

Multiple categorical variables

Concatenated tables

Between- or within-sets?

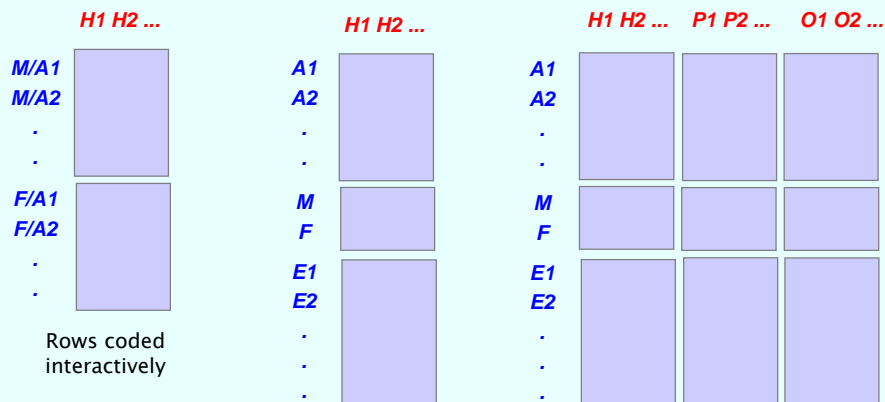
- In simple CA we analyze the relationship between two categorical variables, a row variable and a column variable.
- In order to generalize this to more than two variables, we need to distinguish two situations:
 1. There are several “explanatory” variables and one or more “response” variables, and we are interested in the relationship between these two sets.
 2. 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(EALTH), P(RODUCTS), O(PINIONS)**, etc... with levels **H1, H2, ...; P1, P2, ...; O1, O2, ...**; etc...



Example: data set “women94” – 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?**

Data set "women94" – 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:

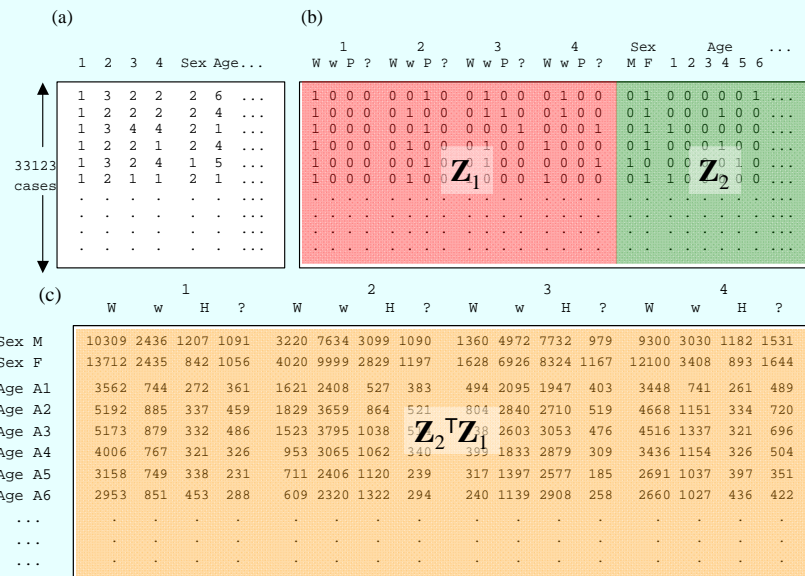
1. before having a child, 1W 1w 1H 1?
2. with a pre-school child 2W 2w 2H 2?
3. when youngest child is still at school 3W 3w 3H 3?
4. when all children have left home 4W 4w 4H 4?

