, 1958, 1969; Renyi, 1959; Csáki and Fisher, 1963; Dembo et al., 2001; Bryc et al., 2004].

To study this problem it is more convenient to work with P and with its diagonal matrices P_c and P_r of row and column marginals. Suppose the scores x and y are in deviations of the mean, i.e. suppose $u_n' P_r x = 0$ and $u_m' P_c y = 0$, where u_n and u_m are vectors of the appropriate length with all their elements equal to one. The variances of the row and column scores are $x' P_r x$ and $y' P_c y$, and the covariance is $x' P y$.

We can maximize the correlation $\rho(x, y)$ by choosing scores x and y with unit variances that maximize the covariance. This is a maximization problem with equality constraints, which can be solved by introducing Lagrange multipliers and differentiating the corresponding Lagrangian. Because we are maximizing a continuous and differentiable function on a compact set, the maximum correlation exists, and is given by a solution of the stationary equations.

One could also compute the minimum correlation coefficient. But clearly if scores $\hat{x}$ and $\hat{y}$ give the maximum correlation $\hat{\rho}$, then $\hat{x}$ and $-\hat{y}$ give the minimum correlation $-\hat{\rho}$. Thus the problem of computing the minimum correlation is not of independent interest. We shall briefly refer to the problem of minimizing $\rho^2(x, y)$ below.

2. Stationary Equations

The Lagrangian is

$$\begin{aligned}(1)\quad \mathbf{L}(x, y, \lambda, \mu, \alpha, \beta) &= \\ &= x'Py - \lambda(x'P_rx - 1) - \mu(y'P_cy - 1) - \alpha u_n'P_rx - \beta u_m'P_cy\end{aligned}$$

Differentiation leads to the stationary equations

$$Py = \lambda P_rx + \alpha P_ru_n, \tag{2a}$$

$$P'x = \mu P_cy + \beta P_cu_m, \tag{2b}$$

$$x'P_rx = 1, \tag{2c}$$

$$y'P_cy = 1, \tag{2d}$$

$$x'P_ru_n = 0, \tag{2e}$$

$$y'P_cu_m = 0. \tag{2f}$$

Of coure there may be additional solutions of these stationary equations which define neither maxima nor minima. This will be analyzed in detail below.

We also study the somewhat simpler system

$$Py = \rho P_r x, \tag{3a}$$

$$P'x = \rho P_c y, \tag{3b}$$

$$x' P_r x = 1, \tag{3c}$$

$$y' P_c y = 1. \tag{3d}$$

The next two theorems show that solutions of (2) can be easily computed from those of (3), and vice versa.

THEOREM 2.1.

If $(\hat{x}, \hat{y}, \hat{\lambda}, \hat{\mu}, \hat{\alpha}, \hat{\beta})$ *solves* (2) *then*

(1) $\hat{\lambda} = \hat{\mu} = \rho(\hat{x}, \hat{y})$,
(2) $\hat{\alpha} = \hat{\beta} = 0$,

and thus $(\hat{x}, \hat{y}, \rho(\hat{x}, \hat{y}))$ *solves* (3).

PROOF. Suppose$(\hat{x}, \hat{y}, \hat{\lambda}, \hat{\mu}, \hat{\alpha}, \hat{\beta})$ solves (2). If we premultiply (2a) by $\hat{x}'$ and (2b) by $\hat{y}'$ we see that

$$\hat{x}' P \hat{y} = \hat{\lambda} \hat{x}' P_r \hat{x} + \hat{\alpha} \hat{x}' P_r u_n = \hat{\lambda},$$
$$\hat{y}' P' \hat{x} = \hat{\mu} \hat{y}' P_c \hat{y} + \hat{\beta} \hat{y}' P_c u_m = \hat{\lambda},$$

and thus $\hat{\lambda} = \hat{\mu} = \hat{x}' P \hat{y} = \rho(\hat{x}, \hat{y})$. If we premultiply (2a) by u_n' and (2b) by u_m' we see that

$$u_n' P \hat{y} = \hat{\lambda} u_n' P_r \hat{x} + \hat{\alpha} u_n' P_r u_n = \hat{\alpha},$$
$$u_m' P' \hat{x} = \hat{\mu} u_m' P_c \hat{y} + \hat{\beta} u_m' P_c u_m = \hat{\beta},$$

and because $u_n' P \hat{y} = u_m' P_c \hat{y} = 0$ and $u_m' P' \hat{x} = u_n' P_r \hat{x} = 0$ we have $\hat{\alpha} = \hat{\beta} = 0$. □

Theorem 2.1 has a partial converse.

