Categorical Data Analysis and Visualisation

Part II: Two-way Contingency Tables

Rosaria Lombardo
Department of Economics,
University of Campania "Luigi Vanvitelli", Italy
rosaria.lombardo@unicampania.it

Eric J. Beh
University of Wollongong, Wollongong, Australia,
Stellenbosch University, Stellenbosch, South Africa
ericb@uow.edu.au

The Contingency Table

- Notation
- Early Examples

Notation

Consider

- A random sample of n individuals/units from which two categorical variables are considered
- Two categorical variables, A and B, are cross-classified to form a two-way contingency table, **N**.
- Let variable A consist of I categories
- Let variable B consist of J categories
- **N** is of size I × J

- Denote n_{ij} as the (i, j)th cell frequency
- Denote $n_{i\bullet}$ as the i'th row marginal frequency
- Denote $n_{\bullet j}$ as the j'th column marginal frequency

A/B	B_1	B_2	$\cdots$	B_j	$\cdots$	B_J	Total
A_1	n_{11}	n_{12}	$\cdots$	n_{1j}	$\cdots$	n_{1J}	$n_{1\bullet}$
A_2	n_{21}	n_{22}	$\cdots$	n_{2j}	$\cdots$	n_{2J}	$n_{2\bullet}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
A_i	n_{i1}	n_{i2}	$\cdots$	n_{ij}	$\cdots$	n_{iJ}	$n_{i\bullet}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
A_I	n_{I1}	n_{I2}	$\cdots$	n_{Ij}	$\cdots$	n_{IJ}	$n_{I\bullet}$
Total	$n_{\bullet 1}$	$n_{\bullet 2}$	$\cdots$	$n_{\bullet j}$	$\cdots$	$n_{\bullet J}$	n

We shall consider the case of more than two categorical variables later.

Quetelet's Contingency Tables

One of the first serious categorical data analysts was

Lambert Adolphe Jacques Quetelet
(1796-1874)

His studies on social aspects in France involved the construction of what we now know to be the contingency table. His analysis of these tables was, by current standards, more than superficial, but he did set up the ground word for data analysis that Galton (who developed linear regression analysis) and Pearson (who is a pioneer of categorical data analysis) continued on with.

Their work lives on even today.

What follows is a Quetelet's table, although he did not refer to it as a contingency table.

Quetelet's Contingency Tables

	1826.	1827.	1828.	1829.	1830.	1831.
Murders in general, -	241	234	227	231	205	266
Gun and pistol, - -	56	64	60	61	57	88
Sabre, sword, stiletto, poniard, dagger, &c.,	15	7	8	7	12	30
Knife, - - - - - - -	39	40	34	46	44	34
Cudgels, cane, &c., -	23	28	31	24	12	21
Stones, - - - - - - -	20	20	21	21	11	9
Cutting, stabbing, and bruising instruments,	35	40	42	45	46	49
Strangulations, - - -	2	5	2	2	2	4
By precipitating and drowning, - - - -	6	16	6	1	4	3
Kicks and blows with the fist, - - - - - -	28	12	21	23	17	26
Fire, - - - - - - - -	..	1	..	1	..	..
Unknown, - - - - -	17	1	2	..	2	2

Quetelet (1842, p. 6)

He not only speaks of the condition of man at the time, but he also hints at the possibility of being able to model such behaviour.

Symmetric Association

- Francis Galton
- Karl Pearson
- Pearson's Chi-squared Statistic

Francis Galton's "Fingerprints"

Consider two categorical variables, A and B. The simplest quested of such variables is whether they are associated with each other. In a very simple form, we address the hypotheses

H_0: A and B are NOT associated (independent)
H_1: A and B are associated

To help address these hypotheses, we "compare" the observed cell values with the cell values that we would expect to get if the rows and columns are independent.

For the (i, j)th cell, the expected cell frequency (if independence were observed) is

$$\text{Expected (i, j)th count} = \frac{\text{(i'th row total)} \times \text{(j'th column total)}}{\text{sample size}}$$

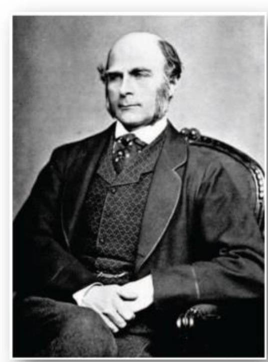
or more formally

$$\frac{n_{i\bullet} n_{\bullet j}}{n}$$

Francis Galton's "Fingerprints"

In fact, Galton (1892, p. 174) was the first (known) to state this result. Part of his work involved determining the association of fingerprint characteristics of 105 fraternal (or dizygotic) male twins. One male twin was "earmarked" as twin A and his brother was twin B.

B children.	A children.			Totals in B children.
	Arches.	Loops.	Whorls.	
Arches . . .	5	12	2	*19*
Loops . . .	4	42	15	*61*
Whorls . . .	1	14	10	*25*
Totals in A children	*10*	*68*	*27*	105



Francis Galton
(1822 – 1911)

Galton (1892, pp. 175 – 176) said

"The squares that run diagonally from the top at the left, to the bottom at the right, contain the double events, and it is with these that we are now concerned. Are entries in those squares larger or not than the randoms . . . The values of 10x19, 68x61, 27x25, all divided by 105?"

$$\frac{n_{i\bullet} n_{\bullet j}}{n}$$

Karl Pearson

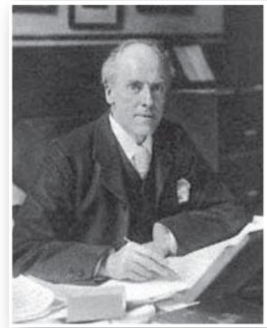
Pearson considered a more general setting than what Galton did and compared all observed cell frequencies with their expected values (under independence) – not just the diagonal elements.

Pearson's approach was to consider looking at the difference between the two:

$$n_{ij} - \frac{n_{i\bullet}n_{\bullet j}}{n}$$

or, equivalently,

$$p_{ij} - p_{i\bullet}p_{\bullet j}$$



Karl Pearson
(1857 – 1936)

Pearson referred to these differences as the cell's *contingency*.

If all of the contingency's are zero then there is *complete independence* between the two categorical variables.

Note: Pearson used the word **compartment** while we now use the word **cell**

The Statistic

Suppose we consider the Galton expectations and tie this in with Pearson's idea of a contingency. The hypothesis

H_0: A and B are NOT associated (independent)
H_1: A and B are associated

can be more formally expressed by

$$H_0:\ n_{ij} = \frac{n_{i\bullet}n_{\bullet j}}{n}$$
$$H_1:\ n_{ij} \neq \frac{n_{i\bullet}n_{\bullet j}}{n}$$

Quantitatively, Pearson proposed the following statistic as a single measure of the strength of the association between the rows and columns of the table

$$X^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{\left(n_{ij} - \frac{n_{i\bullet}n_{\bullet j}}{n}\right)^2}{\frac{n_{i\bullet}n_{\bullet j}}{n}}$$

Here X^2 is a chi-squared random variable with $(I-1)(J-1)$ degrees of freedom.

The Statistic

Several more succinct expressions of this statistic can be derived.

For example . . .

Suppose we express the above null and alternative hypothesis as

$$H_0: \; p_{ij} = p_{i\bullet} p_{\bullet j}$$
$$H_1: \; p_{ij} \neq p_{i\bullet} p_{\bullet j}$$

Then an equivalent expression for Pearson's chi-squared statistic is

$$X^2 = n \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{\left(p_{ij} - p_{i\bullet} p_{\bullet j}\right)^2}{p_{i\bullet} p_{\bullet j}}$$

Here X^2 is also a chi-squared random variable with $(I-1)(J-1)$ degrees of freedom.

Some Properties

- It may seem surprising at first that, in its day, while the classic variance and least squares was known, why didn't Pearson simply consider the sum-of-squares of the contingency's:

$$\sum_{i=1}^{I} \sum_{j=1}^{J} \left(n_{ij} - \frac{n_{i\bullet} n_{\bullet j}}{n}\right)^2 \quad ??$$

To answer this question, suppose we consider n_{ij} to be a Poisson random variable so that

$$E(n_{ij}) = Var(n_{ij}) = \frac{n_{i\bullet} n_{\bullet j}}{n}$$

If there are issues concerning the stability of the expectation/variance equality there are ways in which we can deal with this.

Then normalising the cells leads to, asymptotically,

$$Z_{ij} = \frac{n_{ij} - \frac{n_{i\bullet} n_{\bullet j}}{n}}{\sqrt{\frac{n_{i\bullet} n_{\bullet j}}{n}}} \sim N(0,\ 1)$$

of which the sum-of-squares is his chi-squared statistic (we'll return back to this later).

Some Properties

- The chi-squared statistic remains unchanged even if the rows and/or columns are interchanged, or swapped (Pearson was aware of this)
- The magnitude of the chi-squared statistic is dependent on the sample size, n, selected. Therefore, for a large enough sample size, it is possible to ALWAYS conclude that there exists a statistically significant association between the rows and columns, even if the association is very weak.
- In fact

$$0 \leq X^2 \leq n[\min(I, J) - 1]$$

- So
  - doubling the sample size doubles X^2
  - Increasing the sample size means there will ALWAYS be an n that leads to a statistically significant association

On Dealing with the Sample Size

Pearson's phi-squared statistic

One obvious way of dealing with the impact of the sample size on Pearson's chi-squared statistic is to simply divide it by n:

$$\phi^2 = \frac{X^2}{n} = \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{\left(p_{ij} - p_{i\bullet}p_{\bullet j}\right)^2}{p_{i\bullet}p_{\bullet j}}$$

Pearson (1904, p. 6) referred to this as the *mean-squared contingency*. These days its also called *Pearson's phi-squared statistic* and

- ϕ^2 ranges from 0 (complete independence) to $\min(I, J) - 1$(complete dependence)
- its magnitude is **independent** of the sample size

Other Contingency Based Measures

Pearson's chi-squared statistic

One may derive a number of other measures of association based on Pearson's contingency $p_{ij} - p_{i\bullet}p_{\bullet j}$.

Marcotorchino (1984) discussed a host of less well known measures for an $I \times J$ contingency table, including

Belson's statistic $$B = n^2 \sum_{i=1}^{I}\sum_{j=1}^{J}\left(p_{ij} - p_{i\bullet}p_{\bullet j}\right)^2$$

Jordan's statistic $$J = n \sum_{i=1}^{I}\sum_{j=1}^{J} p_{ij}\left(p_{ij} - p_{i\bullet}p_{\bullet j}\right)^2$$

Variation of Squares $$V = n^2 \sum_{i=1}^{I}\sum_{j=1}^{J}\left(p_{ij} - p_{i\bullet}p_{\bullet j}\right)\left(p_{ij} + p_{i\bullet}p_{\bullet j}\right)$$

- They are all zero when the categorical variables are independent
- B and J are at least zero. V can be negative.
- There is no known distributional property of these measures

15

Example 2: Galton's Fingerprint Data

Pearson's chi-squared statistic

B children.	A children.			Totals in B children.
	Arches.	Loops.	Whorls.	
Arches . . .	5	12	2	19
Loops . . .	4	42	15	61
Whorls . . .	1	14	10	25
Totals in A children	10	68	27	105

```
              Value  MC.P-value
Chi-sq      11.1699       0.031
A.chi-sq    18.8401       0.096
Belson      48.0305       0.195
Jordan       0.0496       0.275
Var.sq     146.9029       0.113
Phi2         0.1064       0.031
```

There might indeed be a reason why the study of these has not continued

16

Asymmetric Association

- Goodman-Kruskal *lambda* Index
- Goodman-Kruskal *tau* Index

A Definition

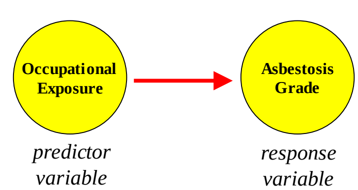
Asymmetric Association

Treat the categorical variables as being *asymmetrically* associated…..

