ORIGIN ≡ 1

CORRESPONDENCE ANALYSIS

Correspondence Analysis is primarily a gaphical technique designed to show relationships between variables constituting rows and columns of a contingency table. For instance, in the example below, rows may represent different kinds of trees in a forest, columns may represent different forest localities, and the numbers represent counts of each tree observed in each locality. A typical result is a biplot graph showing points (here representing trees = rows) indicating relative similarity in the distribution of trees among localities. In addition often represented are vectors (here representing localities = columns) indicating relative similarity in occurrence of trees among localities. Close plotting of a vector (head) and point indicate greater association of tree with specific locality than expected at "random". As in PCA, a measure (here called "inertia") indicates relative quality of fit beteen the reduced dimensions of the plot versus the full dimensions of the problem. Data is from RA Johnson & DW Wichern *Applied Multivariate Statistical Analysis 4th Edition* 1998.

Read in Data:

Eight Tree species (rows) counted within plots of 10 localities (columns).

M := READPRN("c:\DATA\Multivariate\T12-10.DAT")

$r := \text{rows}(M)$ $r = 8$

$c := \text{cols}(M)$ $c = 10$

$i := 1..r$ $j := 1..c$

$l_{r_i} := 1$ $l_{c_j} := 1$

$T := \sum_i \sum_j M_{i,j}$ $T = 292$ **< Total number of observations in matrix M:**

M =

	1	2	3	4	5	6	7	8	9	10
1	9	8	3	5	6	0	5	0	0	0
2	8	9	8	7	0	0	0	0	0	0
3	5	4	9	9	7	7	4	6	0	2
4	3	4	0	6	9	8	7	6	4	3
5	2	2	4	5	6	0	5	0	2	5
6	0	0	0	0	2	7	6	6	7	6
7	0	0	0	0	0	0	7	4	6	5
8	0	0	0	0	0	5	4	8	8	9

Matrix P of frequencies:

$P := \frac{1}{T} M$

^ Matrix P is called the "correspondence matrix"

$$P = \begin{pmatrix} 0.0308 & 0.0274 & 0.0103 & 0.0171 & 0.0205 & 0 & 0.0171 & 0 & 0 & 0 \\ 0.0274 & 0.0308 & 0.0274 & 0.024 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.0171 & 0.0137 & 0.0308 & 0.0308 & 0.024 & 0.024 & 0.0137 & 0.0205 & 0 & 0.0068 \\ 0.0103 & 0.0137 & 0 & 0.0205 & 0.0308 & 0.0274 & 0.024 & 0.0205 & 0.0137 & 0.0103 \\ 0.0068 & 0.0068 & 0.0137 & 0.0171 & 0.0205 & 0 & 0.0171 & 0 & 0.0068 & 0.0171 \\ 0 & 0 & 0 & 0 & 0.0068 & 0.024 & 0.0205 & 0.0205 & 0.024 & 0.0205 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.024 & 0.0137 & 0.0205 & 0.0171 \\ 0 & 0 & 0 & 0 & 0 & 0.0171 & 0.0137 & 0.0274 & 0.0274 & 0.0308 \end{pmatrix}$$

Matrix P_s of Centered & Scaled frequencies:

$R := P \cdot l_c$

$C := P^T \cdot l_r$

Row (R) and Column (C) vectors of sums of matrix P:

$IsD_r := \text{diag}\left(\frac{1}{\sqrt{R}}\right)$

$IsD_c := \text{diag}\left(\frac{1}{\sqrt{C}}\right)$

< Inverse square-root matrices with values along the main diagonal

$$R = \begin{pmatrix} 0.123 \\ 0.11 \\ 0.182 \\ 0.171 \\ 0.106 \\ 0.116 \\ 0.075 \\ 0.116 \end{pmatrix} \qquad C = \begin{pmatrix} 0.09247 \\ 0.09247 \\ 0.08219 \\ 0.10959 \\ 0.10274 \\ 0.09247 \\ 0.13014 \\ 0.10274 \\ 0.09247 \\ 0.10274 \end{pmatrix}$$

