

Correspondence Analysis & Related Methods

Michael Greenacre

SESSION 16: Analyzing concatenated tables (note: no slides for Session 15)

Between- or within-sets?

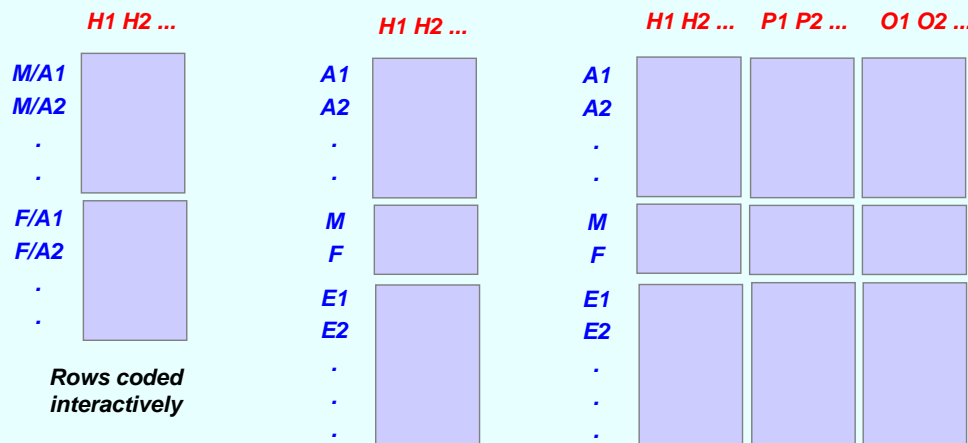
- In simple CA we analyze the relationship between two categorical variables, a row variable and a column variable.
- In order to generalize this to more than two variables, we need to distinguish two situations:
 - There are several “explanatory” variables and one or more “response” variables, and we are interested in the relationship between these two sets.
 - There is a set of “homogeneous” variables, usually all measured on the same scale, and we are interested in their inter-relationships (similar to factor analysis context, but with categorical data).
- We have already had an example of the first situation: in the Spanish Health Survey example: explanatory variables = age and sex, response variable = perceived health; we combined the 7 age categories and 2 sex categories into one **interactively coded** variable with 14 categories and did a simple CA. We first look at this easier approach of stacking – or **concatenating** – tables.

Different ways of concatenating tables

Describing (“explanatory”) variables: **A(GE)**, **S(EX)**, **E(DUCATION)**, etc...

with levels **A1, A2, ...; M, F; E1, E2, ...; etc...**

Variables to be described (“response”): **H(ALTH)**, **P(RODUCTS)**, **O(PINIONS)**, etc... with levels **H1, H2, ...; P1, P2, ...; O1, O2, ...; etc...**



“Salud” data

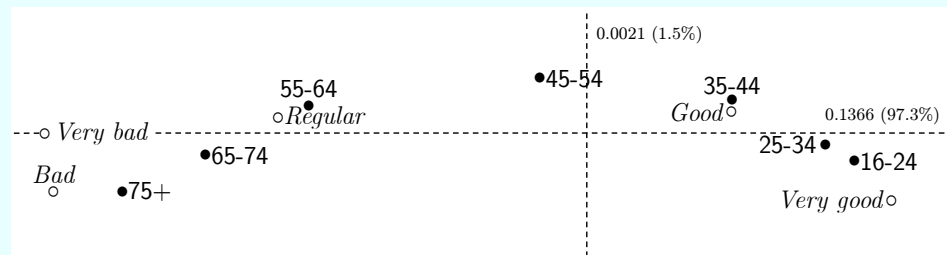
Table 1: Crosstabulation of age groups by perceived health status

AGE GROUP	Very Good	Good	Regular	Bad	Very Bad	SUM
16-24	243	789	167	18	6	1223
25-34	220	809	164	35	6	1234
35-44	147	658	181	41	8	1035
45-54	90	469	236	50	16	861
55-64	53	414	306	106	30	909
65-74	44	267	284	98	20	713
75+	20	136	157	66	17	396
SUM	817	3542	1495	414	103	6371