For example,

"How does the number of years of occupational exposure to asbestos impact upon the workers diagnosed level of asbestosis?"

Occupational Exposure (yrs)	None	Asbestos grade Grade 1	Diagnosed Grade 2	Grade 3	Total
0-9	310	36	0	0	346
10-19	212	158	9	0	379
20-29	21	35	17	4	77
30-39	25	102	49	18	194
40+	7	35	51	28	121
Total	575	366	126	50	1117



We now explore two indices that can be used to quantify the asymmetric association between two categorical variables

18

A Definition

Asymmetric Association

So . . .

We can't (shouldn't) use the Pearson chi-squared statistic

(although in many practical and theoretical studies of contingency tables researchers do)

A/B	B_1	B_2	$\cdots$	B_j	$\cdots$	B_J	Total
A_1	n_{11}	n_{12}	$\cdots$	n_{1j}	$\cdots$	n_{1J}	$n_{1\bullet}$
A_2	n_{21}	n_{22}	$\cdots$	n_{2j}	$\cdots$	n_{2J}	$n_{2\bullet}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
A_i	n_{i1}	n_{i2}	$\cdots$	n_{ij}	$\cdots$	n_{iJ}	$n_{i\bullet}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
A_I	n_{I1}	n_{I2}	$\cdots$	n_{Ij}	$\cdots$	n_{IJ}	$n_{I\bullet}$
Total	$n_{\bullet 1}$	$n_{\bullet 2}$	$\cdots$	$n_{\bullet j}$	$\cdots$	$n_{\bullet J}$	n

- Consider two independent events A and B. Then

$$P(A|B) = P(A) \quad \rightarrow \quad P(A|B) - P(A) = 0$$

For a contingency table, *given the columns*, how do the rows compare?

$$\pi_{ij} = \frac{p_{ij}}{p_{\bullet j}} - p_{i\bullet}$$

If $\pi_{ij} = 0$ the j'th column is not a good predictor of the i'th row category

19

Goodman-Kruskal *lambda* Index

Asymmetric Association

So . . .

We can't (shouldn't) use the Pearson chi-squared statistic

(although in many practical and theoretical studies of contingency tables researchers do)

A/B	B_1	B_2	$\cdots$	B_j	$\cdots$	B_J	Total
A_1	n_{11}	n_{12}	$\cdots$	n_{1j}	$\cdots$	n_{1J}	$n_{1\bullet}$
A_2	n_{21}	n_{22}	$\cdots$	n_{2j}	$\cdots$	n_{2J}	$n_{2\bullet}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
A_i	n_{i1}	n_{i2}	$\cdots$	n_{ij}	$\cdots$	n_{iJ}	$n_{i\bullet}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
A_I	n_{I1}	n_{I2}	$\cdots$	n_{Ij}	$\cdots$	n_{IJ}	$n_{I\bullet}$
Total	$n_{\bullet 1}$	$n_{\bullet 2}$	$\cdots$	$n_{\bullet j}$	$\cdots$	$n_{\bullet J}$	n

- There are various measures of *asymmetric association* discussed by Goodman & Kruskal (1954). For A|B (predicting the rows given the columns)
  - Goodman-Kruskal *lambda index*

$$\lambda_{A|B} = \frac{\sum_{j=1}^{J} p_{mj} - p_{m\bullet}}{1 - p_{m\bullet}}$$

where $p_{mj} = \max\{p_{mj}\}$ (largest proportion in the observed level j) and $p_{m\bullet} = \max\{p_{i\bullet}\}$

20

Goodman-Kruskal *tau* Index

Asymmetric Association

So . . .

We can't (shouldn't) use the Pearson chi-squared statistic

(although in many practical and theoretical studies of contingency tables researchers do)

A/B	B_1	B_2	$\cdots$	B_j	$\cdots$	B_J	Total
A_1	n_{11}	n_{12}	$\cdots$	n_{1j}	$\cdots$	n_{1J}	$n_{1\bullet}$
A_2	n_{21}	n_{22}	$\cdots$	n_{2j}	$\cdots$	n_{2J}	$n_{2\bullet}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
A_i	n_{i1}	n_{i2}	$\cdots$	n_{ij}	$\cdots$	n_{iJ}	$n_{i\bullet}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
A_I	n_{I1}	n_{I2}	$\cdots$	n_{Ij}	$\cdots$	n_{IJ}	$n_{I\bullet}$
Total	$n_{\bullet 1}$	$n_{\bullet 2}$	$\cdots$	$n_{\bullet j}$	$\cdots$	$n_{\bullet J}$	n

- There are various measures of *asymmetric association* discussed by Goodman & Kruskal (1954). For A|B (predicting the rows given the columns)
  - Goodman-Kruskal *tau index*

$$\tau_{A|B} = \frac{\sum_{i=1}^{I}\sum_{j=1}^{J}\frac{p_{ij}^2}{p_{\bullet j}} - \sum_{i=1}^{I} p_{i\bullet}^2}{1 - \sum_{i=1}^{I} p_{i\bullet}^2}$$

21

The Index

Goodman-Kruskal tau Index

- Goodman-Kruskal tau index
  - as a weighted sum of squares of the centred conditional proportions

$$\tau_{A|B} = \frac{1}{1 - \sum_{i=1}^{I} p_{i\bullet}^2}\sum_{i=1}^{I}\sum_{j=1}^{J} p_{\bullet j}\left(\frac{p_{ij}}{p_{\bullet j}} - p_{i\bullet}\right)^2$$

- **PROPERTIES**
  - $0 \leq \tau_{A|B} \leq 1$
  - Light & Margolin (1971) showed that

$$C = (n-1)(I-1)\tau_{A|B} \sim X^2_{(I-1)(J-1)}$$

  - Agresti (1990, p.25) notes that low values of tau do not necessarily mean "low" levels of association, since tau tends to take smaller values as the number of categories increase.

22

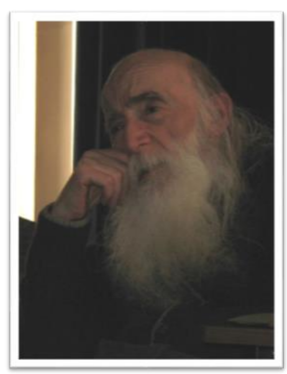
Simple (Symmetrical) Correspondence Analysis

- Benzécri, Greenacre
- Profiles Reciprocal Averaging and the Triplet
- Singular Value Decomposition
- Correspondence Plot and the Biplot

Origins: Jean-Paul Benzécri

The 1960's saw the advances in categorical data analysis take on a geometric form with the development of correspondence analysis.

The "father" of modern day correspondence analysis is French linguist Jean-Paul Benzecri, and with his team of researchers, developed its foundations at the Mathematical Statistics Laboratory, Faculty of Science in Paris, France.



Jean-Paul Benzécri
Paris, 2011

As a result the method of *l'analyse des correspondances*, as coined by Benzécri, is very popular in France. The popularity of correspondence analysis in France resulted in a journal dedicated to the development and application of the technique as well as methods of classification, *Cahiers de l'Analyse des Données*, founded by Benzecri (1976 – 1997) . . . http://www.numdam.org/journals/CAD/

Simple Correspondence Analysis

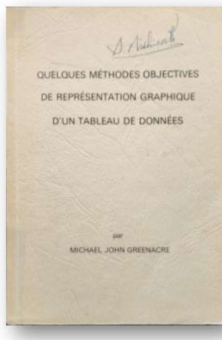
24

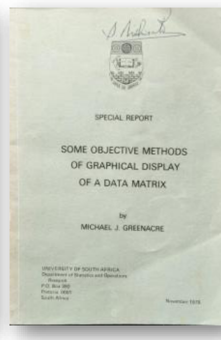
Origins: Michael J. Greenacre

Simple Correspondence Analysis

- Michael Greenacre
  Universitat Pompeu Fabra, Barcelona, Spain
- Former student of Benzécri

  - Greenacre, M. J. (1978), Quelques methodes objectives de representation graphique d'un tableau de donnes, Unpublished PhD thesis, Universite Pierre et Marie Curie, Paris
    - Translation: *Some objective methods of graphical display of a data matrix*





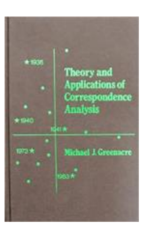
25

Origins: Michael J. Greenacre

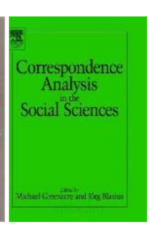
Simple Correspondence Analysis

- Michael Greenacre
  Universitat Pompeu Fabra, Barcelona, Spain
- Former student of Benzécri

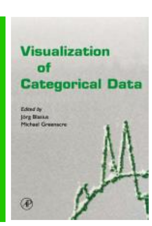
  - Greenacre, M. J. (1978), Quelques methodes objectives de representation graphique d'un tableau de donnes, Unpublished PhD thesis, Universite Pierre et Marie Curie, Paris
    - Translation: *Some objective methods of graphical display of a data matrix*



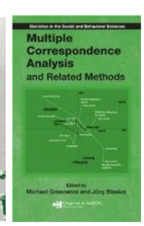
(1984)



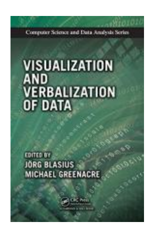
(1994)



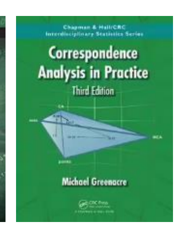
(1998)



(2006)



(2014)

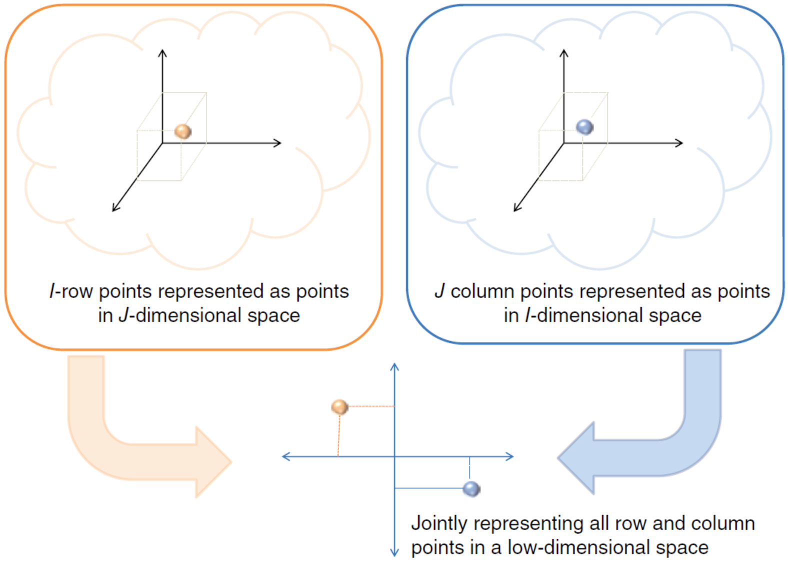


(2017)

You are also invited to consider Beh and Lombardo (2012, 2019) who provide an extensive bibliography on the history and development of correspondence analysis.