Data set “women” – explanatory variables

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

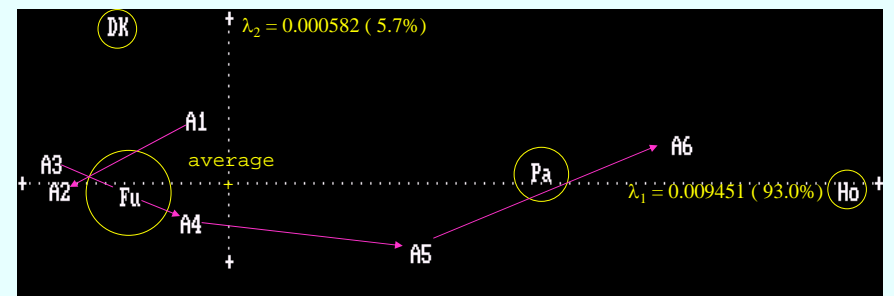
- Country** 24 countries: AUS (Australia), DW (West Germany), DE (East Germany), GB (Great Britain), NI (Northern Ireland), USA, A (Austria), H (Hungary), I (Italy), IRL (Ireland), NL (Netherlands), N (Norway), S (Sweden), CZ (Czechoslovakia), SLO (Slovenia), PL (Poland), BG (Bulgaria), RUS (Russia), 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

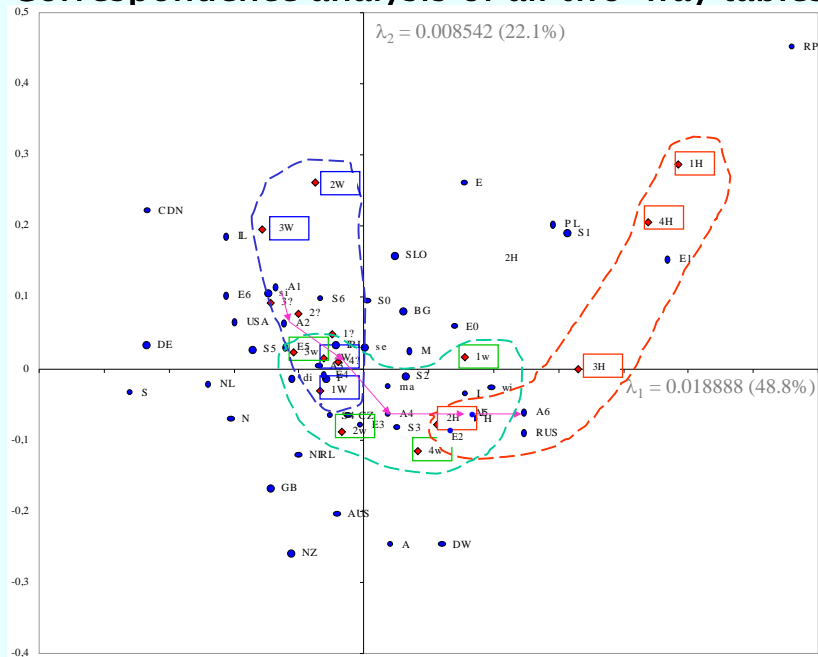


Concatenated format

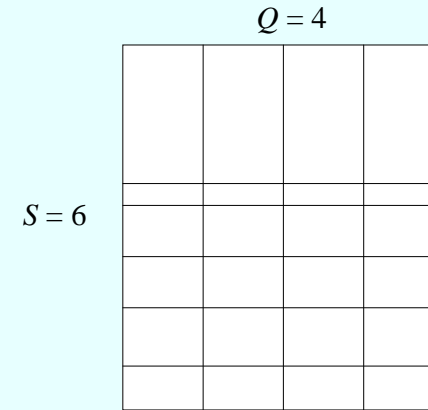
	1				2				3				4			
	W	w	H	?	W	w	H	?	W	w	H	?	W	w	H	?
Sex M	10309	2436	1207	1091	3220	7634	3099	1090	1360	4972	7732	979	9300	3030	1182	1531
Sex F	13712	2435	842	1056	4020	9999	2829	1197	1628	6926	8324	1167	12100	3408	893	1644
Age A1	3562	744	272	361	1621	2408	527	383	494	2095	1947	403	3448	741	261	489
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



