

Correspondence analysis

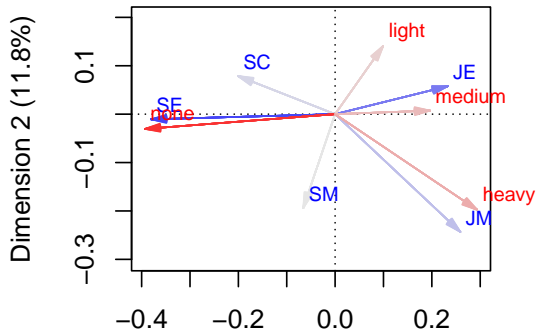
Introduction to correspondence analysis (CA)

- Context: \mathbf{Y} (abundance)
 - Goal: Graphically display the relationships between and/or within the rows and columns

We go from this dataset:

	none	light	medium	heavy
SM	4	2	3	2
JM	4	3	7	4
SE	25	10	12	4
JE	18	24	33	13
SC	10	6	7	2

to this plot:



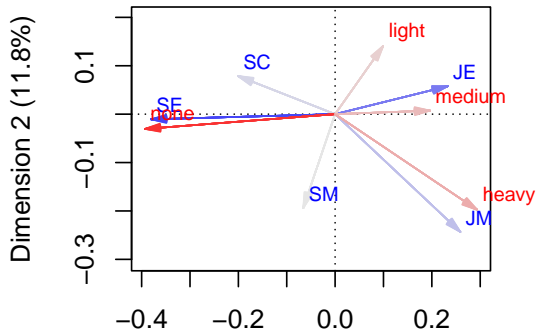
Introduction to correspondence analysis (CA)

- Context: \mathbf{Y} (abundance)
 - Goal: Graphically display the relationships between and/or within the rows and columns

We go from this dataset:

	none	light	medium	heavy
SM	0.36	0.18	0.27	0.18
JM	0.22	0.17	0.39	0.22
SE	0.49	0.20	0.24	0.08
JE	0.20	0.27	0.38	0.15
SC	0.40	0.24	0.28	0.08

to this plot:



Example applications

Datasets

- ▶ Rows are various dams, and columns are counts of waterbird species
- ▶ Rows are various immune compartments (e.g. blood, spleen, lymph), and columns are frequencies of immune cell types (e.g. T cells, B cells, NK cells)
- ▶ Rows are company brands (e.g. Cadbury, Beacon, Lindt), and columns are consumer ratings on a 1-5 scale (e.g. quality, price, taste)

Key characteristics

- ▶ Non-negative
- ▶ Natural zero (i.e. zero means literally nothing and not simply that two quantities are equal, for example)
- ▶ Same units (e.g. counts all in thousands)

The key property of the data is that proportions make sense throughout.

Correspondence matrix, \mathbf{P}

- ▶ Suppose that we have some matrix $\mathbf{X} : I \times J$ where each element
 - ▶ Rows can be thought of as observations and columns as variables
- ▶ The correspondence matrix $\mathbf{P} : I \times J$ is the matrix of overall proportions where

$$P_{ij} = \frac{x_{ij}}{\sum_{i=1}^I \sum_{j=1}^J x_{ij}} = \frac{x_{ij}}{n}$$

We go from \mathbf{X}

	none	light	medium	heavy
SM	4	2	3	2
JM	4	3	7	4
SE	25	10	12	4
JE	18	24	33	13
SC	10	6	7	2

to \mathbf{P}

	none	light	medium	heavy
SM	0.02	0.01	0.02	0.01
JM	0.02	0.02	0.04	0.02
SE	0.13	0.05	0.06	0.02
JE	0.09	0.12	0.17	0.07
SC	0.05	0.03	0.04	0.01

Independence of rows and columns

- ▶ Let \mathbf{r} be the vector of row totals, i.e. $r_i = \sum_{j=1}^J P_{ij} = \mathbf{P}\mathbf{1}$
- ▶ Let \mathbf{c} be the vector of column totals, i.e. $c_j = \sum_{i=1}^I P_{ij} = \mathbf{P}'\mathbf{1}$
- ▶ Then if the rows are independent of the cells, we have that

$$p_{ij} = r_i c_j, \implies \mathbf{P}_{\text{ind}} = \mathbf{r}\mathbf{c}'$$

	none	light	medium	heavy
SM	0.02	0.01	0.02	0.01
JM	0.03	0.02	0.03	0.01
SE	0.08	0.06	0.08	0.03
JE	0.14	0.11	0.15	0.06
SC	0.04	0.03	0.04	0.02

Matrix of residuals

- Under the assumption of independence, we can calculate residuals:

$$\mathbf{P} - \mathbf{P}_{\text{ind}} = \mathbf{P} - \mathbf{r}\mathbf{c}'.$$

- Continuing the smoking example, we then have

\mathbf{P}

	none	light	medium	heavy
SM	0.02	0.01	0.02	0.01
JM	0.02	0.02	0.04	0.02
SE	0.13	0.05	0.06	0.02
JE	0.09	0.12	0.17	0.07
SC	0.05	0.03	0.04	0.01