THEOREM 2.2.

(1) $(u_n, u_m, 1)$ *solves* (3),

(2) *if* $(\hat{x}, \hat{y}, 1)$ *solves* (3) *and* $\hat{x}' P_r u_n = \hat{y}' P_c u_m = 0$ *then* $(\hat{x}, \hat{y}, 1, 1, 0, 0)$ *solves* (2),

(3) *if* $(\hat{x}, \hat{y}, \hat{\rho})$ *solves* (3) *and* $\hat{\rho}^2 < 1$, *then* $\hat{x}' P_r u_n = \hat{y}' P_c u_m = 0$ *and* $(\hat{x}, \hat{y}, \hat{\rho}, \hat{\rho}, 0, 0)$ *solves* (2).

PROOF. We can easily see that (a) and (b) are true by substituting them into the stationary equations. Part (c) is a bit less obvious. Suppose $(\hat{x}, \hat{y}, \hat{\rho})$ solves (3). Premultiplying (3a) by u_n' and (3b) by u_m' gives

$$u_m' P_c \hat{y} = \hat{\rho} u_n' P_r \hat{x},$$
$$u_n' P_r \hat{x} = \hat{\rho} u_m' P_c \hat{y}.$$

This implies $(1 - \hat{\rho}^2) u_n' P_r \hat{x} = (1 - \hat{\rho}^2) u_m' P_c \hat{y} = 0$. Thus if $\hat{\rho}^2 < 1$ the solutions $\hat{x}$ and $\hat{y}$ satisfy $u_n' P_r \hat{x} = u_m' P_c \hat{y} = 0$. Now use part (2b). This proves the second part. □

Summarize

3. Linearizing the regressions

The regression are linear for standardized scores x and y if the conditional expectations are on straight lines throught the origin. This means

$$D^{-1} F y = \lambda x,$$
$$E^{-1} F' x = \lambda y.$$

But these are exactly the stationary equations for optimizing the correlation coefficient. Thus any solution of the stationary equations give scores linearizing the regressions, and vice versa.

Another way of saying this: if we have scores linearizing the regressions, then a small perturbation of the scores will not change the correlation coefficient.

4. Maximizing Correlation Ratio’s

Perfect correlation.

5. Reciprocal Averaging and the Centroid Principle

6. Computation

7. Monotonicity Restrictions

8. Examples

CHAPTER 3

Chi-square Decomposition of Bivariate Tables

1. Different solutions

The following theorem generalizes part (2c) of Theorem 2.2, which is the special case in which one of the solutions is $(u_n, u_m, 1)$.

THEOREM 1.1. *Suppose* $(\hat{x}, \hat{y}, \hat{\rho})$ *and* $(\tilde{x}, \tilde{y}, \tilde{\rho})$ *are two solutions to* (3). *If* $\hat{\rho} \neq \tilde{\rho}$ *then* $\hat{x}' P_r \tilde{x} = \hat{y}' P_c \tilde{y} = 0$.