Table 2: Row percentages calculated from Table 1

AGE GROUP	Very Good	Good	Regular	Bad	Very Bad	SUM
16-24	19.9	64.5	13.7	1.5	0.5	100.0
25-34	17.8	65.6	13.3	2.8	0.5	100.0
35-44	14.2	63.6	17.5	4.0	0.8	100.0
45-54	10.5	54.5	27.4	5.8	1.9	100.0
55-64	5.8	45.5	33.7	11.7	3.3	100.0
65-74	6.2	37.4	39.8	13.7	2.8	100.0
75+	5.1	34.3	39.6	16.7	4.3	100.0
AVERAGE	12.8	55.6	23.5	6.5	1.6	100.0

CA of perceived health status

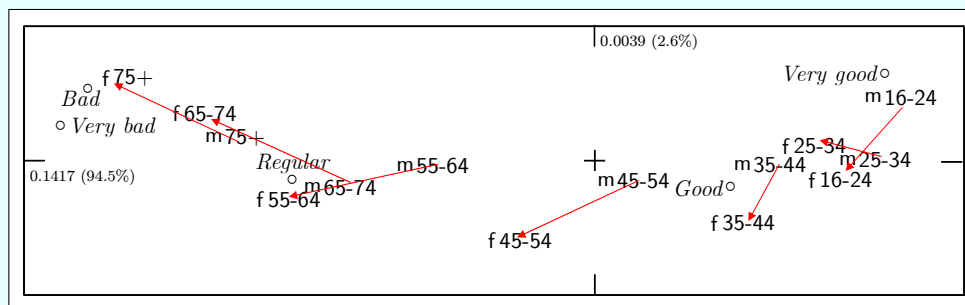


Perceived health status: male - female

Table 3: Age group and sex interactively crosstabulated with health status

AGE GROUP	Very Good	Good	Regular	Bad	Very Bad	SUM
MALES						
16-24	145	402	84	5	3	639
25-34	112	414	74	13	2	615
35-44	80	331	82	24	4	521
45-54	54	231	102	22	6	415
55-64	30	219	119	53	12	433
65-74	18	125	110	35	4	292
75+	9	67	65	25	8	174
FEMALES						
16-24	98	387	83	13	3	584
25-34	108	395	90	22	4	619
35-44	67	327	99	17	4	514
45-54	36	238	134	28	10	446
55-64	23	195	187	53	18	476
65-74	26	142	174	63	16	421
75+	11	69	92	41	9	222
SUM	817	3542	1495	414	103	6371

Perceived health status: male - female



Example: data set "women" - response variables

- ISSP 1994 survey on Family & Changing Gender Roles
- 33123 respondents
- 24 countries (former West and East Germany still kept separate)
- We focus on four questions related to women's participation in work outside the home
- Should women "work full-time" (W), "work part-time" (w), "stay at home" (H), or "unsure/don't know" (?) at these four different periods of married life:

1. before having a child, with possible responses **1W 1w 1H 1?**
2. with a pre-school child **2W 2w 2H 2?**
3. when youngest child is still at school **3W 3w 3H 3?**
4. when all children have left home **4W 4w 4H 4?**

Example: data set "women" - explanatory variables

We also have various explanatory variables, from which we select the following six:

Country 24 countries: *AUS* (Australia), *DW* (West Germany), *DE* (East Germany), *GB* (Great Britain), *NI* (Northern Ireland), *USA*, *A* (Austria), *H* (Hungary), *I* (Italy), *IRL* (Ireland), *NL* (Netherlands), *N* (Norway), *S* (Sweden), *CZ* (Czechoslovakia), *SLO* (Slovenia), *PL* (Poland), *BG* (Bulgaria), *RUS* (Russia), *NZ* (New Zealand), *CDN* (Canada), *RP* (Philippines), *IL* (Israel), *J* (Japan), *E* (Spain)

Sex 2 categories: *M*, *F*

Age 6 groups: *A1* (up to 25), *A2* (26-35), *A3* (36-45), *A4* (46-55), *A5* (56-65), *A6* (66 and over)