$P_s := IsD_r \cdot \left(P - R \cdot C^T\right) \cdot IsD_c$ **< centered & scaled correspondence matrix.**

$$P_s = \begin{pmatrix} 0.1819 & 0.1498 & 0.0014 & 0.0311 & 0.07 & -0.1068 & 0.0085 & -0.1125 & -0.1068 & -0.1125 \\ 0.1715 & 0.2055 & 0.1938 & 0.1092 & -0.1061 & -0.1007 & -0.1194 & -0.1061 & -0.1007 & -0.1061 \\ 0.0026 & -0.0238 & 0.1302 & 0.0775 & 0.039 & 0.0555 & -0.0646 & 0.0139 & -0.1295 & -0.0864 \\ -0.0442 & -0.017 & -0.1186 & 0.013 & 0.0997 & 0.0919 & 0.0113 & 0.0223 & -0.017 & -0.0552 \\ -0.0299 & -0.0299 & 0.0532 & 0.0509 & 0.0923 & -0.0991 & 0.0281 & -0.1044 & -0.0299 & 0.0595 \\ -0.1038 & -0.1038 & -0.0978 & -0.113 & -0.0468 & 0.1273 & 0.0438 & 0.0785 & 0.1273 & 0.0785 \\ -0.0835 & -0.0835 & -0.0787 & -0.0909 & -0.088 & -0.0835 & 0.1431 & 0.0677 & 0.1627 & 0.1066 \\ -0.1038 & -0.1038 & -0.0978 & -0.113 & -0.1094 & 0.0613 & -0.0118 & 0.1411 & 0.1603 & 0.1724 \end{pmatrix}$$

Singular Value Decomposition of matrix P_s:

$\gamma_U := \text{reverse}\left(\text{sort}\left(\text{eigenvals}\left(P_s \cdot P_s^T\right)\right)\right)$

$\gamma_V := \text{reverse}\left(\text{sort}\left(\text{eigenvals}\left(P_s^T \cdot P_s\right)\right)\right)$

$U^{\langle i \rangle} := \text{eigenvec}\left(P_s \cdot P_s^T, \gamma_{U_i}\right)$

$V^{\langle j \rangle} := \text{eigenvec}\left(P_s^T \cdot P_s, \gamma_{V_j}\right)$

$\lambda := \sqrt{\gamma_U}$ **< singular values**

$$\gamma_U = \begin{pmatrix} 0.537 \\ 0.096 \\ 0.072 \\ 0.046 \\ 0.011 \\ 4.539 \times 10^{-3} \\ 3.882 \times 10^{-3} \\ 0 \end{pmatrix} \quad \gamma_V = \begin{pmatrix} 0.53673 \\ 0.09618 \\ 0.0721 \\ 0.04555 \\ 0.01107 \\ 0.00454 \\ 0.00388 \\ 0 \\ 0 \\ 0 \end{pmatrix} \quad \lambda = \begin{pmatrix} 0.7326 \\ 0.3101 \\ 0.2685 \\ 0.2134 \\ 0.1052 \\ 0.0674 \\ 0.0623 \\ 3.0039 \times 10^{-9} \end{pmatrix}$$

$$U = \begin{pmatrix} -0.39042 & 0.08311 & 0.47814 & -0.4562 & -0.03768 & -0.33691 & -0.40707 & 0.35112 \\ -0.53268 & 0.4985 & -0.40802 & -0.09251 & -0.07379 & 0.34196 & 0.24645 & 0.33104 \\ -0.19986 & -0.38891 & -0.40889 & 0.36216 & 0.43908 & -0.32169 & -0.18076 & 0.42604 \\ 0.06979 & -0.5382 & 0.17259 & -0.31809 & -0.05439 & 0.1596 & 0.61224 & 0.4138 \\ -0.08195 & 0.01512 & 0.42706 & 0.7086 & -0.41601 & 0.16851 & -0.03071 & 0.32583 \\ 0.40046 & -0.08307 & -0.14776 & -0.18661 & -0.00416 & 0.58951 & -0.55865 & 0.34123 \\ 0.36336 & 0.48503 & 0.32324 & 0.09371 & 0.62981 & -0.01639 & 0.21722 & 0.27449 \\ 0.46892 & 0.24757 & -0.315 & -0.07263 & -0.47713 & -0.51421 & 0.07627 & 0.34123 \end{pmatrix}$$