26

Dimension Reduction



Profiles

Suppose we have following two row categories with five entries each (so there are five categories)

	Col 1	Col 2	Col 3	Col 4	Col 5	Total
Row 1	2	4	6	8	10	30
Row 2	20	40	60	80	100	300
. . .	. . .	. . .	. . .	. . .	. . .	. . .

For the first row – "Row 1" – it has a total of 30 classifications, while for the second its 300. While the totals of each row are different, their relative proportions are identical:

$$\left(\frac{1}{15,} \frac{2}{15,} \frac{3}{15,} \frac{4}{15,} \frac{5}{15,}\right)$$

This array of relative cell frequencies is referred to as a *profile*.

Profiles

Simple Correspondence Analysis

In the more general case, for the i'th *row profile* is

$$\left(\frac{n_{i1}}{n_{i\bullet}}, \frac{n_{i2}}{n_{i\bullet}}, \cdots, \frac{n_{ij}}{n_{i\bullet}}, \cdots, \frac{n_{iJ}}{n_{i\bullet}}\right) = \left(\frac{p_{i1}}{p_{i\bullet}}, \frac{p_{i2}}{p_{i\bullet}}, \cdots, \frac{p_{ij}}{p_{i\bullet}}, \cdots, \frac{p_{iJ}}{p_{i\bullet}}\right)$$

A/B	B_1	B_2	$\cdots$	B_j	$\cdots$	B_J	Total
A_1	n_{11}	n_{12}	$\cdots$	n_{1j}	$\cdots$	n_{1J}	$n_{1\bullet}$
A_2	n_{21}	n_{22}	$\cdots$	n_{2j}	$\cdots$	n_{2J}	$n_{2\bullet}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
A_i	n_{i1}	n_{i2}	$\cdots$	n_{ij}	$\cdots$	n_{iJ}	$n_{i\bullet}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
A_I	n_{I1}	n_{I2}	$\cdots$	n_{Ij}	$\cdots$	n_{IJ}	$n_{I\bullet}$
Total	$n_{\bullet 1}$	$n_{\bullet 2}$	$\cdots$	$n_{\bullet j}$	$\cdots$	$n_{\bullet J}$	n

29

Profiles

Simple Correspondence Analysis

In the more general case, for the i'th *row profile* is

$$\left(\frac{n_{i1}}{n_{i\bullet}}, \frac{n_{i2}}{n_{i\bullet}}, \cdots, \frac{n_{ij}}{n_{i\bullet}}, \cdots, \frac{n_{iJ}}{n_{i\bullet}}\right) = \left(\frac{p_{i1}}{p_{i\bullet}}, \frac{p_{i2}}{p_{i\bullet}}, \cdots, \frac{p_{ij}}{p_{i\bullet}}, \cdots, \frac{p_{iJ}}{p_{i\bullet}}\right)$$

A/B	B_1	B_2	$\cdots$	B_j	$\cdots$	B_J	Total
A_1	n_{11}	n_{12}	$\cdots$	n_{1j}	$\cdots$	n_{1J}	$n_{1\bullet}$
A_2	n_{21}	n_{22}	$\cdots$	n_{2j}	$\cdots$	n_{2J}	$n_{2\bullet}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
A_i	n_{i1}	n_{i2}	$\cdots$	n_{ij}	$\cdots$	n_{iJ}	$n_{i\bullet}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
A_I	n_{I1}	n_{I2}	$\cdots$	n_{Ij}	$\cdots$	n_{IJ}	$n_{I\bullet}$
Total	$n_{\bullet 1}$	$n_{\bullet 2}$	$\cdots$	$n_{\bullet j}$	$\cdots$	$n_{\bullet J}$	n

Similarly, the j'th *column profile* is

$$\left(\frac{n_{1j}}{n_{\bullet j}}, \frac{n_{2j}}{n_{\bullet j}}, \cdots, \frac{n_{ij}}{n_{\bullet j}}, \cdots, \frac{n_{Ij}}{n_{\bullet j}}\right) = \left(\frac{p_{1j}}{p_{\bullet j}}, \frac{p_{2j}}{p_{\bullet j}}, \cdots, \frac{p_{ij}}{p_{\bullet j}}, \cdots, \frac{p_{Ij}}{p_{\bullet j}}\right)$$

30

Profiles

If there is **no association (independence)** between the row and column variables then these profiles simplify to

$$\left(\frac{n_{\bullet 1}}{n}, \frac{n_{\bullet 2}}{n}, \cdots, \frac{n_{\bullet j}}{n}, \cdots, \frac{n_{\bullet J}}{n}\right) = (p_{\bullet 1}, p_{\bullet 2}, \cdots, p_{\bullet j}, \cdots, p_{\bullet J})$$

and

$$\left(\frac{n_{1\bullet}}{n}, \frac{n_{2\bullet}}{n}, \cdots, \frac{n_{i\bullet}}{n}, \cdots, \frac{n_{I\bullet}}{n}\right) = (p_{1\bullet}, p_{2\bullet}, \cdots, p_{i\bullet}, \cdots, p_{I\bullet})$$

respectively.

This suggests we may alternatively consider the *centred* row and centred column profiles as a means of detecting any departures from independence.

Centred Row Profiles

The i'th *centred row profile* element is

$$\left(\frac{p_{i1}}{p_{i\bullet}} - p_{\bullet 1}, \frac{p_{i2}}{p_{i\bullet}} - p_{\bullet 2} \;\cdots\;, \frac{p_{iJ}}{p_{i\bullet}} - p_{\bullet J}\right)$$

Note these centred row profiles are **centred around zero** so that:

$$\begin{aligned}\sum_{j=1}^{J}\left(\frac{p_{ij}}{p_{i\bullet}} - p_{\bullet j}\right) &= \frac{1}{p_{i\bullet}}\sum_{j=1}^{J} p_{ij} - \sum_{j=1}^{J} p_{\bullet j} \\ &= \frac{1}{p_{i\bullet}} p_{i\bullet} - 1 \\ &= 0\end{aligned}$$

Centred Column Profiles

Simple Correspondence Analysis

The j'th *centred column profile* element is

$$\left(\frac{p_{1j}}{p_{\bullet j}} - p_{1\bullet}, \frac{p_{2j}}{p_{\bullet j}} - p_{2\bullet} \; \cdots \;, \frac{p_{Ij}}{p_{\bullet j}} - p_{I\bullet}\right)$$

Note these centred column profiles are **centred around zero** so that:

$$\sum_{i=1}^{I}\left(\frac{p_{ij}}{p_{\bullet j}} - p_{i\bullet}\right) = \frac{1}{p_{\bullet j}}\sum_{i=1}^{J} p_{ij} - \sum_{i=1}^{I} p_{\bullet j}$$

$$= \frac{1}{p_{\bullet j}} p_{\bullet j} - 1$$

$$= 0$$

33

Example: Sekiloff's Asbestos Data

Simple Correspondence Analysis

Centred Row & Column Profile Matrices

Occupational Exposure (yrs)	None	Asbestos grade Grade 1	Diagnosed Grade 2	Grade 3	Total
0-9	310	36	0	0	346
10-19	212	158	9	0	379
20-29	21	35	17	4	77
30-39	25	102	49	18	194
40+	7	35	51	28	121
Total	575	366	126	50	1117

```
> solve(dI)%*%P - rep(1, times= 5)%*%t(apply(P, 2, sum))
           None      Grade 1     Grade 2      Grade 3
[1,]  0.38118205 -0.22361714 -0.11280215 -0.044762757
[2,]  0.04459504  0.08922316 -0.08905545 -0.044762757
[3,] -0.24204444  0.12688207  0.10797707  0.007185295
[4,] -0.38590573  0.19810981  0.13977517  0.048020748
[5,] -0.45692047 -0.03840719  0.30868545  0.186642201
>
> solve(dJ)%*%t(P) - rep(1, times= 4)%*%t(apply(P, 1, sum))
            0-9       10-19       20-29      30-39       40+
[1,]  0.2293722  0.02939395 -0.03241291 -0.1302012 -0.09615196
[2,] -0.2113976  0.09239229  0.02669377  0.1050090 -0.01269746
[3,] -0.3097583 -0.26787313  0.06598599  0.2152094  0.29643603
[4,] -0.3097583 -0.33930170  0.01106535  0.1863205  0.45167413
>
```

Centred Row Profiles

Centred Column Profiles

34

Scoring the Categories

The key objective is to find the set of row **scores**

$$\mathbf{a} = (a_1, a_2, \dots, a_i, \dots, a_I)^T$$

and the set of column scores

$$\mathbf{b} = (b_1, b_2, \dots, b_j, \dots, b_J)^T$$

so that the correlation between **a** and **b** is maximised.

These scores can be found using a technique called *reciprocal averaging* (also *dual scaling*, *optimal scaling*, *homogeneity analysis* and other terms) and is related to *canonical correlation analysis*.

We won't go into any great detail.

We could treat *ordinal* categorical variables by imposing a constraint that

$$a_1 < a_2 < \cdots < a_I$$

(for increasing row categories, say) but we won't do that here.

The Triplet

But we do impose the property that

$$\sum_{i=1}^{I} p_{i\bullet} a_i = 0 \qquad \sum_{i=1}^{I} p_{i\bullet} a_i^2 = 1$$

$$\sum_{j=1}^{J} p_{\bullet j} b_j = 0 \qquad \sum_{j=1}^{J} p_{\bullet j} b_j^2 = 1$$

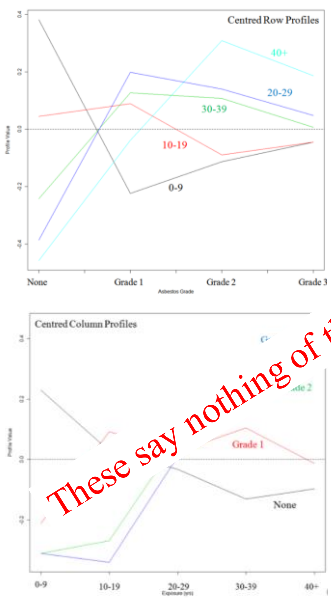
To ensure we get a solution (which we will always do).
The correlation between **a** and **b** is

$$\lambda = \sum_{i=1}^{I} \sum_{j=1}^{J} p_{ij} a_i b_j$$

Note: we are only talking about a *one-dimensional* solution for the row and column scores. Soon we will turn our attention to a *multi-dimensional* solution.