Marital status 5 groups: *ma* (married), *wi* (widowed), *di* (divorced), *se* (separated), *si* (single)

Education 7 groups: *E0* (none), *E1* (incomplete primary), *E2* (primary), *E3* (incomplete secondary), *E4* (secondary), *E5* (incomplete tertiary), *E6* (tertiary)

Social class 7 groups: *S0* (other), *S1* (lower class), *S2* (working class), *S3* (upper working/lower middle), *S4* (middle), *S5* (upper middle), *S6* (upper)

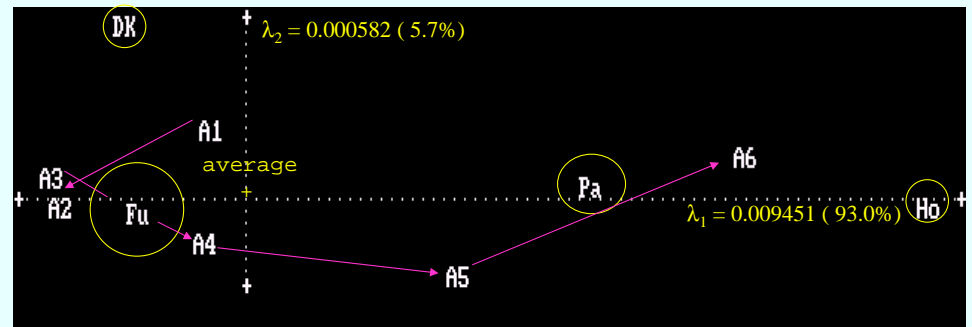
(a) Original (b) Indicator and (c) Concatenated

(a)	(b)
1 2 3 4 Sex Age ...	1 2 3 4 Sex Age ...
W w P ? W w P ? W w P ? W w P ? M F 1 2 3 4 5 6	W w P ? W w P ? W w P ? W w P ? M F 1 2 3 4 5 6
1 3 2 2 2 6 ...	1 0 0 0 0 0 1 0 0 1 0 0 0 0 1 0 0 0 0 1 ...
1 2 2 2 2 4 ...	1 0 0 0 0 1 0 0 0 1 1 0 0 0 1 0 0 0 0 1 0 0 ...
1 3 4 4 2 1 ...	1 0 0 0 0 0 1 0 0 0 0 1 0 0 0 1 0 0 1 1 0 0 0 0 0 ...
1 2 2 1 2 4 ...	1 0 0 0 0 1 0 0 0 1 0 0 0 1 0 0 0 0 1 0 0 0 1 0 0 ...
1 3 2 4 1 5 ...	1 0 0 0 0 0 1 0 0 1 0 0 0 0 0 1 1 0 0 0 0 0 1 0 ...
1 2 1 1 2 1 ...	1 0 0 0 0 1 0 0 0 1 0 0 0 1 0 0 0 0 0 1 1 0 0 0 0 ...
...	...