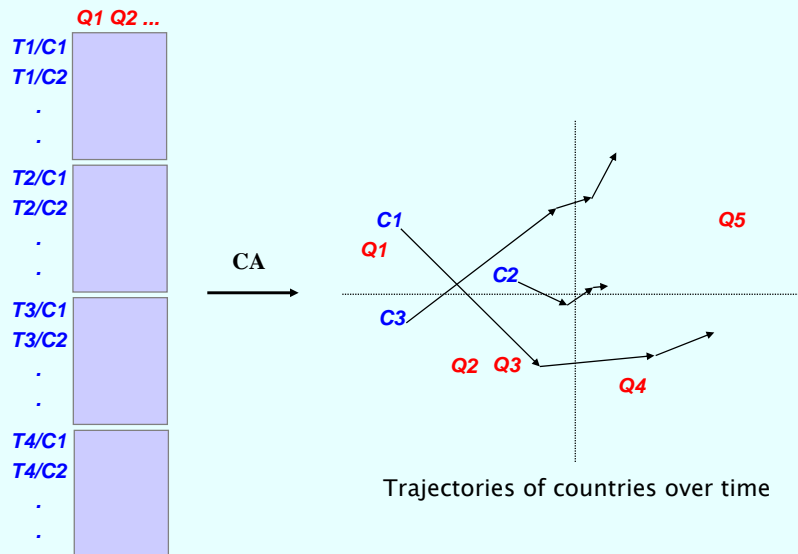
Inertia=Average inertia over tables



Inertia of concatenated table = Average of inertias in QS subtables

Trend studies

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



Data set “women02”

Substantive variables: *Do you strongly agree/ agree/ neither...nor.../ disagree/ strongly disagree to these statements...*

- A:** a working mother can establish a warm relationship with her child
- B:** a pre-school child suffers if his or her mother works
- C:** when a woman works the family life suffers
- D:** what women really want is a home and kids
- E:** running a household is just as satisfying as a paid job
- F:** work is best for a woman's independence
- G:** a man's job is to work; a woman's job is the household
- H:** working women should get paid maternity leave

Demographic variables

- g:** gender (1 = male, 2 = female)
- m:** marital status (1 = married/living as married, 2 = widowed, 3 = divorced, 4 = separated, but married, 5 = single, never married)
- e:** education (0 = no formal education, 1 = lowest education, 2 = above lowest education, 3 = higher secondary completed, 4 = above higher secondary level, below full university, 5 = university degree completed)
- a:** age (1 = 16-25 years, 2 = 26-35, 3 = 36-45, 4 = 46-55, 5 = 56-65, 6 = 66 and older)

Sample: Spanish sample (year 2002); $N=2471$ (including missing values)

Data file for "women02"

```

A B C D E F G H g m e a
2 4 3 3 4 1 4 1 1 1 3 4
2 4 3 9 4 1 4 1 1 1 3 3
3 2 2 3 4 1 3 2 2 2 1 6
3 9 2 2 2 1 3 1 1 1 1 6
9 1 2 2 3 2 3 1 2 1 1 4
2 4 4 4 2 2 5 1 1 5 4 2
1 3 2 3 2 4 4 2 2 1 5 5
2 9 4 9 9 1 3 1 1 5 3 5
2 4 2 3 4 4 5 1 2 1 5 3
2 4 2 2 4 2 4 1 2 1 2 3
. . . . . . . . . . . . . .
. . . . . . . . . . . . . .
. . . . . . . . . . . . . .