$\mathbf{P} - \mathbf{r}\mathbf{c}'$

	none	light	medium	heavy
SM	0.003	-0.003	-0.003	0.003
JM	-0.009	-0.006	0.006	0.009
SE	0.046	-0.010	-0.023	-0.014
JE	-0.051	0.018	0.025	0.008
SC	0.011	0.001	-0.005	-0.006

- Residuals are naturally larger for the more abundant rows (employee ranks)

Standardised residuals

- ▶ To avoid the more abundant rows and columns from dominating downstream analyses, we normalise by row and column size.
- ▶ For each residual $P_{ij} - P_{\text{ind}_{ij}}$, we standardise by

$$\frac{P_{ij} - P_{\text{ind}_{ij}}}{\sqrt{r_i c_j}} = \frac{P_{ij} - r_i c_j}{\sqrt{r_i c_j}}$$

- ▶ Define the diagonal matrices $\mathbf{D}_r = \text{diag}(\mathbf{r})$ and $\mathbf{D}_c = \text{diag}(\mathbf{c})$.
- ▶ We then have that the matrix of standardised residuals is given by

$$\mathbf{S} = \mathbf{D}_r^{-1/2}(\mathbf{P} - \mathbf{P}_{\text{ind}})\mathbf{D}_c^{-1/2}.$$

Motivation for this form of residual I: Count distribution

- ▶ Consider the following residual:

$$\frac{\text{Observed} - \text{Expected}}{\sqrt{\text{Expected}}}$$

- ▶ This is equal to a residual with mean 0 and unit variance if the data are Poisson distributed, as for the Poisson distribution, the mean is equal to the variance.
 - ▶ Considering that we are dealing with abundance (e.g. count) data, the Poisson distribution seems appropriate.
 - ▶ We are accounting for the mean-variance relationship.

Motivation for this form of residual II: The χ^2 statistic

- ▶ Now if we square the residuals, we get a χ^2 statistic:

$$\frac{(\text{Observed} - \text{Expected})^2}{\text{Expected}}$$

- ▶ This tracks the deviation from the model with mean Expected and variance Expected.
- ▶ We can calculate this for all the elements in the matrix P under the assumption that P arose under independent rows and columns.
 - ▶ This yields the overall χ^2 statistic when we add them up:

$$\chi^2 = \sum_{i=1}^I \sum_{j=1}^J \frac{(\text{Observed}_{ij} - \text{Expected}_{ij})^2}{\text{Expected}_{ij}}$$

Role of independence assumption

- ▶ The assumption of independence is likely not true, but that is in fact the point:
 - ▶ We've created a way to highlight observation-variable (row-column) combinations that are more common than expected
 - ▶ In particular, the abundance of the row or column is now cancelled out
- ▶ We are *not* performing inference for a test of association between rows and columns
- ▶ Rather, this deviation-from-independence information (i.e. \mathbf{S}) will be used to represent rows in a lower-dimensional space that we can then interpret

Correspondence analysis as PCA on a transformed matrix

- ▶ What we've done up until this point is, essentially, transform the data appropriately
 - ▶ Calculated proportions ($\mathbf{X} \rightarrow \mathbf{P}$)
 - ▶ "Centred" it ($\mathbf{P} - \mathbf{rc}'$)
 - ▶ "Standardised" it ($\mathbf{S} = \mathbf{D}_r^{-1/2}(\mathbf{P} - \mathbf{rc}')\mathbf{D}_c^{-1/2}$)
- ▶ We then use the SVD to obtain a low-rank approximation to \mathbf{S}
 - ▶ Find the low-rank $s < \min(I - 1, J - 1)$ rank matrix $\tilde{\mathbf{S}}$ such that the sum of squared differences is minimised
 - ▶ $\tilde{\mathbf{S}} = \sum_{k=1}^s \lambda_k \mathbf{u}_k \mathbf{v}_k'$
 - ▶ PCA creates the same approximating matrix as the SVD, and so CA is essentially PCA on a transformed matrix
- ▶ Since this is an exploratory technique, we plot the points (transformed again) in a biplot
 - ▶ We have various options for how exactly to calculate the points, which we'll discuss

Views of correspondence analysis

- ▶ The previous description of CA is a bit handwavy, but serves to (help) give an intuition for what CA is doing
- ▶ There are two formal approaches to (or ways of developing) correspondence analysis:
 - ▶ Matrix approximation
 - ▶ Profile approximation

Matrix approximation view of correspondence analysis I

- ▶ The matrix approximation view to CA regards it as solving the following weighted least squares problem:

Matrix approximation angle on CA

- ▶ Find a reduced rank matrix $\hat{\mathbf{P}}$ such that

$$\sum_{i=1}^I \sum_{j=1}^J \frac{(p_{ij} - \hat{p}_{ij})^2}{r_i c_j}$$

is minimised.

- ▶ We upweight errors arising from more commonly observed rows and cells.
 - ▶ This makes sense from the mean-variance relationship we mentioned before.