Example: Selikoff's Asbestos Data

Simple Correspondence Analysis



Years of exposure:

- “20 – 29” and “30 – 39” a… …ilarly distributed
- “0 – 9” is distr… …erently to “40+”
- “10 – 1… …vary much when co… …h independence

…e grades of asbestosis severity

Asbestosis severity levels

- “Grade 2” and “Grade 3” appear similarly distributed
- “None” is distributed differently to “Grade 2” and “Grade 3”

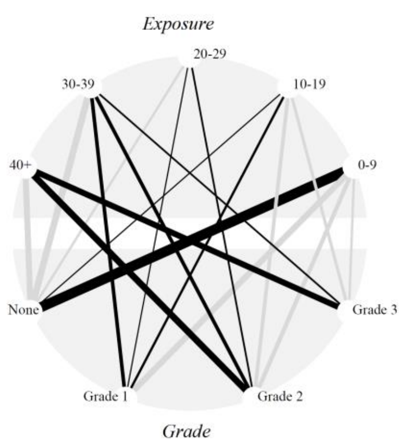
across the different years of exposure

These say nothing of the association **between** the variables

Example: Selikoff's Asbestos Data

Simple Correspondence Analysis

Cobweb Diagram



By using standardised residuals

- Black line indicates *positive* association
- Grey line indicates a *negative* association.
- Thick lines indicate a strong association*.
- Thin lines weak association*.

*Positive and negative association

Using the R function `cobweb()` that appears in the Appendix of Upton (2017)

Generalised Singular Value Decomposition

Simple Correspondence Analysis

Correspondence analysis may be performed using the *generalised singular value decomposition* (GSVD) of the matrix of *Pearson residuals*

For rectangular matrix of Pearson residuals, **Z**,

$$\mathbf{Z} = \mathbf{D}_I^{-1}\left(\mathbf{P} - \mathbf{r}\,\mathbf{c}^T\right)\mathbf{D}_J^{-1}$$

the GSVD of **Z** is

$$\mathbf{Z} = \mathbf{A}\mathbf{D}_\lambda\mathbf{B}^T$$

where

$$\mathbf{A}^T\mathbf{D}_I\mathbf{A} = \mathbf{I}_M \qquad \mathbf{B}^T\mathbf{D}_J\mathbf{B} = \mathbf{I}_M$$

and

$$\mathbf{r} = (p_{1\bullet}, p_{2\bullet}, \ldots, p_{I\bullet})^T$$
$$\mathbf{c} = \left(p_{\bullet 1}, p_{\bullet 2}, \ldots, p_{\bullet J}\right)^T$$

Note that the (i, j)th element of **Z** is $\quad Z_{ij} = \dfrac{p_{ij} - p_{i\bullet}p_{\bullet j}}{p_{i\bullet}p_{\bullet j}}$

39

Generalised Singular Value Decomposition

Simple Correspondence Analysis

Correspondence analysis may be performed using the *generalised singular value decomposition* (GSVD) of the matrix of *Pearson residuals*

For rectangular matrix of Pearson residuals, **Z**,

$$\mathbf{Z} = \mathbf{D}_I^{-1}\left(\mathbf{P} - \mathbf{r}\,\mathbf{c}^T\right)\mathbf{D}_J^{-1}$$

the GSVD of **Z** is

$$\mathbf{Z} = \mathbf{A}\mathbf{D}_\lambda\mathbf{B}^T$$

where

$$\mathbf{A}^T\mathbf{D}_I\mathbf{A} = \mathbf{I}_M \qquad \mathbf{B}^T\mathbf{D}_J\mathbf{B} = \mathbf{I}_M$$

while

$$\mathbf{D}_\lambda = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_M)$$

contains elements that are the singular values of **Z** so that

$$\phi^2 = \frac{X^2}{n} = \sum_{m=1}^{M} \lambda_m^2$$

40

Standard Coordinates

Suppose we consider the left singular vectors which are the scores for the *row categories*

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} & \cdots & a_{1M} \\ a_{21} & a_{22} & \cdots & a_{2m} & \cdots & a_{2M} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{im} & \cdots & a_{iM} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{I1} & a_{I2} & \cdots & a_{Im} & \cdots & a_{IM} \end{pmatrix}$$

The i'th row of **A** can be used as the coordinate of the i'th row category in an $M = \min(I, J) - 1$ dimensional space

These are referred to as row *standard coordinates*

$$\mathbf{F} = \mathbf{A} \quad \Rightarrow \quad f_{im} = a_{im}$$

Standard Coordinates

Suppose we consider the right singular vectors which are the scores for the *column categories*

$$\mathbf{B} = \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1m} & \cdots & b_{1M} \\ b_{21} & b_{22} & \cdots & b_{2m} & \cdots & b_{2M} \\ \vdots & \vdots & & \vdots & & \vdots \\ b_{j1} & b_{j2} & \cdots & b_{jm} & \cdots & b_{jM} \\ \vdots & \vdots & & \vdots & & \vdots \\ b_{J1} & b_{J2} & \cdots & b_{Jm} & \cdots & b_{JM} \end{pmatrix}$$

The j'th column of **B** can be used as the coordinate of the j'th column category in an $M = \min(I, J) - 1$ dimensional space.

These are referred to as *column standard coordinates*

$$\mathbf{G} = \mathbf{B} \quad \Rightarrow \quad g_{jm} = b_{jm}$$

Principal Coordinates

We can graphically depict the association between the row and column categories by considering the following columns of the following matrices of coordinates

$$\tilde{\mathbf{F}} = \mathbf{AD}_{\lambda} \qquad \text{(Row Principal Coordinates)}$$

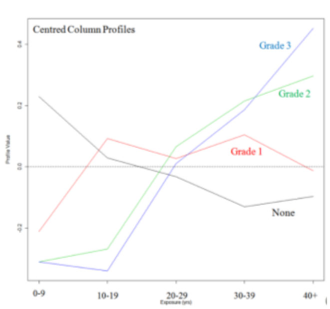
$$\tilde{\mathbf{G}} = \mathbf{BD}_{\lambda} \qquad \text{(Column Principal Coordinates)}$$

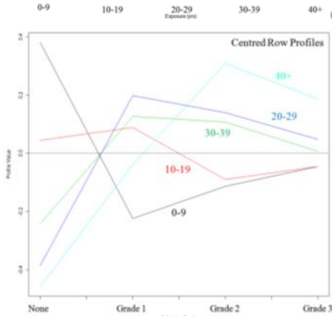
These are referred to as *principal coordinates*.

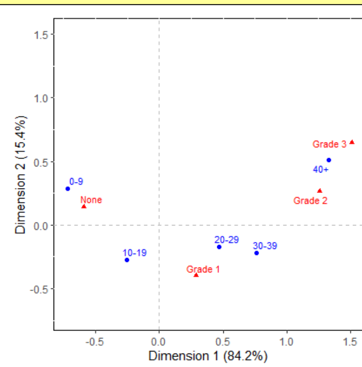
Note that

- $\tilde{\mathbf{F}}$ is an I × M column matrix ⇒ $\tilde{f}_{im} = a_{im}\lambda_m$ is the (i, m)'th value
  - is the principal coordinate of the i'th row along the m'th dimension of the plot
- $\tilde{\mathbf{G}}$ is a J × M column matrix ⇒ $\tilde{g}_{jm} = b_{jm}\lambda_m$ is the (j, m)'th value
  - is the principal coordinate of the j'th column along the m'th dimension of the plot

Example: Selikoff's Asbestos Data







The technique used to obtain this graphical summary of the association between the categorical variables is called *correspondence analysis*.

This plot is commonly referred to as a *correspondence plot*.

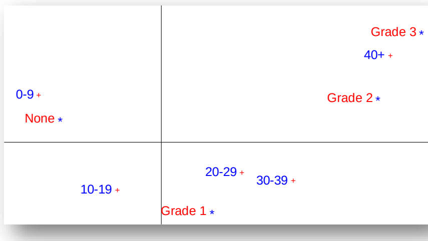
An M-dimensional plot is commonly called an *optimal correspondence plot*

Example: Selikoff's Asbestos Data

Simple Correspondence Analysis

```
> ca.out <- CAvariants(asbestos)
> plot(ca.out, plottype = "classic")

    RESULTS for CA Correspondence Analysis:
```



Inertias, percent inertias and cumulative percent inertias of the row space

	inertia	inertiapc	cuminertiapc
1	0.489	84.22	84.22
2	0.089	15.35	99.57
3	0.003	0.43	100.00
4	0.000	0.00	100.00

$X^2/n = 0.581$

Inertia values = Eigenvalues

Explained inertia of each axis => contribution of each axis to chi-squared

First axis: 84.2%,
Second axis: 15.4%

Cumulative contribution of each axis to chi-squared

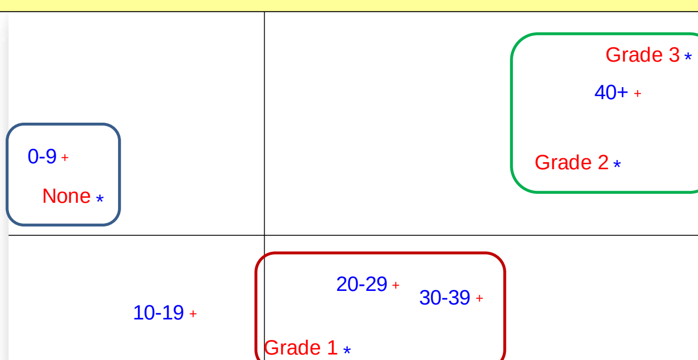
One-dimension : 84.22%,
Two-dimensions: 99.57%

Three-dimensions: 100.00%
(optimal correspondence plot)

45

Example: Selikoff's Asbestos Data

Simple Correspondence Analysis

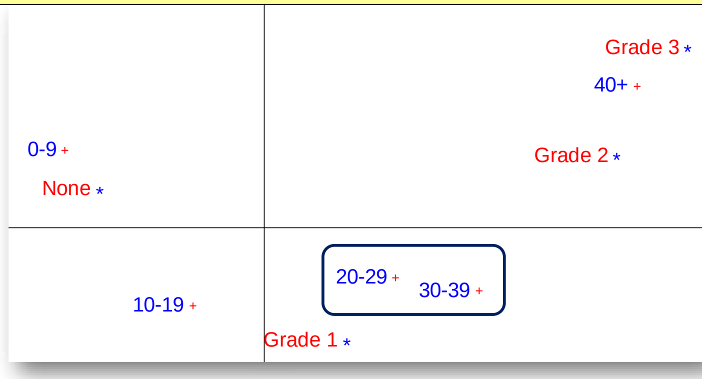


- Less than 10 years of occupational exposure and no asbestosis are associat...
- Mild asbestosis is associat... ...ccupational exposure to asbestos
- ...e grades of asbestosis are associated with at least 40 years of occupational exposure to asbestos

Note that we are NOT implying a CAUSAL association here

46

Example: Selikoff's Asbestos Data



Distance between two row points when in principal coordinates

$$d_I^2(i,i') = \sum_{j=1}^{J} \frac{1}{p_{\bullet j}} \left(\frac{p_{ij}}{p_{i\bullet}} - \frac{p_{i'j}}{p_{i'\bullet}} \right)^2 = \sum_{j=1}^{J} \frac{1}{p_{\bullet j}} \left[\left(\frac{p_{ij}}{p_{i\bullet}} - p_{\bullet j} \right) - \left(\frac{p_{i'j}}{p_{i'\bullet}} - p_{\bullet j} \right) \right]^2 = \sum_{m=1}^{M} \left(\tilde{f}_{im} - \tilde{f}_{i'm} \right)^2$$

Cannot measure the distance between a row point and a column point

Simple Correspondence Analysis

47

Biplot Coordinates

We can graphically depict the association between the row and column categories by considering the following matrices of coordinate