PROOF.

$$P\hat{y} = \hat{\rho} P_r \hat{x}, \tag{4a}$$

$$P'\hat{x} = \hat{\rho} P_c \hat{y}, \tag{4b}$$

and

$$P\tilde{y} = \tilde{\rho} P_r \tilde{x}, \tag{4c}$$

$$P'\tilde{x} = \tilde{\rho} P_c \tilde{y}. \tag{4d}$$

Premultiply (4a) by $\tilde{x}'$ and (4b) by $\tilde{y}'$. Use (4c) and (4d) to obtain

$$\tilde{\rho} \tilde{y}' P_c \hat{y} = \hat{\rho} \tilde{x}' P_r \hat{x},$$
$$\tilde{\rho} \tilde{x}' P_r \hat{x} = \hat{\rho} \tilde{y}' P_c \hat{y},$$

which implies $(\hat{\rho}^2 - \tilde{\rho}^2)\hat{x}' P_r \tilde{x} = (\hat{\rho}^2 - \tilde{\rho}^2)\hat{y}' P_c \tilde{y} = 0$. Thus if $\hat{\rho}^2 \neq \tilde{\rho}^2$ we have $\hat{x}' P_r \tilde{x} = \hat{y}' P_c \tilde{y} = 0$. □

Thus different solutions, with different correlation coefficients, are orthogonal to each other. In particular, this implies our previous result that all solutions with a correlation coefficient less than one are centered. It also implies there cannot be more than m solutions with different correlation coefficients (because we cannot have more than m orthogonal vector of length m).

What remains to be done is to look at different solutions with the same correlation coefficient.

THEOREM 1.2. *Suppose* $(\hat{x}, \hat{y}, \hat{\rho})$ *and* $(\tilde{x}, \tilde{y}, \hat{\rho})$ *are two solutions to* (3). *Then* $(\alpha\hat{x} + \beta\tilde{x}, \alpha\hat{y} + \beta\tilde{y}, \hat{\rho})$ *is a solution of* (3) *for all* α *and* β.

PROOF. Simple substitution. □

Thus all (unnormalized) solutions with correlation equal to $\hat{\rho}$ form a linear subspace. Solutions with a different correlation $\tilde{\rho}$ form another subspace, orthogonal to the first one. Again, the dimensionality of the different solution subspaces cannot add up to more than m. It also follows that if the dimensionality of the solution subspace corresponding with $\hat{\rho}$ is larger than one, then we can choose an orthonormal basis of solutions in this space. Thus if the dimensionality is d we can choose $x_1, \cdots, x_d$ such that $x_i' P_r x_j = 0$ for all $1 \leq i \neq j \leq d$. And the same for the row scores $y_1, \cdots, y_d$.

Or, in summary, different solutions either are orthogonal (when they correspond to different correlations) or they can be chosen to be orthogonal (when they correspond with the same correlation coefficient).

2. More on Existence

So far we have shown that scores giving the maximal correlation exists, and that additional solutions, if they exist, are orthogonal to each other. This is actually enough to give a general answer to the existence problem.

THEOREM 2.1. *We can find an* $n \times m$ *matrix* X*, an* $m \times m$ *matrix* Y*, and an* $m \times m$ *non-negative diagonal matrix* S*, with elements non-increasing* *along the diagonal, such that*

$$PY = P_r X S,$$
$$P'X = P_c Y S,$$
$$X' P_r X = I,$$
$$Y' P_c Y = I.$$

PROOF. The proof is constructive and builds up X and Y column by column. In step $s = 0$, we start with the solution $(u_, u_m, 1)$. In step $1 <= s <= m - 1$ we find the next column of X and Y by maximizing the correlation coefficient $\rho(x, y)$ over all scores x and y, with the additional condition that x and y are orthogonal to the previous $s - 1$ solutions $\hat{x}_0, \cdots, \hat{x}_{s-1}$ and $\hat{y}_0, \cdots, \hat{y}_{s-1}$ that we have already computed. □

$$X' P Y = R$$

$$Y^{-1} = Y' E$$

$$P = D X R Y' E$$

SVD

3. The Burt Table, Indicator Matrices, and Canonical Correlation

4. Approximating the Benzécri Distances

Also define the column sums $p_{\bullet j}$, the row sums $p_{i\bullet}$ and the normalized rows $pj|i = \frac{p_{ij}}{p_{i\bullet}}$ and normalized columns $p_{i|j} = \frac{p_{ij}}{p_{\bullet j}}$.