Matrix approximation view of correspondence analysis II

Since

$$\sum_{i=1}^I \sum_{j=1}^J \frac{(p_{ij} - \hat{p}_{ij})^2}{r_i c_j} = \text{tr}[(\mathbf{D}_r^{-1/2}(\mathbf{P} - \hat{\mathbf{P}})\mathbf{D}_c^{-1/2})(\mathbf{D}_r^{-1/2}(\mathbf{P} - \hat{\mathbf{P}})\mathbf{D}_c^{-1/2})'],$$

we have that

$$\begin{aligned} \sum_{i=1}^I \sum_{j=1}^J \frac{(p_{ij} - \hat{p}_{ij})^2}{r_i c_j} &= \text{tr}[(\mathbf{D}_r^{-1/2}\mathbf{P}\mathbf{D}_c^{-1/2} - \mathbf{D}_r^{-1/2}\hat{\mathbf{P}}\mathbf{D}_c^{-1/2})(\mathbf{D}_r^{-1/2}\mathbf{P}\mathbf{D}_c^{-1/2} - \mathbf{D}_r^{-1/2}\hat{\mathbf{P}}\mathbf{D}_c^{-1/2})'], \\ &= \text{tr}[(\mathbf{D}_r^{-1/2}\mathbf{P}\mathbf{D}_c^{-1/2} - \hat{\mathbf{P}}^*)(\mathbf{D}_r^{-1/2}\mathbf{P}\mathbf{D}_c^{-1/2} - \hat{\mathbf{P}}^*)'], \end{aligned}$$

where $\hat{\mathbf{P}}^* = \mathbf{D}_r^{-1/2}\hat{\mathbf{P}}\mathbf{D}_c^{-1/2}$.

- Thus the weighted least squares problem is essentially the same as an unweighted least squares problem, which we know how to solve - using the SVD on $\mathbf{D}_r^{-1/2}\mathbf{P}\mathbf{D}_c^{-1/2}$.

Results from matrix approximation view I

1. The reduced rank s approximation to \mathbf{P} is given by

$$\sum_{k=1}^s \tilde{\lambda}_k (\mathbf{D}_r^{1/2} \tilde{\mathbf{u}}_k) (\mathbf{D}_c^{1/2} \tilde{\mathbf{v}}_k)',$$

where $\tilde{\lambda}_k$, $\tilde{\mathbf{u}}_k$ and $\tilde{\mathbf{v}}_k$ arise from the SVD of $\mathbf{D}_r^{-1/2} \mathbf{P} \mathbf{D}_c^{-1/2}$ and the approximation error is $\sum_{k=s+1}^J \tilde{\lambda}_k^2$.

2. We always have that

$$\tilde{\lambda}_k (\mathbf{D}_r^{1/2} \tilde{\mathbf{u}}_1) (\mathbf{D}_c^{1/2} \tilde{\mathbf{v}}_1) = \mathbf{r} \mathbf{c}'.$$

Results from matrix approximation view II

3. The reduced rank $K > 1$ approximation to $\mathbf{P} - \mathbf{r}\mathbf{c}'$ is given by

$$\sum_{k=1}^K \lambda_k (\mathbf{D}_r^{1/2} \mathbf{u}_k) (\mathbf{D}_c^{1/2} \mathbf{v}_k)',$$

where λ_k , \mathbf{u}_k and \mathbf{v}_k arise from the SVD of $\mathbf{D}_r^{-1/2}(\mathbf{P} - \mathbf{r}\mathbf{c}')\mathbf{D}_c^{-1/2}$.

4. $\mathbf{D}_r^{-1/2}\mathbf{P}\mathbf{D}_c^{-1/2}$ and $\mathbf{D}_r^{-1/2}(\mathbf{P} - \mathbf{r}\mathbf{c}')\mathbf{D}_c^{-1/2}$ share singular vectors and singular values, in that $\tilde{\lambda}_{k+1} = \lambda_k$, $\tilde{u}_{k+1} = u_k$ and $\tilde{v}_{k+1} = v_k$.

Comment on the previous results

- ▶ It's a lot - but I mention them because they shed light on a couple of things in CA, rather than that you memorise them all.
- ▶ The main point is that CA is an application of PCA on a transformed matrix, which is equivalent to a least-squares problem whose solution we find by the singular -vectors and -values of the matrix of standardised residuals,
$$\mathbf{S} = \mathbf{D}_r^{-1/2}(\mathbf{P} - \mathbf{rc}')\mathbf{D}_c^{-1/2}.$$

Inertia

- ▶ For $\mathbf{S} = \mathbf{D}_r^{-1/2}(\mathbf{P} - \mathbf{r}\mathbf{c}')\mathbf{D}_c^{-1/2}$, total inertia is the weighted sum of squares of residuals:

$$\text{tr}[\mathbf{S}\mathbf{S}'] = \sum_i \sum_j \frac{(p_{ij} - r_i c_j)^2}{r_i c_j} = \sum_{k=1}^{J-1} \lambda_k^2$$

- ▶ λ_k is the k -th singular value of \mathbf{S}
 - ▶ This follows from the fact that $\mathbf{r}\mathbf{c}'$ is in fact the best rank-one approximation to \mathbf{P}
 - ▶ It is therefore the first term in the SVD approximation, and the rank one SVD matrix's approximation error is equal to $\sum_{k=2}^J \tilde{\lambda}_k^2 = \sum_{k=1}^{J-1} \lambda_k^2$.
- ▶ The inertia captured by the first s components represents the variation captured in \mathbf{S} , and the proportion of variation captured is given by $(\sum_{k=1}^s \lambda_k^2) / (\sum_{k=1}^{J-1} \lambda_k^2)$.