- $\tilde{\mathbf{F}} = \mathbf{AD}_\lambda \Rightarrow \tilde{f}_{im} = a_{im}\lambda_m$ (Row Principal Coordinates)
- $\mathbf{G} = \mathbf{B} \Rightarrow g_{jm} = b_{jm}$ (Column Standard Coordinates)

These are referred to as *row (isometric) biplot coordinates.*

Similarly, for the *column (isometric) biplot coordinates:*

- $\mathbf{F} = \mathbf{A} \Rightarrow f_{im} = a_{im}$ (Row Standard Coordinates)
- $\tilde{\mathbf{G}} = \mathbf{BD}_\lambda \Rightarrow \tilde{g}_{jm} = b_{jm}\lambda_m$ (Column Principal Coordinates)

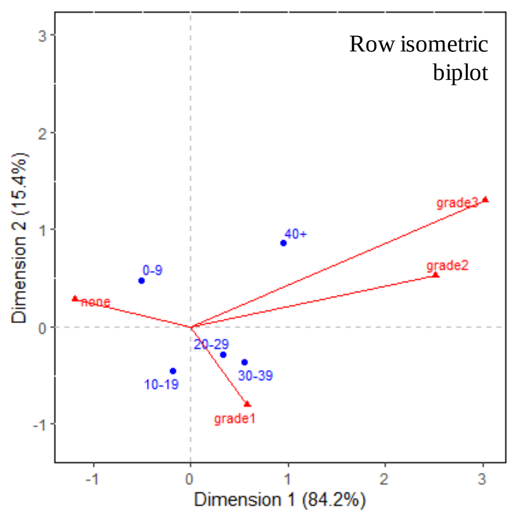
Simple Correspondence Analysis

48

Example: Selikoff's Asbestos Data

Simple Correspondence Analysis

```
> ris <- CAvariants(asbestos, catype = "CA")
> plot(ris, plottype = "biplot", biptype = "row", scaleplot = 2)
```



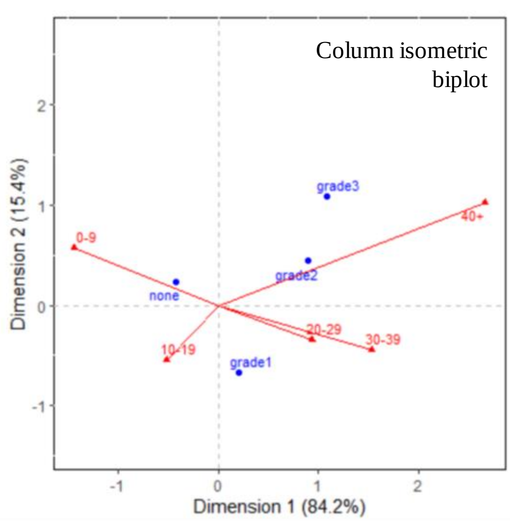
- 0 – 9yrs of exposure **strongly** associated with no asbestosis
- 10-19yrs of exposure equally likely to be associated with no or lowest grade asbestosis
- 20 – 39 years of exposure associated with Grade 1 asbestosis
- Most severe grade of asbestosis **strongly** linked to 40+yrs of exposure
- *Strong Association between **Grade 3** and 40+yrs exposure*

49

Example: Selikoff's Asbestos Data

Simple Correspondence Analysis

```
> ris <- CAvariants(asbestos, catype = "CA")
> plot(ris, plottype = "biplot", biptype = "column", scaleplot = 2)
```



- 0 – 9yrs of exposure **strongly** associated with no asbestosis
- Grade 1 asbestosis equally linked to 10 – 39 years of exposure to asbestos fibres
- 20 – 29yrs and 30-39 yrs of exposure contributes equally to the association
- 40+yrs of exposure to asbestos fibres **strongly** linked to Grade 2 asbestosis
- *Strong Association between **Grade 2** and 40+yrs exposure*

50

Non-symmetrical Correspondence Analysis

Singular Value Decomposition and the Quantification of the Variables

Origins: D'Ambra & Lauro

- Not only we can quantify the asymmetric association between two categorical variables, but we can also visualise it.
- For two symmetrically associated variables, we discussed (simple) correspondence analysis
- For two asymmetrically associated variables, we now discuss non-symmetrical correspondence analysis (NSCA)
- For NSCA, the total inertia of a contingency table isn't measured in terms of X^2. Instead we consider either
  - Goodman-Kruskal τ index, or
  - The C – statistic

Developed by Lauro & D'Ambra (1984) and D'Ambra & Lauro (1989). Discussed in detail by Kroonenberg & Lombardo (1999), Lombardo, Beh & D'Ambra (2007), Beh & Lombardo (2014, Chp 5; 2021, Chp 3), Lombardo, Kroonenberg & Beh (2016) and Lombardo, Beh & Kroonenberg (2021)

Prof Luigi D'Ambra
Uni of Naples

Prof Carlo Lauro
Uni of Naples

A Recap

Non-Symmetrical Correspondence Analysis

Suppose we use the Goodman-Kruskal index, $\tau_{A|B}$, ("rows given columns") as our measure of association. Recall that this implies that we are treating the column variable and row variable as our predictor and response variable, respectively.

The Goodman – Kruskal tau index is of the form

$$\tau_{A|B} = \frac{1}{1 - \sum_{i=1}^{I} p_{i\bullet}^2} \sum_{i=1}^{I}\sum_{j=1}^{J} p_{\bullet j}\left(\frac{p_{ij}}{p_{\bullet j}} - p_{i\bullet}\right)^2$$

Goodman & Krusal (1954)

Since the term out front does not reflect the association captured (via the cell proportions) we can ignore it. Doing so, we have the statistic

$$\tau_{num} = \sum_{i=1}^{I}\sum_{j=1}^{J} p_{\bullet j}\left(\frac{p_{ij}}{p_{\bullet j}} - p_{i\bullet}\right)^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\left[\sqrt{p_{\bullet j}}\left(\frac{p_{ij}}{p_{\bullet j}} - p_{i\bullet}\right)\right]^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\tilde{\pi}_{ij}^2$$

where τ_{num} is the numerator of $\tau_{A|B}$, and

$$\tilde{\pi}_{ij} = \sqrt{p_{\bullet j}}\left(\frac{p_{ij}}{p_{\bullet j}} - p_{i\bullet}\right) = \pi_{ij}\sqrt{p_{\bullet j}}$$

53

Generalised Singular Value Decomposition

Non-Symmetrical Correspondence Analysis

In matrix, notation, π_{ij} is the (i, j)'th element of the matrix

$$\mathbf{\Pi} = \mathbf{D}_J^{-1}\mathbf{P}^T - \mathbf{1}_J\,\mathbf{r}^T$$

Then NSCA can be performed by applying a GSVD to $\mathbf{\Pi}$ such that

$$\mathbf{D}_J^{-1}\mathbf{P}^T - \mathbf{1}_J\,\mathbf{r}^T = \mathbf{A}\mathbf{D}_\lambda\mathbf{B}^T$$

where

$$\tilde{\mathbf{A}}^T\tilde{\mathbf{A}} = \mathbf{I}_M \qquad \mathbf{B}^T\mathbf{D}_J\mathbf{B} = \mathbf{I}_M$$

and $\mathbf{1}_J$ is a vector of 1's of length J so that

$$\tau_{num} = \sum_{m=1}^{M}\lambda_m^2$$

predictor

response

Asbestos grade Diagnosed	Occupational exposure (years) 0–9	10–19	20–29	30–39	40+	Total
None	310	212	21	25	7	575
Grade 1	36	158	35	102	35	366
Grade 2	0	9	17	49	51	126
Grade 3	0	0	4	18	28	50
Total	346	379	77	194	121	1117

54

Principal & Biplot Coordinates

Non-Symmetrical Correspondence Analysis

- *Principal Coordinates*

  $\tilde{\mathbf{F}} = \tilde{\mathbf{A}}\mathbf{D}_\lambda$ (Row Coordinates)

  $\tilde{\mathbf{G}} = \mathbf{B}\mathbf{D}_\lambda$ (Column Coordinates)

- *Column Biplot Coordinates*

  (Correct way to portray the asymmetric association of rows *given* columns)

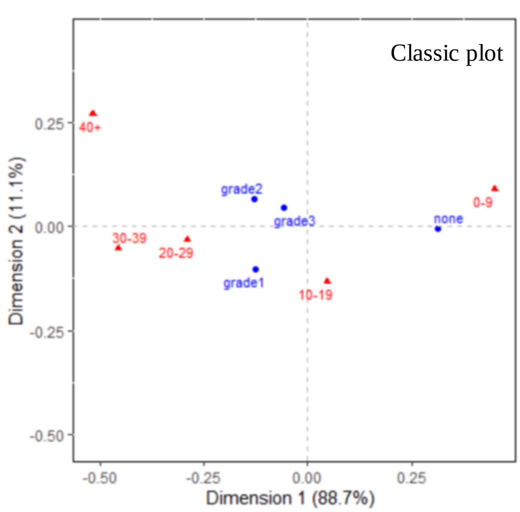
- $\mathbf{F} = \mathbf{A}$ $\Rightarrow$ $f_{im} = a_{im}$ (Row Standard Coordinates)
- $\tilde{\mathbf{G}} = \mathbf{B}\mathbf{D}_\lambda$ $\Rightarrow$ $\tilde{g}_{jm} = b_{jm}\,\lambda_m$ (Column Principal Coordinates)

55

Example: Selikoff's Asbestos Data

Non-Symmetrical Correspondence Analysis

```
> ris <- CAvariants(asbestos, catype = "NSCA")
> plot(ris, plottype = "classic")
```



Given that a worker is exposed to asbestos fibres for

- **Less than 10yrs** they are unlikely to be diagnosed with asbestosis
- **10 – 19** yrs, they are likely to be diagnosed with **no**, or **Grade 1** asbestosis
- 20 – 29yrs, they are likely to be diagnosed with any Grade of asbestosis
- 40+ yrs, they are likely to be diagnosed with the two most severe grades of asbestosis

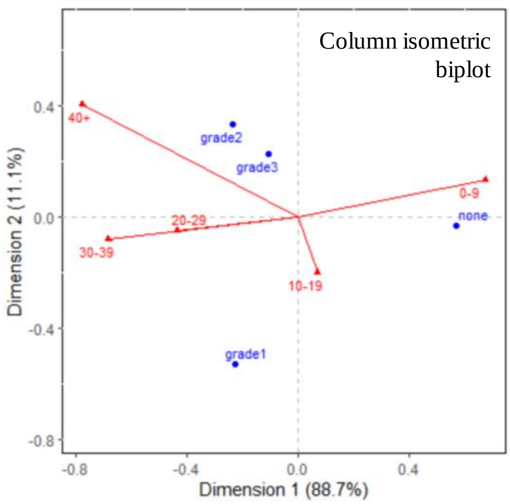
Predictor = *Grade* (columns)
Response = *Years* (rows)

56

Example: Selikoff's Asbestos Data

Non-Symmetrical Correspondence Analysis

```
> ris <- CAvariants(asbestos, catype = "NSCA")
> plot(ris, plottype = "biplot", biptype = "column", scaleplot = 1.5)
```



Given that a worker is exposed to asbestos fibres for

- **Less than 10yrs** they are unlikely to be diagnosed with asbestosis
- **10 – 19** yrs, they are likely to be diagnosed with **no**, or **Grade 1** asbestosis
- 20 – 29yrs, they are likely to be diagnosed with any Grade of asbestosis
- 40+ yrs, they are likely to be diagnosed with the two most severe grades of asbestosis

Predictor = *Grade* (columns)
Response = *Years* (rows)

57

Ordinal Correspondence analysis

Bivariate Moment Decomposition
and the
Quantification of the Variables

Some Approaches

Ordinal Correspondence Analysis

When it comes to ordinal categorical variables many approaches force the scores to be ordered.