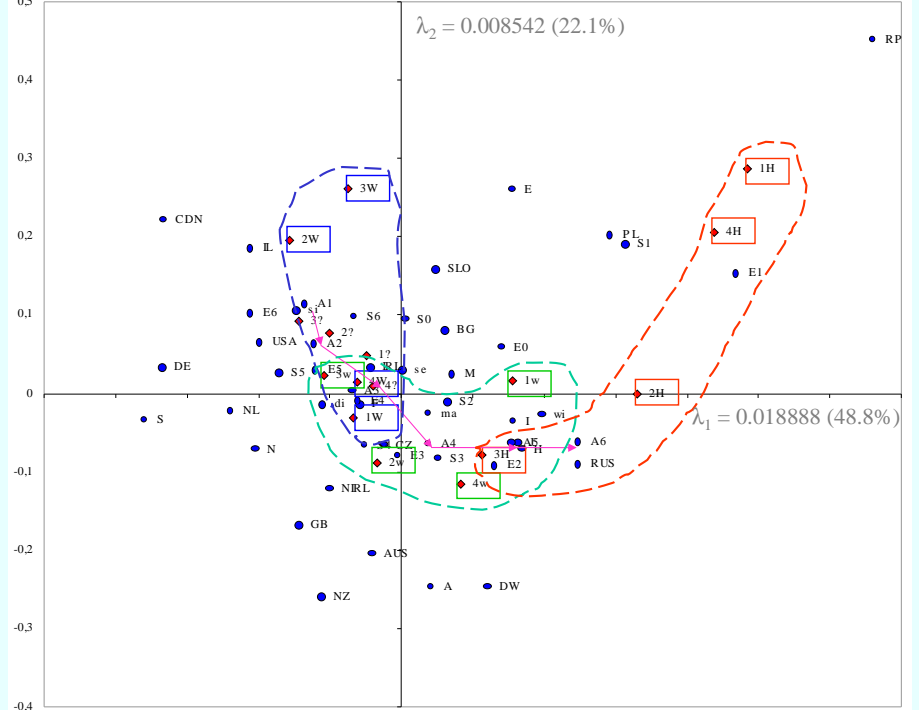
(c)	1	2	3	4
	W w H ?	W w H ?	W w H ?	W w H ?
Sex M	10309 2436 1207 1091	3220 7634 3099 1090	1360 4972 7732 979	9300 3030 1182 1531
Sex F	13712 2435 842 1056	4020 9999 2829 1197	1628 6926 8324 1167	12100 3408 893 1644
Age A1	3562 744 272 361	1621 2408 527 383	494 2095 1947 403	3448 741 261 489
Age A2	5192 885 337 459	1829 3659 864 521	804 2840 2710 519	4668 1151 334 720
Age A3	5173 879 332 486	1523 3795 1038 514	738 2603 3053 476	4516 1337 321 696
Age A4	4006 767 321 326	953 3065 1062 340	399 1833 2879 309	3436 1154 326 504
Age A5	3158 749 338 231	711 2406 1120 239	317 1397 2577 185	2691 1037 397 351
Age A6	2953 851 453 288	609 2320 1322 294	240 1139 2908 258	2660 1027 436 422
...

Concatenated format

	1	2	3	4
	W w H ?	W w H ?	W w H ?	W w H ?
Sex M	10309 2436 1207 1091	3220 7634 3099 1090	1360 4972 7732 979	9300 3030 1182 1531
Sex F	13712 2435 842 1056	4020 9999 2829 1197	1628 6926 8324 1167	12100 3408 893 1644
Age A1	3562 744 272 361	1621 2408 527 383	494 2095 1947 403	3448 741 261 489
Age A2	5192 885 337 459	1829 3659 864 521	804 2840 2710 519	4668 1151 334 720
Age A3	5173 879 332 486	1523 3795 1038 514	738 2603 3053 476	4516 1337 321 696
Age A4	4006 767 321 326	953 3065 1062 340	399 1833 2879 309	3436 1154 326 504
Age A5	3158 749 338 231	711 2406 1120 239	317 1397 2577 185	2691 1037 397 351
Age A6	2953 851 453 288	609 2320 1322 294	240 1139 2908 258	2660 1027 436 422
...



Correspondence analysis of all two-way tables

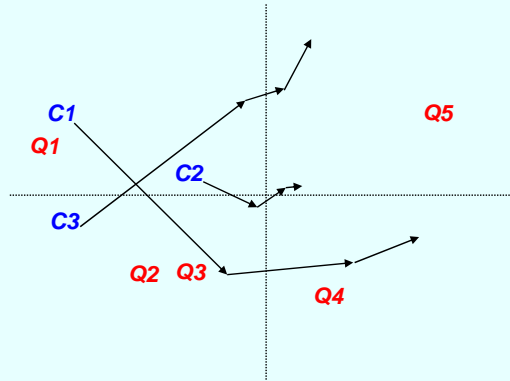


Trend studies

Tables (e.g., C(OUNTRIES) by Q(UESTION) repeated over T(IME) T1, T2, ...

	Q1	Q2	...
T1/C1			
T1/C2			
.			
.			
T2/C1			
T2/C2			
.			
.			
T3/C1			
T3/C2			
.			
.			
T4/C1			
T4/C2			
.			
.			

CA



Trajectories of countries over time