The Benzécri, or chi-square, distance between rows i and k of P is

$$\delta_{ik}^2 = T \sum_{j=1}^{m} \frac{(p_{j|i} - p_{j|k})^2}{p_{\bullet j}}$$

The delta method shows that if row and column variables are independent, then $\delta_{ik}^2 \overset{\mathcal{L}}{\Rightarrow} \chi_{m-1}^2$.

Let

$$h_{ij} = \frac{\sqrt{T}\, p_{ij}}{p_{i\bullet}\sqrt{p_{\bullet j}}}.$$

Then δ_{ik}^2 is the squared Euclidean distance between rows i and k of H. The matrix H can also be written as $H = D^{-1}FE^{-1/2}$, and thus, using unit vectors e_i and e_k,

$$\delta_{ik}^2 = (e_i - e_k)'D^{-1}FE^{-1}F'D^{-1}(e_i - e_k).$$

But $D^{-1}FE^{-1} = X\Lambda Y'$ and thus

$$\delta_{ik}^2 = (e_i - e_k)'X\Lambda^2 X'(e_i - e_k),$$

which means the Benzécri distances are squared Euclidean distances between the rows of $\overline{X} = X\Lambda$. In fact

$$\delta_{ik}^2 = (\overline{x}_i - \overline{x}_k)'(\overline{x}_i - \overline{x}_k).$$

The same reasoning can obviously be applied to the columns.

5. Decomposing Chi-square

6. The Bivariate Normal, Polynomiality

7. Horseshoes

8. Joint Plots

CHAPTER 4

Graph Drawing

1. Graph Drawing

A contingency table or cross table can be interpreted as the adjacency matrix of a weighted bipartite graph. The graph has $m+n$ vertices and mn edges, with the edge connecting row-vertex i and column-vertex j having weight f_{ij}. If $f_{ij}=0$ then the edge has weight zero, and can be simply considered to be absent.

In graph drawing [Battista et al., 1998] we want to draw a picture of the graph that is easy to read and clearly shows the most important relationships. Vertices are generally presented as point, and edges as lines. The general of a "nice" or "clear" drawing can be quantified in many different ways, but we shall interpret it to mean that edges with a large weight in the graph should generally be short in the drawing.

More precisely we want to minimize

$$\sigma(X,Y)=\sum_{i=1}^{n}\sum_{j=1}^{m} f_{ij}d_{ij}^2(X,Y),$$

where $d_{ij}^2(X,Y)$ is the squared distance between row i of X and row j of Y. Expanding gives

$$\sigma(X,Y)=\mathbf{tr}\ X'DX+\mathbf{tr}\ Y'EY-2\mathbf{tr}\ X'FY.$$

If we want to minimize the loss function we need a normalization condition to rule out the trivial solution in which both X and Y are zero. There are three obvious choices for normalization. We can require $X'DX = TI$ or $Y'EY = TI$ or $X'DX + Y'EY = 2TI$. In principle we could also require $X'DX = Y'EY = TI$, but these last conditions are obviously more restrictive.

Let us first compute the solution requiring $X'DX = TI$. There are no restrctions on Y, which means that the optimum Y for given X is $\hat{Y} = E^{-1}F'X$. Substituting this, and using $X'DX = TI$, shows that

$$\sigma(X, \hat{Y}) = pT - \mathbf{tr}\; X'FE^{-1}F'X$$

and thus ...

Requiring $Y'EY = TI$ and not restricting X gives, by the same argument, $F'D^{-1}FY = Y\Lambda$ and $X = D^{-1}FY$.

Requiring $X'DX + Y'EY = 2TI$ gives the equations

$$FY = DX\Lambda,$$
$$F'X = EY\Lambda.$$

2. Binary Data

If the data matrix F is binary then the usual introductions of Correspondence Analysis, using correlation and chi-square, do not make much sense. The same is true for the Benzécri distances. But the graph drawing rationale still applies, and in fact becomes more convincing.