- For example, for *increasingly ordered row categories*, the elements of

$$\mathbf{a} = (a_1, a_2, \dots, a_i, \dots, a_I)^T$$

  are arranged so that they increase: $a_1 < a_2, < \cdots < a_i < \cdots < a_I$
- Schriever (1983), Ritov & Gilula (1993), Parsa & Smith (1993), Yang & Huh (1999)

Two problems:

- For an M-dimensional solution, which set of scores are re-arranged. Typically, the scores along the first dimension, but why not scores along other dimensions?
- Forcing the scores to be ordered forces the "correspondences" to behave in a certain way (a_1 will ALWAYS be associated with b_1)

59

Some Approaches

Ordinal Correspondence Analysis

When it comes to ordinal categorical variables many approaches force the scores to be ordered.

- For example, for *increasingly ordered row categories*, the elements of

$$\mathbf{a} = (a_1, a_2, \dots, a_i, \dots, a_I)^T$$

  are arranged so that they increase: $a_1 < a_2, < \cdots < a_i < \cdots < a_I$
- Schriever (1983), Ritov & Gilula (1993), Parsa & Smith (1993), Yang & Huh (1999)

Instead of doing this . . .

- Don't perform a SVD on the matrix of Pearson residuals and force the scores to be arranged in a particular way, instead . . .
- Perform a *bivariate moment decomposition* on the matrix. Consists of
  - *Orthogonal polynomials* instead of singular vectors
  - *Generalised correlations* instead of singular values

60

Bivariate Moment Decomposition

We have a *doubly ordered* contingency table; ordinal row & column variable.

Rather than applying an SVD to *Pearson's residuals*

$$\mathbf{Z} = \mathbf{D}_I^{-1}(\mathbf{P} - \underline{\mathbf{r}}\,\underline{\mathbf{c}}^T)\mathbf{D}_J^{-1}$$

apply a BMD so that

$$\mathbf{Z} = \boldsymbol{\mathcal{A}}\mathbf{G}\boldsymbol{\mathcal{B}}^T$$

where $\boldsymbol{\mathcal{A}}^T\mathbf{D}_I\boldsymbol{\mathcal{A}} = \mathbf{I}_I$ and $\boldsymbol{\mathcal{B}}^T\mathbf{D}_J\boldsymbol{\mathcal{B}} = \mathbf{I}_J$. (More details: Beh 1997, Beh & Lombardo 2014).

Here,

- $\boldsymbol{\mathcal{A}}$ is the $I \times (I-1)$ matrix of *row orthogonal polynomials*
- $\boldsymbol{\mathcal{B}}$ is the $J \times (J-1)$ matrix of *column orthogonal polynomials*
- $\mathbf{G}$ is a $(I-1) \times (J-1)$ rectangular matrix of *generalised correlations* with (u,v)'th element G_{uv}; $u = 1, 2, \dots, I-1$ and $v = 1, 2, \dots, J-1$
- The chi-squared statistic is $X^2 = n\sum_{i=1}^{I-1}\sum_{j=1}^{J-1} G_{uv}^2$ (Lancaster, 1953)

Example: Selikoff's Asbestos Data

Occupational Exposure (yrs)	None	Asbestos grade Grade 1	Diagnosed Grade 2	Grade 3	Total
0-9	310	36	0	0	346
10-19	212	158	9	0	379
20-29	21	35	17	4	77
30-39	25	102	49	18	194
40+	7	35	51	28	121
Total	575	366	126	50	1117

Features

- The row categories are increasing in order
- The column categories are increasing in order
- We shall use *natural scores* to reflect the order of both variables
- So that Pearson's product moment correlation, and it's non-linear variants, are used to assess the structure of the association

Generalised Correlations

Ordinal Correspondence Analysis

The matrix **G** is NOT diagonal. It contains generalised correlations with (u, v)'th element

$$G_{uv} = \sum_{i=1}^{I}\sum_{j=1}^{J} p_{ij} a_u(i) b_v(j)$$

- G_{12} is the linear-by-quadratic correlation: association between row location differences and column dispersion differences
- G_{21} is the quadratic-by-linear correlation: association between row dispersion differences and column location differences
- G_{22} is the quadratic-by-quadratic correlation: association between row and column dispersion differences
- Higher order correlations tend to be small, harder to interpret

Note: $G_{uv}\sqrt{n}$ is asymptotically standard normally distributed

More info: Rayner & Best (1996), Best & Rayner (1996), Rayner & Beh (2009)

63

Example: Selikoff's Asbestos Data

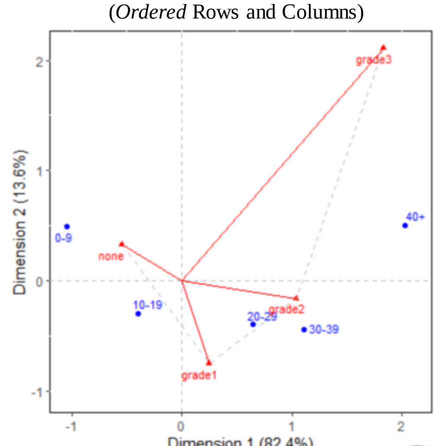
Ordinal Correspondence Analysis

```
> ris = CAvariants(asbestos, catype = "DOCA")
> plot(ris, plottype = "biplot", biptype = "row", scaleplot = 1.5, invproj = F)
```

Interpretation

- Similar to nominal analysis but
- . . . generalised correlations:
  - $\mathbf{G_{11} = 0.689}$
    (Pearson's product moment correlation)
  - $G_{12} = -0.085$
    (linear-by-quadratic)
  - $G_{21} = 0.024$

Row Isometric Biplot
(*Ordered* Rows and Columns)



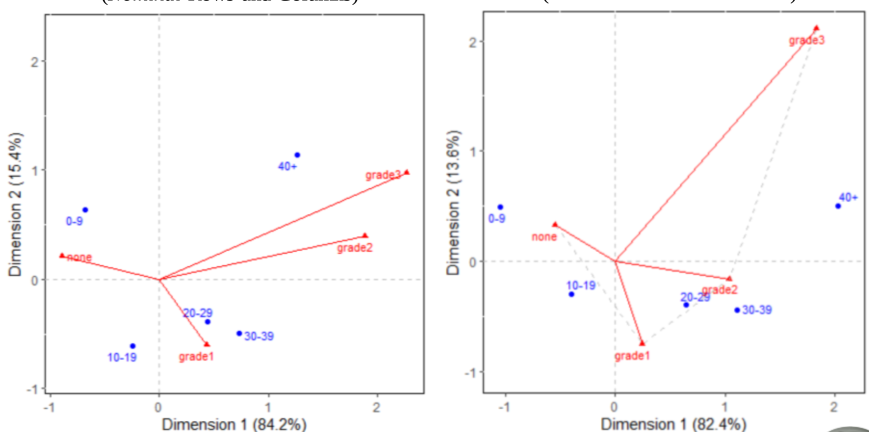
64

Example: Selikoff's Asbestos Data

Ordinal Correspondence Analysis

```
> ris = CAvariants(asbestos, catype = "DOCA")
> plot(ris, plottype = "biplot", biptype = "row", scaleplot = 1.5)
```

Row Isometric Biplot (*Nominal* Rows and Columns) | Row Isometric Biplot (*Ordered* Rows and Columns)



65

See Lombardo, Beh and Kroonenberg (2016) for more on biplots for doubly and singly ordered analysis

Overdispersion and Correspondence Analysis Strategies

Adjusted standardised residuals,
generalised standardised residuals
and the
Cressie-Read family of Divergence Statistics

Standardised Residual

Overdispersion

Correspondence analysis may alternatively, and equivalently, be performed by applying a SVD to the matrix of *standardised residuals*

$$\mathbf{Z} = \mathbf{D}_I^{-1/2}(\mathbf{P} - \mathbf{r}\,\mathbf{c}^T)\mathbf{D}_J^{-1/2}$$

so that

$$\mathbf{Z} = \widetilde{\mathbf{A}}\mathbf{D}_\lambda\widetilde{\mathbf{B}}^{\mathrm{T}}$$

where $\widetilde{\mathbf{A}}^{\mathrm{T}}\widetilde{\mathbf{A}} = \mathbf{I}_M$ and $\widetilde{\mathbf{B}}^{\mathrm{T}}\widetilde{\mathbf{B}} = \mathbf{I}_M$.

The (i,j)'th element of $\mathbf{Z}$ is the *standardised residual*

$$\mathrm{Z}_{\mathrm{ij}} = \frac{\mathrm{p}_{\mathrm{ij}} - \mathrm{p}_{\mathrm{i}\bullet}\mathrm{p}_{\bullet\mathrm{j}}}{\sqrt{\mathrm{p}_{\mathrm{i}\bullet}\mathrm{p}_{\bullet\mathrm{j}}}}$$

67

Standardised Residual

Overdispersion

The standardised residuals is based on the assumption that

$$\mathrm{n}_{\mathrm{ij}} \sim \mathrm{Poisson}(\mathrm{n}\mathrm{p}_{\mathrm{i}\bullet}\mathrm{p}_{\bullet\mathrm{j}})$$

so that

$$\mathrm{E}(\mathrm{n}_{\mathrm{ij}}) = \mathrm{Var}(\mathrm{n}_{\mathrm{ij}}) = \frac{\mathrm{n}_{\mathrm{i}\bullet}\mathrm{n}_{\bullet\mathrm{j}}}{\mathrm{n}}$$

$$\Rightarrow \mathrm{E}(\sqrt{\mathrm{n}}\,\mathrm{Z}_{\mathrm{ij}}) = 0 \qquad \mathrm{Var}(\sqrt{\mathrm{n}}\,\mathrm{Z}_{\mathrm{ij}}) = 1$$

. . . *but* . . .

68

Adjusted Standardised Residual

Overdispersion

Haberman (1973) points out that, under independence

$$\text{Var}\left(\sqrt{n}\,Z_{ij}\right) = (1 - p_{i\bullet})(1 - p_{\bullet j}) < 1$$

Agresti (2002, p. 81) comments that, for $\sqrt{n}\,Z_{ij}$,

"their asymptotic variances are less than 1.0, averaging $[(I-1)(J-1)]$/(number of cells)".

Agresti (2002, pp. 588 – 589) provides a proof of this result.

Therefore, the standardised residual may be amended to yield

$$\tilde{Z}_{ij} = \frac{p_{ij} - p_{i\bullet}p_{\bullet j}}{\sqrt{p_{i\bullet}p_{\bullet j}(1 - p_{i\bullet})(1 - p_{\bullet j})}}$$

and is the *adjusted standardised residual* and are asymptotically standard normally distributed; see Beh (2012) for details on CA using $\tilde{Z}_{ij}$.