```

2471
rows

R script for "women02"

```

# read original data for Spain, 2002
women02 <- read.table("clipboard", header=T)

# getting the stacked matrix from the Burt matrix
women02.B <- mjca(women02)$Burt
rownames(women02.B)
colnames(women02.B)

women02.stack <- women02.B[49:69,1:48]
rownames(women02.stack)
colnames(women02.stack)

# checking frequencies of each substantive category
apply(women02.stack, 2, sum)/4

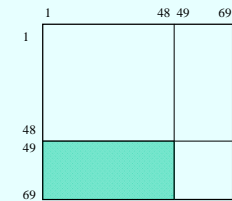
# combine H4 and H5 (H5 frequency very low)
women02.stack[,46] <- women02.stack[,46]+women02.stack[,47]
women02.stack <- women02.stack[,-47]

# CA of whole stacked matrix
plot(ca(women02.stack))

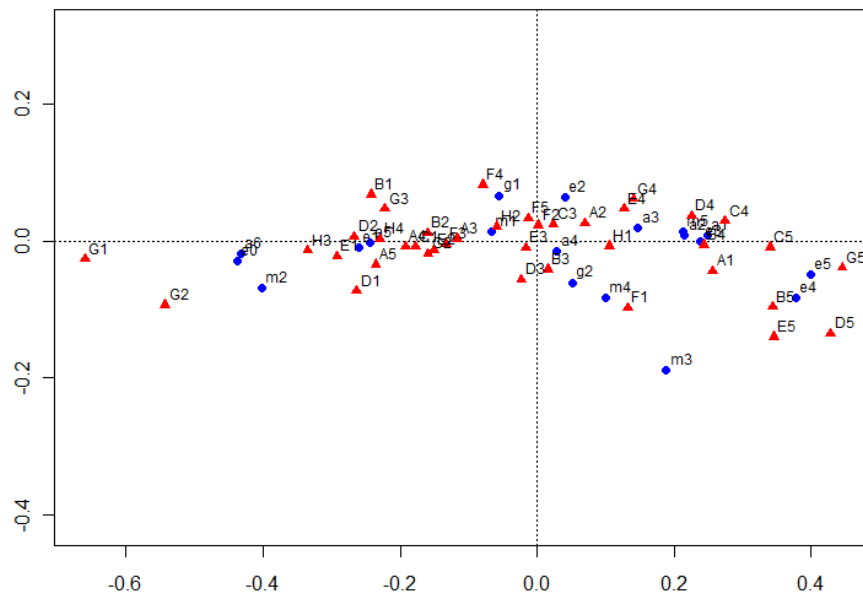
# checking frequencies of each demographic category
apply(women02.stack, 1, sum)/8

# subset analysis (subsetting out the missing categories)
plot(ca(women02.stack, subsetrow=seq(1,21)[-c(8,15)],
+ subsetcol=seq(1,47)[-c(6,12,18,24,30,36,42,47)]))

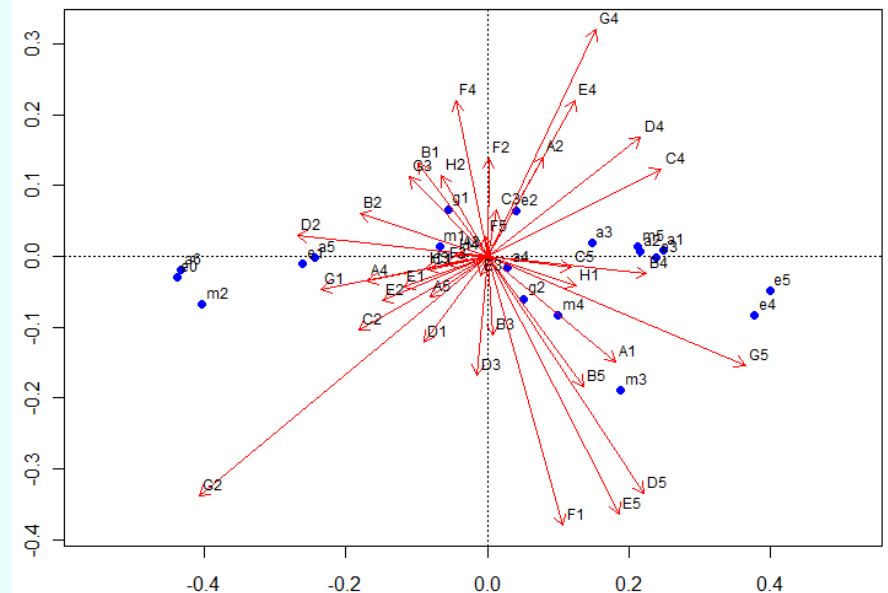
```



Subset analysis of "women02"



Subset analysis: contribution biplot



Contribution biplot (cleaned up - categories removed inside box)