Examining the raw data using CA concepts

```
> CrossTable(smokelong$Rank, smokelong$smoke)
```

```
Cell Contents
-----
N
Chi-square contribution
N / Row Total
N / Col Total
N / Table Total
-----
```

Total Observations in Table: 193

smokelong\$Rank	1	2	3	4	Total
1	4	2	3	2	11
SM	0.079	0.124	0.081	0.232	0.057
	0.364	0.182	0.273	0.182	
	0.066	0.044	0.048	0.080	
	0.021	0.010	0.016	0.010	
4	4	3	7	4	18
JM	0.502	0.341	0.256	1.194	0.093
	0.222	0.167	0.389	0.222	
	0.066	0.067	0.113	0.160	
	0.021	0.016	0.036	0.021	
3	25	10	12	4	51
SE	4.893	0.301	1.173	1.028	0.264
	0.490	0.196	0.235	0.078	
	0.410	0.222	0.194	0.160	
	0.130	0.052	0.062	0.021	
4	18	24	33	13	88
JE	3.463	0.591	0.792	0.225	0.456
	0.205	0.273	0.375	0.148	
	0.295	0.533	0.532	0.520	
	0.093	0.124	0.171	0.067	
5	10	6	7	2	25
SEC	0.557	0.005	0.132	0.474	0.130
	0.400	0.240	0.280	0.080	
	0.164	0.133	0.113	0.080	
	0.052	0.031	0.036	0.010	
Column Total	61	45	62	25	193
	0.316	0.233	0.321	0.130	

Third row = row profile=row percentages =distribution of row frequencies across columns

Row margins = percentage of overall mass/total accounted for by each row. So JE and SE contribute more than other rows.

Fourth entry in every column cell = column profiles = column percentages=distribution of column frequencies across rows

Column margins = percentage of overall mass/total accounted for by each column. Even distribution across columns, except for column 4.

$$\left(\frac{o_{ij} - e_{ij}}{\sqrt{e_{ij}}} \right)^2$$

Calculating the solution to the least-squares problem

```
> #CA from first principles
```

```
> N<-smoke> N
```

	none	light	medium	heavy
SM	4	2	3	2
JM	4	3	7	4
SE	25	10	12	4
JE	18	24	33	13
SC	10	6	7	2

Show how 1 unit of mass is distributed across cells

```
> P<-N/sum(N)
```

```
> P
```

	none	light	medium	heavy
SM	0.02072539	0.01036269	0.01554404	0.01036269
JM	0.02072539	0.01554404	0.03626943	0.02072539
SE	0.12953368	0.05181347	0.06217617	0.02072539
JE	0.09326425	0.12435233	0.17098446	0.06735751
SC	0.05181347	0.03108808	0.03626943	0.01036269

Or across rows, = row margins or "masses"

```
> rm<-apply(P,1,sum)
```

```
> rm
```

	SM	JM	SE	JE	SC
	0.05699482	0.09326425	0.26424870	0.45595855	0.12953368

```
> cm<-apply(P,2,sum)
```

```
> cm
```

	none	light	medium	heavy
	0.3160622	0.2331606	0.3212435	0.1295337

Or across columns, = column margins or "masses"

```
> Dr<-diag(rm)
```

```
> Dr
```

	[,1]	[,2]	[,3]	[,4]	[,5]
[1,]	0.05699482	0.00000000	0.00000000	0.00000000	0.00000000
[2,]	0.00000000	0.09326425	0.00000000	0.00000000	0.00000000
[3,]	0.00000000	0.00000000	0.26424870	0.00000000	0.00000000
[4,]	0.00000000	0.00000000	0.00000000	0.45595855	0.00000000
[5,]	0.00000000	0.00000000	0.00000000	0.00000000	0.1295337

```
> Dc<-diag(cm)
```

```
> Dc
```

	[,1]	[,2]	[,3]	[,4]
[1,]	0.3160622	0.00000000	0.00000000	0.00000000
[2,]	0.00000000	0.2331606	0.00000000	0.00000000
[3,]	0.00000000	0.00000000	0.3212435	0.00000000
[4,]	0.00000000	0.00000000	0.00000000	0.1295337

```
> S<-diag(sqrt(1/rm))*%(as.matrix(P)-rm%*%t(cm))*%diag(sqrt(1/cm))
```

```
> svdS<-svd(S)
```

```
> svdS
```

```
$d
```

```
[1] 2.734211e-01 1.000859e-01 2.033652e-02 5.571478e-03
```

Taking SVD of weighted difference between observed and expected relative frequencies.

```
$u
```

	[,1]	[,2]	[,3]	[,4]
[1,]	-0.05742524	-0.46212293	0.8332653	-0.09239316
[2,]	0.28923816	-0.74239515	-0.5061482	-0.17736175
[3,]	-0.71554563	-0.05475038	-0.1303234	-0.68022273
[4,]	0.57530335	0.38957951	0.1097504	-0.67979717
[5,]	-0.26469630	0.28376408	-0.1430158	0.18756108

```
$v
```