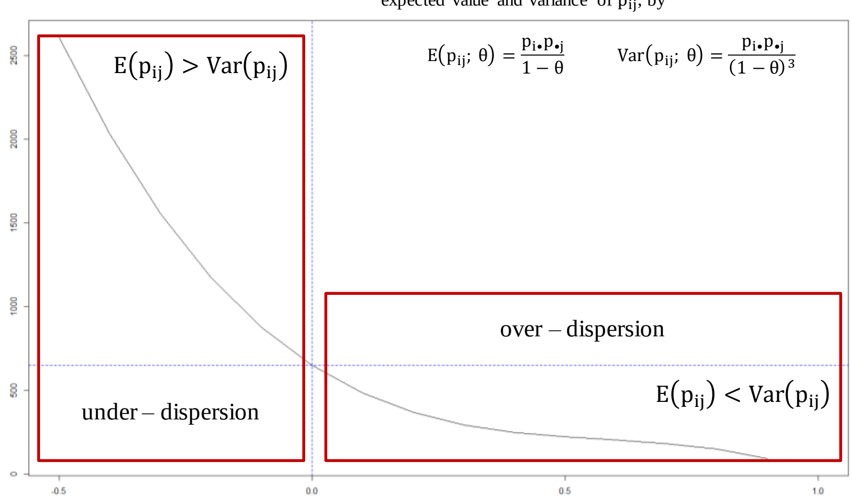
69

Example

Overdispersion

For our contingency table . . .

For the generalised Poisson distribution, and a given θ, the expected value and variance of p_{ij}, by



70

Power Transformations of Profiles

CR Family of Divergence Statistics

- *Recall* . . . at the core of CA lies, in part, analysing those rows (and columns) that are similar, or different. Done by comparing their *profiles*.
- The *centred profile* of the ith row category, say, is

$$\left(\frac{p_{i1}}{p_{i\bullet}} - p_{\bullet 1}, \frac{p_{i2}}{p_{i\bullet}} - p_{\bullet 2}, . \ . \ . \ , \frac{p_{iJ}}{p_{i\bullet}} - p_{\bullet J}\right)$$

- But what happens if we consider a power transformation of the profiles:

$$\left(\left(\frac{p_{i1}}{p_{i\bullet}}\right)^{\delta} - p_{\bullet 1}^{\delta}, \left(\frac{p_{i2}}{p_{i\bullet}}\right)^{\delta} - p_{\bullet 2}^{\delta}, . \ . \ . \ , \left(\frac{p_{iJ}}{p_{i\bullet}}\right)^{\delta} - p_{\bullet J}^{\delta}\right)$$

- Greenacre (2009) proposed *log-ratio analysis* (LRA) – "power family 2"
  - He confined $\delta \in [0, 1]$
  - No link to any interpretable total inertia or distance measures were given

71

Cressie-Read Family of Divergence Statistics

CR Family of Divergence Statistics

For some δ, a family of chi-squared statistics with (I – 1)(J – 1) degrees of freedom:

$$\mathrm{CR}(\delta) = \frac{2n}{\delta(\delta+1)} \sum_{i=1}^{I} \sum_{j=1}^{J} p_{ij} \left[\left(\frac{p_{ij}}{p_{i\bullet} p_{\bullet j}}\right)^{\delta} - 1\right]$$

where $\delta \in (-\infty, \infty)$ (Cressie & Read, 1984).

Read and Cressie (1988, p. 96) point out

> Departures involving large ratios of the alternative to null expected frequencies in one or two cells are best detected using large values of δ, say $\delta = 5$

72

Special Cases

Likelihood Ratio Test Statistic

$$G^2 = CR(0) = 2n\sum_{i=1}^{I}\sum_{j=1}^{J} p_{ij}\ln\left(\frac{p_{ij}}{p_{i\bullet}p_{\bullet j}}\right)$$

Modified Likelihood Ratio Statistic

$$M^2 = CR(-1) = 2n\sum_{i=1}^{I}\sum_{j=1}^{J} p_{i\bullet}p_{\bullet j}\ln\left(\frac{p_{i\bullet}p_{\bullet j}}{p_{ij}}\right)$$

Modified Chi-Squared Statistic

$$N^2 = CR(-2) = n\sum_{i=1}^{I}\sum_{j=1}^{J}\frac{\left(p_{ij} - p_{i\bullet}p_{\bullet j}\right)^2}{p_{ij}}$$

The Freeman-Tukey Statistic

$$T^2 = CR\left(-\frac{1}{2}\right) = 4n\sum_{i=1}^{I}\sum_{j=1}^{J}\left(\sqrt{p_{ij}} - \sqrt{p_{i\bullet}p_{\bullet j}}\right)^2$$

. . . are also chi-squared random variables with df = (I – 1)(J – 1)

CR Family of Divergence Statistics

73

The "Approximation"

Read and Cressie (1988, pp. 94 – 95) show that

$$CR(\delta) \approx CR^*(\delta) = \frac{n}{\delta^2}\sum_{i=1}^{I}\sum_{j=1}^{J}\frac{\left(p_{ij}^{\delta} - \left(p_{i\bullet}p_{\bullet j}\right)^{\delta}\right)^2}{\left(p_{i\bullet}p_{\bullet j}\right)^{2\delta-1}}$$

$$= n\sum_{i=1}^{I}\sum_{j=1}^{J} p_{i\bullet}p_{\bullet j}\left[\frac{1}{\delta}\left(\left(\frac{p_{ij}}{p_{i\bullet}p_{\bullet j}}\right)^{\delta} - 1\right)\right]^2$$

It can be shown that, exactly,

$$M^2 = CR^*(0) \qquad T^2 = CR^*\left(\frac{1}{2}\right) \qquad X^2 = CR^*(1)$$

CR Family of Divergence Statistics

74

The "Approximation"

Read and Cressie (1988, pp. 94 – 95) show that

$$CR(\delta) \approx CR^*(\delta) = \frac{n}{\delta^2}\sum_{i=1}^{I}\sum_{j=1}^{J}\frac{\left(p_{ij}^{\delta} - (p_{i\bullet}p_{\bullet j})^{\delta}\right)^2}{(p_{i\bullet}p_{\bullet j})^{2\delta-1}}$$

$$= n\sum_{i=1}^{I}\sum_{j=1}^{J} p_{i\bullet}p_{\bullet j}\left[\frac{1}{\delta}\left(\left(\frac{p_{ij}}{p_{i\bullet}p_{\bullet j}}\right)^{\delta} - 1\right)\right]^2$$

Read and Cressie (1988, p. 97) also suggested that $\delta = 2/3$ works well for assessing deviations from independence giving the *approximation* to the Cressie-Read statistic

$$CR^*\left(\frac{2}{3}\right) = \frac{9n}{5}\sum_{i=1}^{I}\sum_{j=1}^{J} p_{i\bullet}p_{\bullet j}\left(\left(\frac{p_{ij}}{p_{i\bullet}p_{\bullet j}}\right)^{2/3} - 1\right)^2$$

CR Family of Divergence Statistics

75

Divergence Residual

Beh & Lombardo (2022b) define the *divergence residual* of the (i,j)'th cell our contingency table by

$$r_{ij}(\delta) = \frac{1}{\delta}\left(\left(\frac{p_{ij}}{p_{i\bullet}p_{\bullet j}}\right)^{\delta} - 1\right)$$

for a given δ so that the total inertia is

$$\phi^2(\delta) = \frac{CR^*(\delta)}{n} = \sum_{i=1}^{I}\sum_{j=1}^{J} r_{ij}^2(\delta)$$

- **Classical Correspondence Analysis**

$$r_{ij}(1) = \frac{p_{ij} - p_{i\bullet}p_{\bullet j}}{\sqrt{p_{i\bullet}p_{\bullet j}}} \qquad \phi^2(1) = \frac{X^2}{n}$$

- **"Freeman-Tukey" Correspondence Analysis**

$$r_{ij}\left(\frac{1}{2}\right) = 2\left(\sqrt{p_{ij}} - \sqrt{p_{i\bullet}p_{\bullet j}}\right) \qquad \phi^2\left(\frac{1}{2}\right) = \frac{T^2}{n}$$

Cuadras and Cuadras (2006) . . . *Hellinger Distance Decomposition (HDD)* method,
Beh, Lombardo & Alberti (2018) . . . *"Freeman-Tukey" CA (FTCA)*

CR Family of Divergence Statistics

76

Divergence Residual

Beh & Lombardo (2022b) define the *divergence residual* of the (i, j)'th cell our contingency table by

$$r_{ij}(\delta) = \frac{1}{\delta}\left(\left(\frac{p_{ij}}{p_{i\bullet}p_{\bullet j}}\right)^{\delta} - 1\right)$$

for a given δ so that the total inertia is

$$\phi^2(\delta) = \frac{CR^*(\delta)}{n} = \sum_{i=1}^{I}\sum_{j=1}^{J} r_{ij}^2(\delta)$$

- **Correspondence Analysis & M^2?**

$$r_{ij}(0) = \sqrt{p_{i\bullet}p_{\bullet j}}\ln\left(\frac{p_{ij}}{p_{i*}p_{*j}}\right) \qquad \phi^2(0) = \frac{M^2}{n}$$

Greenacre (2009) – *Log-ratio analysis (LRA)*

- **Correspondence Analysis & CR?**

$$r_{ij}\left(\frac{2}{3}\right) = \frac{3}{2}\left(\left(\frac{p_{ij}}{p_{i*}p_{*j}}\right)^{2/3} - 1\right) \qquad \phi^2\left(\frac{1}{2}\right) = \frac{CR^2}{n}$$

Example: Nobel Prize Data

Since 1901, the Nobel Prize has been awarded in the fields of physics, physiology and medicine, peace, literature and economics.