The loss function σ now is the total squared length of the vertices in the graph. We have

$$\sigma(X, Y) = \sum\sum\{d_{ij}^2(X, Y) \mid f_{ij} = 1\}$$

Again we can choose different normalizations sich as $X'F_{\rightarrow}X = kI$ or $Y'F_{\downarrow}Y = kI$, or both.

3. Dédoublement

4. Bounded Data

5. Paired Comparisons and Rank Orders

CHAPTER 5

Data Sets

1. Galton Data

0	0	0	0	0	0	0	0	0	0	0	1	3	0
0	0	0	0	0	0	0	1	2	1	2	7	2	4
0	0	0	0	1	3	4	3	5	10	4	9	2	2
1	0	1	0	1	1	3	12	18	14	7	4	3	3
0	0	1	16	4	17	27	20	33	25	20	11	4	5
1	0	7	11	16	25	31	34	48	21	18	4	3	0
0	3	5	14	15	36	38	28	38	19	11	4	0	0
0	3	3	5	2	17	17	14	13	4	0	0	0	0
1	0	9	5	7	11	11	7	7	5	2	1	0	0
1	1	4	4	1	5	5	0	2	0	0	0	0	0
1	0	2	4	1	2	2	1	1	0	0	0	0	0

2. Pearson Data

3. Senate Data

4. Sleeping Bag Data

5. Mammals Data

6. GALO Data

Bibliography

G. Di Battista, P. Eades, R. Tamassia, and I. Tollis. *Graph Drawing; Algorithms for Geometric Representations of Graphs*. Prentice Hall, 1998.

W. Bryc, A. Dembo, and A. Kagan. On the Maximum Correlation Coefficient. *Theory of Probability and its Applications*, 2004.

P. Csáki and J. Fisher. On the General Notion of Maximal Correlation. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, 8:27–51, 1963.

A. Dembo, A. Kagan, and L. Shepp. Remarks on the Maximum Correlation Coefficient. *Bernoulli*, 7:343–350, 2001.

R.A. Fisher. The Precision of Discriminant Functions. *Annals of Eugenics*, 10:422–429, 1940.

H. Gebelein. Das Statistische Problem der Korrelation als Variations- und Eigenwertproblem und sein Zusammenhang mit der Ausgleichsrechnung. *Zeitschrift für Angewandte Mathematik und Mechanik*, 21:364–379, 1941.

W.G. Gilby and K. Pearson. On the Significance of the Teacher's Appreciation of General Intelligence. *Biometrika*, 8:94–108, 1911.

D.V. Glass. *Social Mobility in Britain*. Free Press, 1954.

H.O. Hirschfeld. A Connection between Correlation and Contingency. *Proceedings Cambridge Philosophical Society*, 31:520–524, 1935.

H.O. Lancaster. The Structure of Bivariate Distributions. *Annals of Mathematical Statistics*, 29:719–736, 1958.

H.O. Lancaster. *The Chi-squared Distribution*. Wiley, New York, New York, 1969.

K. Maung. Discriminant Analysis of Tocher's Eye Colour Data for Scottish School Children. *Annals of Eugenics*, 11:64–76, 1941a.

K. Maung. Measurement of Association in a Contingency Table with Special Reference to the Pigmentation of Hair and Eye Colour of Scottish School Children. *Annals of Eugenics*, 11:189–223, 1941b.

K. Pearson. On Certain Points Connected with Scale Order in the Case of a Correlation of Two Characters which for Some Arrangement Give a Linear Regression Line. *Biometrika*, 5:176–178, 1906.

A. Renyi. On Measures of Dependence. *Acta Mathematica Academy Sciences Hungary*, 10:441–451, 1959.

J.F. Tocher. Pigmentation Survey of School Children in Scotland. *Biometrika*, 6(2/3):129–235, Sep 1908.

J.F. Tocher. Pigmentation Survey of School Children in Scotland. *Biometrika*, 6(4):1–67, Mar 1909.

H. Waite. The Teachers Estimation of the General Intelligence of School Children. *Biometrika*, 8:79–93, 1911.