	[,1]	[,2]	[,3]	[,4]
[1,]	-0.8087001	-0.17127755	-0.0246170	0.5621941
[2,]	0.1756411	0.68056865	0.5223178	0.4828671
[3,]	0.4069601	0.04167443	-0.7151246	0.5667835
[4,]	0.3867013	-0.71116353	0.4638695	0.3599079

Principal coordinates for plotting, though they usually are scaled.

```
> smoke_F<-diag(1/sqrt(rm))*%svdS$u*%diag(svdS$d)
```

```
> smoke_F
```

	[,1]	[,2]	[,3]	[,4]
[1,]	-0.06576838	-0.19373700	0.070981028	-2.156217e-17
[2,]	0.25895842	-0.24330457	-0.033705190	-3.235734e-17
[3,]	-0.38059489	-0.01065991	-0.005155757	-7.372506e-17
[4,]	0.23295191	0.05774391	0.003305371	-5.609021e-17
[5,]	-0.20108912	0.07891123	-0.008081076	2.903500e-17

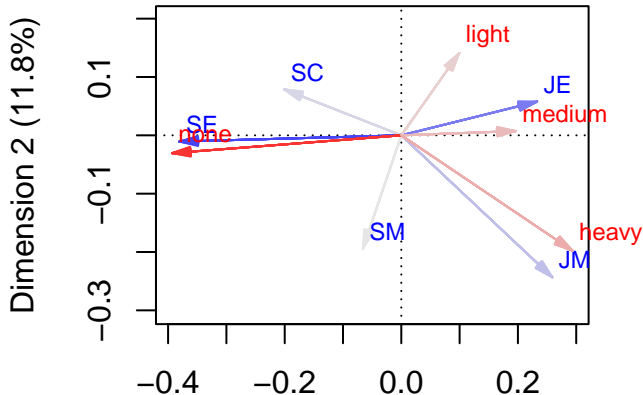
```
> smoke_G<-diag(1/sqrt(cm))*%svdS$v
```

```
> smoke_G
```

	[,1]	[,2]	[,3]	[,4]
[1,]	-1.4384714	-0.30465911	-0.04378737	1
[2,]	0.3637463	1.40943267	1.08170100	1
[3,]	0.7180168	0.07352795	-1.26172451	1
[4,]	1.0744451	-1.97595989	1.28885615	1

Displaying the results of a correspondence analysis

- ▶ We've discussed the first three steps in performing CA:
 1. Calculating the matrix of standardised residuals, $\mathbf{S} = \mathbf{D}_r^{-1/2}(\mathbf{P} - \mathbf{rc}')\mathbf{D}_c^{-1/2}$.
 2. Calculating SVD of this (to obtain its least-squares approximation).
 3. Assessing the quality of the dimensionality reduction (via inertia).
- ▶ We now turn our attention to displaying the rows and columns:



Choosing coordinate scales

- ▶ A biplot is a graphical display of a matrix $\mathbf{X}_{m \times n}$ such that

$$\mathbf{X}_{m \times n} = \mathbf{F}_{m \times k}(\mathbf{G}_{n \times k})' \text{ for } k \leq \min(m, n),$$

where each row of \mathbf{F} is a lower-dimensional representation of a row in \mathbf{X} and each column of \mathbf{G} is a lower-dimensional representation of a column in \mathbf{X} .

- ▶ Here we term \mathbf{F} and \mathbf{G} the *row* and *column* coordinates, respectively.
- ▶ Clearly, in our case $\mathbf{X} = \mathbf{S} = \mathbf{D}_r^{-1/2}(\mathbf{P} - \mathbf{r}\mathbf{c}')\mathbf{D}_c^{-1/2}$.

Principal vs standard coordinates

- ▶ We need to choose the scaling for our row coordinates. Typical choices:
 - ▶ *Principal* coordinates:
 - ▶ Rows: $\mathbf{F} = \mathbf{D}_r^{-1/2} \mathbf{U} \mathbf{\Lambda}$
 - ▶ Columns: $\mathbf{G} = \mathbf{D}_c^{-1/2} \mathbf{V} \mathbf{\Lambda}$
 - ▶ *Standard* coordinates:
 - ▶ Rows: $\mathbf{F} = \mathbf{D}_r^{-1/2} \mathbf{U}$
 - ▶ Columns: $\mathbf{G} = \mathbf{D}_c^{-1/2} \mathbf{V}$
- ▶ When rows are plotted using principal coordinates, the χ^2 distance between the rows is optimally displayed. Similarly for columns.

Symmetric vs asymmetric biplots

- ▶ In symmetric biplots, the both rows and columns are plotted using principal coordinates, or both rows and columns are plotted using standard coordinates.
 - ▶ This is great for displaying relationships between rows, or between columns. But:
 - ▶ Relationships between rows and columns are not necessarily displayed correctly.
 - ▶ It's not a true biplot, as \mathbf{X} is not even approximately equal to $\mathbf{F}\mathbf{G}'$.
- ▶ In asymmetric biplots, one uses principal coordinates and the other uses standard coordinates.
 - ▶ The motivation is that the relationships between rows and columns are more accurately displayed.
 - ▶ However:
 - ▶ The standard and principal coordinates may be on fairly different scales, depending on the size of the singular values.
 - ▶ Neither distances between rows nor distances between columns are optimally displayed.
- ▶ However, at the end of the day it's all quite fuzzy as even principal coordinates are approximations.
 - ▶ It seems to me like dividing Λ up between the row and column coordinates (e.g. $\Lambda^{0.5}$ for both) is the best option unless you have a particular interest in either the rows or columns.