We examine the association between the *Prize* awarded (between 1901 and 2022) and the *Country* of affliation. Note:

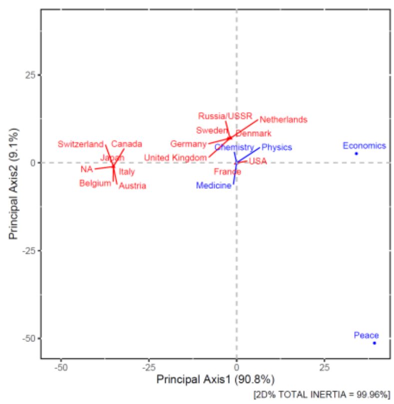
"NA" is not a country but represents recipients who, according to the website, are affiliated with an institution that is not centrally located in a single country

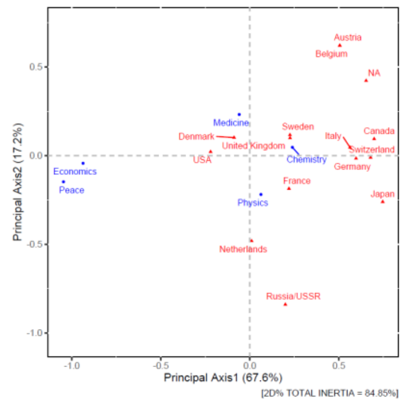
Prize/Country	Physics	Chemistry	Economics	Peace	Medicine
USA	113	83	77	3	127
UK	28	30	6	0	32
Germany	24	40	1	0	17
France	17	11	2	1	10
Switzerland	10	7	0	0	8
Japan	10	7	0	0	4
Sweden	5	5	1	0	6
NA	3	5	0	0	7
Russia/USSR	11	2	1	0	1
Netherlands	6	1	1	0	2
Canada	3	3	0	0	3
Denmark	3	1	1	0	4
Italy	3	1	0	0	3
Austria	1	1	0	0	4
Belgium	1	1	0	0	4

Source: https://www.nobelprize.org/prizes/facts/lists/affiliations.php

Example: Nobel Prize Data

CR Family of Divergence Statistics





$\delta = 0$

- *Total Inertia:*

Modified log-likelihood statistic=272.11

Log-ratio analysis

$\delta = 1/2$

- *Total Inertia:*

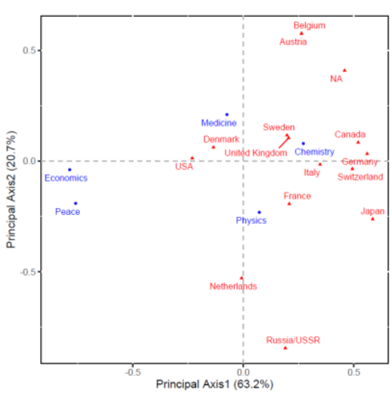
Freeman-Tukey statistic =142.73

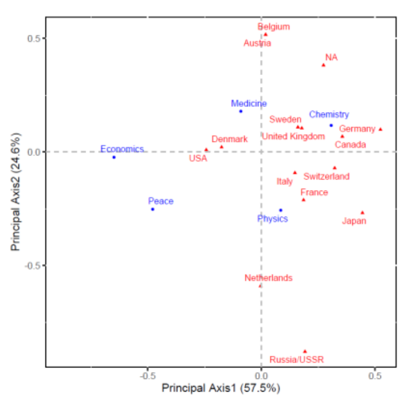
HDD, or FTCA

79

Example: Nobel Prize Data

CR Family of Divergence Statistics





$\delta = 2/3$

- *Total Inertia:*

Cressie-Read statistic=118.91

$\delta = 1$

- *Total Inertia:*

Pearson's statistic=106.19

Classical CA

80

Further Links

CR Family of Divergence Statistics

Keep in mind: Any value of δ can be considered for the Cressie-Read family of divergence statistics

Hellinger Distance Decomposition (HDD)

- Rao (1995) described the pro's of the Hellinger distances ($\delta = 1/2$) over $\delta = 1$
- Cuadras and Cuadras (2006) proposed Hellinger Distance Decomposition (HDD) in their "parametric correspondence analysis" approach
- HDD is equivalent to our approach when $\delta = 1/2$ (Freeman – Tukey)
  - Beh, Lombardo & Alberti (2018) linked the Hellinger distance to the Freeman-Tukey statistic
- Cuadras and Cuadras (2006)
  - did not link their HDD to the Freeman-Tukey statistic
  - only compared their HDD ($\delta = 1/2$) with CA ($\delta = 1$)

81

Further Links

CR Family of Divergence Statistics

Keep in mind: Any value of δ can be considered for the Cressie-Read family of divergence statistics

Log-Ratio Analysis

- Greenacre (2009) described two applications of power transformations in CA
  - "Power family 1" – involves the transformation of the cells . . . p_{ij}^{δ}
  - "Power family 2" – involves the transformation of profiles . . . $\left(p_{ij}/p_{i\bullet}\right)^{\delta}$
- Greenacre's (2009) "Power family 2" . . .
  - examined differences in CA when $\delta = 0$ (LRA) and $\delta = 1$ (CA)
  - LRA is equivalent to our technique when $\delta = 0$
  - did not examine what would happen for values of δ that lie outside of $[0,\ 1]$
  - did not link LRA to any measure of association (including M^2)

82

References

Bibliographic Information

1. Agresti, A. (2002), *Categorical Data Analysis* (2nd ed), Wiley.
2. Beh, E. J. (1997), Simple correspondence analysis of ordinal cross-classifications using orthogonal polynomials, *Biometrical Journal*, 39, 589 – 613
3. Beh, E. J. (2012), Simple correspondence analysis using adjusted residuals, *Journal of Statistical Planning and Inference*, 142, 965 – 973
4. Beh, E. J. & Lombardo, R. (2012), A genealogy of correspondence analysis, *The Australian and New Zealand Journal of Statistics*, 54, 137 – 168
5. Beh, E. J. & Lombardo, R. (2014), *Correspondence Analysis: Theory, Practice and New Strategies*, Chichester: Wiley
6. Beh, E. J. & Lombardo, R. (2019), A genealogy of correspondence analysis: Part 2 – The variants, *Electronic Journal of Applied Statistical Analysis*, 12, 552 – 603
7. Beh, E. J. & Lombardo, R. (2021), *An Introduction to Correspondence Analysis*, Chichester: Wiley.
8. Beh, E. J. & Lombardo, R. (2022a), Features of the polynomial biplot for ordered contingency tables, *Journal of Computational and Graphical Statistics*, 31, 403 – 412
9. Beh, E. J. & Lombardo, R. (2022b), Correspondence analysis using the Cressie-Read divergence statistic (in review)
10. Beh, E. J., Lombardo, R. & Alberti, G. (2018), Correspondence analysis and the Freeman-Tukey statistic: A study of archaeological data, *Computational Statistics and Data Analysis*, 128, 73 – 86
11. Benzécri, J-P. (1973), *L'Analyse des Données: 1. La Taxinomie*, Dunad.
12. Benzécri, J-P. (1973), *L'Analyse des Données: 1. L'Analyse des Correspondances*, Dunad.
13. Benzécri, J.-P. (1992), *Correspondence Analysis Handbook*, Marcel Dekker.
14. Best, D.J. and Rayner, J.C.W. (1996) Nonparametric analysis for doubly ordered two-way contingency tables. *Biometrics*, 52, 1153 – 1156.
15. Blasius, J. & Greenacre, M. (eds.) (1998), *Visualization of Categorical Data*, Academic Press.
16. Blasius, J. & Greenacre, M. (2014), *Visualization and Verbalization of Data*, CRC Press.
17. Cressie, N. & Read, T. R. C. (1984), Multinomial goodness-of-fit tests, *Journal of the Royal Statistical Society (Series B)*, 46, 440 – 464.

83

References

Bibliographic Information

18. Cuadras, C. M. & Cuadras, D. (2006), A parametric approach to correspondence analysis, *Linear Algebra and its Applications*, 417, 64 – 74.
19. D'Ambra, L. & Lauro, N. C. (1989) Non-symmetrical correspondence analysis for three-way contingency table, in *Multiway Data Analysis* (eds. R. Coppi & S. Bolasco), Elsevier, pp. 301 – 315.
20. Galton, F. (1892), *Finger Prints*, Macmillan
21. Goodman, L. A. & Kruskal, W. H. (1954) Measures of association for cross classifications. *Journal of the American Statistical Association* 49, 732--764.
22. Greenacre, M. J. (1978), *Some Objective Methods of Graphical Display of a Data Matrix*, PhD Thesis, L'Universite Pierre et Marie Curie (Paris VI), France.
23. Greenacre, M. J. (1984), *Theory and Applications of Correspondence Analysis*, Academic Press.
24. Greenacre, M. (2009), Power transformations in correspondence analysis, *Computational Statistics and Data Analysis*, 53, 3107 – 3116.
25. Greenacre, M. (2017), *Correspondence Analysis in Practice* (3rd ed), Chapman & Hall/CRC.
26. Greenacre, M. & Blasius, J. (eds.) (1994), *Correspondence Analysis in the Social Sciences*, Academic Press.
27. Greenacre, M. & Blasius, J. (eds.) (2006), *Multiple Correspondence Analysis and Related Methods*, Chapman & Hall/CRC.
28. Haberman, S. J. (1973), The analysis of residuals in cross-classified data, *Biometrics*, 205 – 220.
29. Kroonenberg, P. M. & Lombardo, R. (1999), Nonsymmetric correspondence analysis: a tool for analysing contingency tables with a dependence structure, *Multivariate Behavioral Research*, 34, 367 – 396.
30. Lancaster, H. O. (1953), A reconciliation of χ^2, considered from metrical and enumerative aspects, *Sankhyā*, 13, 1 – 10.
31. Lauro, N. C. & D'Ambra, L. (1984), L'Analyse non symm trique des correspondances, in *Data Analysis and Informatics III* (ed. E. Diday), Elsevier, pp. 433 – 446.
32. Light, R. J. & Margolin, B. H. (1971), An analysis of variance for categorical data, *Journal of the American Statistical Association*, 66, 534 – 544.
33. Lombardo, R., Beh, E. J. & D'Ambra, L. (2007), Non-symmetric correspondence analysis with ordinal variables, *Computational Statistics and Data Analysis*, 52, 566 – 577.

84

References

Bibliographic Information

34. Lombardo, R., Beh, E. J. & Kroonenberg, P. M. (2021), Symmetrical and non-symmetrical variants of three-way correspondence analysis for ordered variables, *Statistical Science*, 36, 542 – 561
35. Lombardo, R., Kroonenberg, P. M. & Beh, E. J. (2016), Modelling trends in ordered three-way non-symmetrical correspondence analysis, *Proceedings of the 48th Scientific Meeting of the Italian Statistical Society*, 14 pages.
36. Marcotorchino, F. (1984), *Utilisation des comparaisons par paires en statistique des contingences: Partie II*, Report # F 071, Etude du Centre Scientifique, IBM, France.
37. Nishisato, S. (2022), *Optimal Quantification and Symmetry*, Springer.
38. Parsa, A. R. & Smith, W. B. (1993), Scoring under ordered constraints in contingency tables, *Communications in Statistics – Theory and Methods*, 22, 3537 – 3551.
39. Pearson, K. (1904), On the theory of contingency and its relation to association and normal correlation. *Drapers Memoirs. Biometric Series*, Vol. 1. London.
40. Quetelet, A. (1842), *A Treatise on Man and the Development of His Faculties*. William and Robert Chambers, Edinburgh.
41. Rao, C. R. (1995) A review of canonical coordinates and an alternative to correspondence analysis using Hellinger distance. *Qüestiió*, 19, 23—63
42. Rayner, J. C. W. & Beh, E. J. (2009), Towards a better understanding of correlation, *Statistica Neerlandica*, 63, 324 – 333.
43. Rayner, J. C. W. & Best, D. J. (1996), Smooth extensions of Pearson's product moment correlation and Spearman's rho, *Statistics & Probability Letters*, 30, 171 – 177.
44. Read, T. R. C. & Cressie, N. A. C. (1988), *Goodness-of-fit Statistics for Discrete Multivariate Data*, Springer-Verlag.
45. Ritov, Y. & Gilula, Z. (1983), Analysis of contingency tables by correspondence models subject to order constraints, Journal of the American Statistical Association, 88, 1380 – 1387.
46. Schriever, B. F. (1983), Scaling of order dependent categorical variables with correspondence analysis, *International Statistical Review*, 51, 225 – 238.
47. Upton, G. J. G. (2017), *Categorical Data Analysis by Example*, Wiley.
48. Yang, K.-S. & Huh, M.-H. (1999), Correspondence analysis of two-way contingency tables with ordered column categories, Journal of the Korean Statistical Society, 28, 347 – 358.

85