Terminology for scaling choices:

- ▶ Both principal: symmetric
- ▶ Row principal and column standard: row-principal
- ▶ Column principal and row standard: column-principal

ca.plot function

Example: symmetric plot

Symmetric Plots:

NORMALIZATION of coordinates determines whether and how the similarity of the row categories, the similarity of the column categories and the relationship between the row and column variables can be interpreted in terms of the row and column coordinates and the origin of the plot.

The Euclidean distance between the row points approximates the chi-squared distances between the corresponding row profiles (i.e., the distributions on the column variables are similar). . So categories mapped close together have similar row profiles

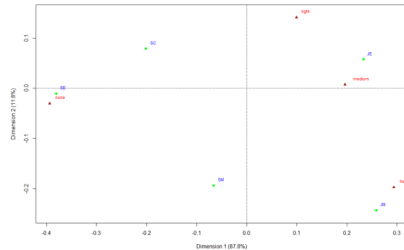
The Euclidean distance between a row point and the origin approximates the chi-square distance from the row profile to the row centroid, indicating how different a category is from the population.

Similarly for column points.

Cannot interpret distances between rows and columns except in a very general sense.

```
> plot(ca(smoke),map="symmetric",col=c("green","brown"))
```

- Rows and columns plotted using principal coordinates but scaled so that each have variance equal to singular values.
- Can interpret distances between rows and distances between columns but not distances between rows and columns.



- First axis contrasts “none” to other smoking categories
- First axis contrasts SE to JE&JM
- SE lies in proximate area to “none”
- JE lies in proximate area to “medium”
- JM lies in proximate area to “heavy”

But cannot interpret actual distances between rows and cols

Example: asymmetric plot I

Asymmetric biplots

ROWPRINCIPAL

You can choose to plot the rows using principal coordinates and the columns as scaled standard coordinates. This is referred to as a row-preserving metric and represent the columns in row space. Distances between columns will not be correct.

COLUMN PRINCIPAL

You can choose to plot the columns using principal coordinates and the rows as scaled standard coordinates. This is referred to as a column-preserving metric and represent the rows in column space. Distances between rows will not be correct.

The angles between arrows representing rows and columns give an indication of associations. The more acute, the more strongly associated.

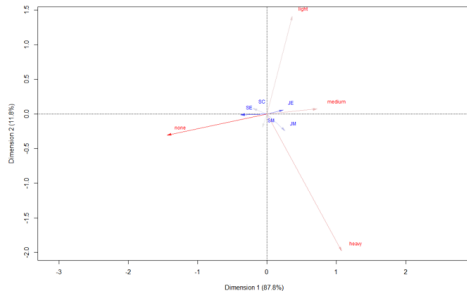
A further scaling according to Greenacre allows for interpretation of contribution of columns or rows to dimensions.

Example: asymmetric plot II

Row principal plots:

```
> plot(ca(smoke), mass = TRUE, contrib = "absolute", map = "rowprincipal", arrows = c(TRUE, TRUE))
```

Rows in principal coordinates, column in standard coordinates



Based on angles.

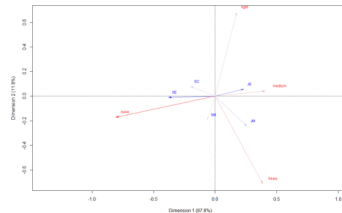
So SE close to “none”.

And JE close to “medium”.

Whereas SM and JM closer to “heavy”.

```
> plot(ca(smoke), mass = TRUE, contrib = "absolute",  
map = "rowgreen", arrows = c(TRUE, TRUE))
```

Rows in principal and columns in standard coordinates, columns scaled to reflect contribution to inertia through distance from origin. Allows interpretation of contribution of columns to axes.



“none and “medium” close to dimension 1

“light” close to dimension 2

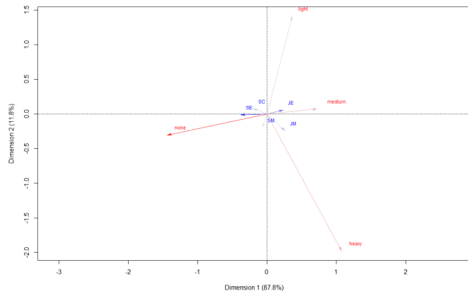
Column points closer to origin contributed less than column point further away.

Example: asymmetric plot III

Row principal plots:

```
> plot(ca(smoke), mass = TRUE, contrib = "absolute", map = "rowprincipal", arrows = c(TRUE, TRUE))
```

Rows in principal coordinates, column in standard coordinates



Based on angles.

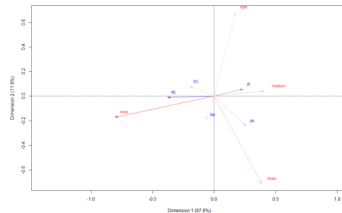
So SE close to “none”.

And JE close to “medium”.

Whereas SM and JM closer to “heavy”.

```
> plot(ca(smoke), mass = TRUE, contrib = "absolute",  
map = "rowgreen", arrows = c(TRUE, TRUE))
```

Rows in principal and columns in standard coordinates, columns scaled to reflect contribution to inertia through distance from origin. Allows interpretation of contribution of columns to axes.



“none” and “medium” close to dimension 1

“light” close to dimension 2

Column points closer to origin contributed less than column point further away.

Example: quality of approximation

```
> summary(ca(smoke))
```

Principal inertias (eigenvalues):

dim	value	%	cum%	scree plot
1	0.074759	87.8	87.8	*****
2	0.010017	11.8	99.5	***
3	0.000414	0.5	100.0	

Total: 0.085190 100.0

λ^2 analysis

Quality in each dimension = the correlation of point with dimension

Ctr refers to contribution of each row/ column to specific dimension

Rows:

	name	mass	qlt	inr	k=1	cor	ctr	k=2	cor	ctr
1	SM	57	893	31	-66	92	3	-194	800	214
2	JM	93	991	139	259	526	84	-243	465	551
3	SE	264	1000	450	-381	999	512	-11	1	3
4	JE	456	1000	308	233	942	331	58	58	152
5	SC	130	999	71	-201	865	70	79	133	81

Row totals*1000

Qlt = Quality =

$$\frac{\text{squared distance of point from origin in 2 dimensions}}{\text{squared distance of point from origin in maximum \# dimensions}}$$

Columns:

	name	mass	qlt	inr	k=1	cor	ctr	k=2	cor	ctr
1	none	316	1000	577	-393	994	654	-30	6	29
2	lght	233	984	83	99	327	31	141	657	463
3	medm	321	983	148	196	982	166	7	1	2
4	hevy	130	995	192	294	684	150	-198	310	506

inr=Relative Inertia

=proportion of inertia accounted for by respective point

$k = 1$ & $k = 2$ are the plotting coordinates

SM not well presented by dimension 1 because cor very low
SE not well presented by dimension 2 because cor very low

Multiple correspondence analysis (MCA)

For multilevel contingency tables when you are cross-tabulating more than two variables.

There are different approaches for adjusting Correspondence Analysis to cope with this, including

1. CA on an indicator matrix
2. CA on a Burt matrix
3. Adjusted MCA through rescaling
4. Joint Correspondence Analysis

For example, consider 10 observations on 3 categorical variables, w , x and y , each with 2 categories:

Then you can form the indicator matrix as follows:

$$Z = \begin{matrix} & w_1 & w_2 & x_1 & x_2 & y_1 & y_2 \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \\ 10 \end{matrix} & \begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \end{pmatrix} \end{matrix}$$

And the Burt matrix corresponding to Z is:

$$B = Z^T Z = \begin{matrix} & w_1 & w_2 & x_1 & x_2 & y_1 & y_2 \\ \begin{matrix} w_1 \\ w_2 \\ x_1 \\ x_2 \\ y_1 \\ y_2 \end{matrix} & \begin{pmatrix} 7 & 0 & 4 & 3 & 3 & 4 \\ 0 & 3 & 1 & 2 & 3 & 0 \\ 4 & 1 & 5 & 0 & 3 & 2 \\ 3 & 2 & 0 & 5 & 3 & 2 \\ 3 & 3 & 3 & 3 & 6 & 0 \\ 4 & 0 & 2 & 2 & 0 & 4 \end{pmatrix} \end{matrix}$$

There were 7
observations with
 $w_1=1$,
4 observations with
 $w_1=1$ and $x_1=1$, etc.

That summarises the number of observations with the pairwise frequencies.

MCA: example introduction

wg93 {ca}R Documentation

International Social Survey Program on Environment 1993

This data frame contains 871 records of four questions on attitude towards science with responses on a five-point scale (1=agree strongly to 5=disagree strongly) and three demographic variables (sex, age and education).

The questions were:

A: We believe too often in science, and not enough in feelings and faith.

B: Overall, modern science does more harm than good.

C: Any change humans cause in nature, no matter how scientific, is likely to make things worse.

D: Modern science will solve our environmental problem with little change to our way of life.

Source: ISSP (1993). International Social Survey Program: Environment. <http://www.issp.org>

```
> data("wg93")
```

```
> head(wg93)
```

	A	B	C	D	sex	age	<u>edu</u>
1	2	3	4	3	2	2	3
2	3	4	2	3	1	3	4
3	2	3	2	4	2	3	2
4	2	2	2	2	1	2	3
5	3	3	3	3	1	5	2
6	3	4	4	5	1	3	2

So we have 4 categorical variables
each with 5 levels,
Plus some additional demographic
characteristics.

The 4 categorical variables can each
turn into 5 binary indicator variables,
for example A1(0/1),
A2(0/1), A3(0/1), A4(0/1), A5(0/1), etc.
The first observation will have A2=1
and all the other A-variables =0, etc.

MCA: example application

```
> summary(mjca(wg93[,1:4],lambda="adjusted"))
```

Principal inertias (eigenvalues):

dim	value	%	cum%	scree plot
1	0.076455	44.9	44.9	*****
2	0.058220	34.2	79.1	*****
3	0.009197	5.4	84.5	**
4	0.005670	3.3	87.8	*
5	0.001172	0.7	88.5	
6	7e-06000	0.0	88.5	

Total: 0.170246

Use mjca
function
with
"adjusted"
option

First 2
dimensions
account for 79%
of associations
among
responses to 4
questions.

Now only have
column variables
since all variables
have been
turned into
column variables.

Some relatively
low correlations
of columns with
dimensions 1
and 2

For the BURT matrix, the main diagonal is a cross-tabulation of each variable with itself, thus the solution will overestimate the total inertia.

To correct for this, MCA rescales the coordinates to best fit the pairwise cross-tabulations on the off-diagonals of the Burt matrix. Use the lambda="adjusted" option which is also the default option.

Another option is to use Joint Correspondence Analysis that uses a different algorithm to find an optimal least squares fit to the off-diagonal elements.

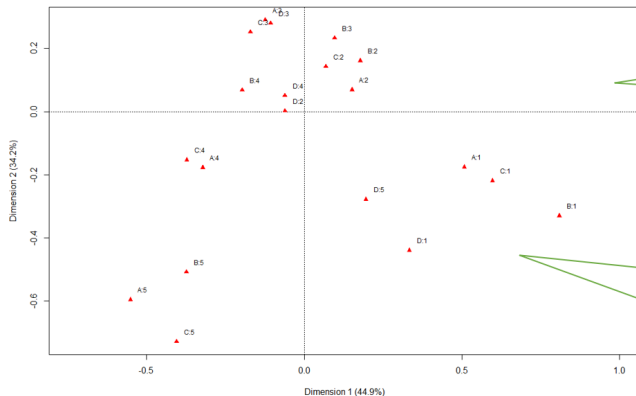
D3 and D4 not
well represented
in first 2
dimensions

Columns:

	name	mass	glt	inr	k=1 cor	ctr	k=2 cor	ctr
1	A:1	34	963	55	508	860	115	-176 103 18
2	A:2	92	659	38	151	546	28	69 113 7
3	A:3	59	929	47	-124	143	12	289 786 84
4	A:4	51	798	50	-322	612	69	-178 186 28
5	A:5	14	799	60	-552	369	55	-596 430 84
6	B:1	20	911	62	809	781	174	-331 131 38
7	B:2	50	631	47	177	346	21	161 285 22
8	B:3	59	806	45	96	117	7	233 690 55
9	B:4	81	620	41	-197	555	41	68 65 6
10	B:5	40	810	60	-374	285	74	-509 526 179
11	C:1	44	847	60	597	746	203	-219 101 36
12	C:2	91	545	38	68	101	6	143 444 32
13	C:3	57	691	48	-171	218	22	252 473 62
14	C:4	44	788	52	-373	674	80	-153 114 18
15	C:5	15	852	60	-406	202	32	-728 650 136
16	D:1	17	782	56	333	285	25	-440 497 57
17	D:2	57	126	42	-61	126	3	2 0 0
18	D:3	58	688	48	-106	87	9	280 601 78
19	D:4	65	174	43	-61	103	3	51 71 3
20	D:5	43	869	50	196	288	22	-278 581 57

MCA: example plot

```
plot(mca(wg93[,1:4]), mass = TRUE, contrib = "absolute", map = "colprincipal", arrows = c(TRUE, TRUE))
```

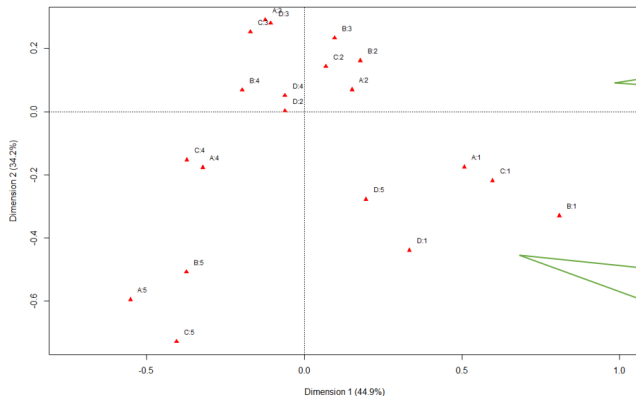


All column variables so can interpret proximity to each other. Horse shoe effect often seen

Looks as if first dimension orders level of response to all questions, grouping 5's together, etc to 1's on right hand side. And second dimension contrasts extreme responses (1 and 5) to more moderate responses.

Summary

```
plot(mica(wg93[,1:4]), mass = TRUE, contrib = "absolute", map = "colprincipal", arrows = c(TRUE, TRUE))
```



All column variables so can interpret proximity to each other. Horse shoe effect often seen

Looks as if first dimension orders level of response to all questions, grouping 5's together, etc to 1's on right hand side. And second dimension contrasts extreme responses (1 and 5) to more moderate responses.