$$V = \begin{pmatrix} 0.38774 & -0.21076 & 0.06164 & -0.40292 & -0.0582 & -0.32693 & 0.42471 & 0.16413 & 0.16413 & 0.16417 \\ 0.38558 & -0.24276 & 0.01057 & -0.43452 & -0.19499 & 0.19683 & -0.26351 & 0.43824 & 0.43824 & 0.4382 \\ 0.34953 & -0.18206 & -0.40795 & 0.57182 & 0.23432 & 0.11673 & 0.32936 & 0.34774 & 0.34774 & 0.34774 \\ 0.30064 & 0.13553 & -0.05396 & 0.26465 & 0.00063 & 0.08255 & -0.66442 & 0.17881 & 0.17882 & 0.17884 \\ 0.11083 & 0.58167 & 0.48562 & 0.15977 & -0.23333 & -0.16067 & 0.07718 & 0.39287 & 0.39287 & 0.39288 \\ -0.20218 & 0.54005 & -0.46257 & -0.26868 & -0.09782 & 0.39427 & 0.26684 & 0.2543 & 0.25431 & 0.25431 \\ -0.18519 & -0.07562 & 0.50896 & 0.02909 & 0.60261 & 0.19554 & 0.15201 & 0.32748 & 0.32748 & 0.32747 \\ -0.31395 & 0.06439 & -0.33937 & -0.15671 & 0.33659 & -0.65731 & -0.25071 & 0.36671 & 0.36671 & 0.36671 \\ -0.42003 & -0.34836 & 0.0394 & -0.11651 & -0.06249 & 0.37722 & -0.14565 & 0.29052 & 0.29052 & 0.29056 \\ -0.35495 & -0.28969 & 0.0345 & 0.33932 & -0.59944 & -0.20021 & 0.12616 & 0.28805 & 0.28805 & 0.28804 \end{pmatrix}$$

Inertia:

$$TI := \sum_i \gamma U_i \qquad TI = 0.77$$

$$\gamma U = \begin{pmatrix} 0.5367 \\ 0.0962 \\ 0.0721 \\ 0.0455 \\ 0.0111 \\ 0.0045 \\ 0.0039 \\ 0 \end{pmatrix} \quad \lambda = \begin{pmatrix} 0.7326 \\ 0.3101 \\ 0.2685 \\ 0.2134 \\ 0.1052 \\ 0.0674 \\ 0.0623 \\ 3.0039 \times 10^{-9} \end{pmatrix} \quad \frac{1}{TI}\cdot\lambda^2 = \begin{pmatrix} 0.697 \\ 0.1249 \\ 0.0936 \\ 0.0591 \\ 0.0144 \\ 0.0059 \\ 0.005 \\ 0 \end{pmatrix}$$

< Proportion of Inertia for each singular value squared $\lambda^2 = \gamma U$

$$0.697 + 0.1249 = 0.822$$

^ first two axes "account' for 82.2% of the fit

Correspondence analysis coordinates:

$D_r := \text{diag}(R) \qquad Ds_r := \text{diag}(\sqrt{R})$

$D_c := \text{diag}(C) \qquad Ds_c := \text{diag}(\sqrt{C})$

< diagonal R & C matrices and diagonal square root matrices

$m := \text{rows}(P_s) \qquad k := \text{cols}(P_s)$

$i := 1 .. k - m \qquad k - m = 2 \qquad Z_{m,i} := 0$

< additional columns of zeros for use with matrix algebra

$\Lambda := \text{augment}(\text{diag}(\lambda), Z)$

$F := D_r^{-1}\cdot(Ds_r\cdot U)\cdot\Lambda$

Matrix F represents vector positions for ROWS of M

Note: typically the first two (or three) columns of F and G are plotted together...

See plot below

$$F = \begin{pmatrix}
-0.8146 & 0.0734 & 0.3656 & -0.2773 & -0.0113 & -0.0646 & -0.0722 & 3.0039\times10^{-9} & 0 & 0 \\
-1.1789 & 0.467 & -0.3309 & -0.0596 & -0.0234 & 0.0696 & 0.0464 & 3.0039\times10^{-9} & 0 & 0 \\
-0.3437 & -0.2831 & -0.2577 & 0.1814 & 0.1084 & -0.0509 & -0.0264 & 3.0039\times10^{-9} & 0 & 0 \\
0.1236 & -0.4034 & 0.112 & -0.1641 & -0.0138 & 0.026 & 0.0922 & 3.0039\times10^{-9} & 0 & 0 \\
-0.1843 & 0.0144 & 0.3519 & 0.4641 & -0.1343 & 0.0348 & -0.0059 & 3.0039\times10^{-9} & 0 & 0 \\
0.8598 & -0.0755 & -0.1163 & -0.1167 & -0.0013 & 0.1164 & -0.102 & 3.0039\times10^{-9} & 0 & 0 \\
0.9698 & 0.548 & 0.3162 & 0.0729 & 0.2414 & -0.004 & 0.0493 & 3.0039\times10^{-9} & 0 & 0 \\
1.0068 & 0.225 & -0.2479 & -0.0454 & -0.1471 & -0.1015 & 0.0139 & 3.0039\times10^{-9} & 0 & 0
\end{pmatrix}$$

$G := D_c^{-1}\cdot(Ds_c\cdot V)\cdot\Lambda^T$

Matrix G represents vector positions for COLUMNS of M

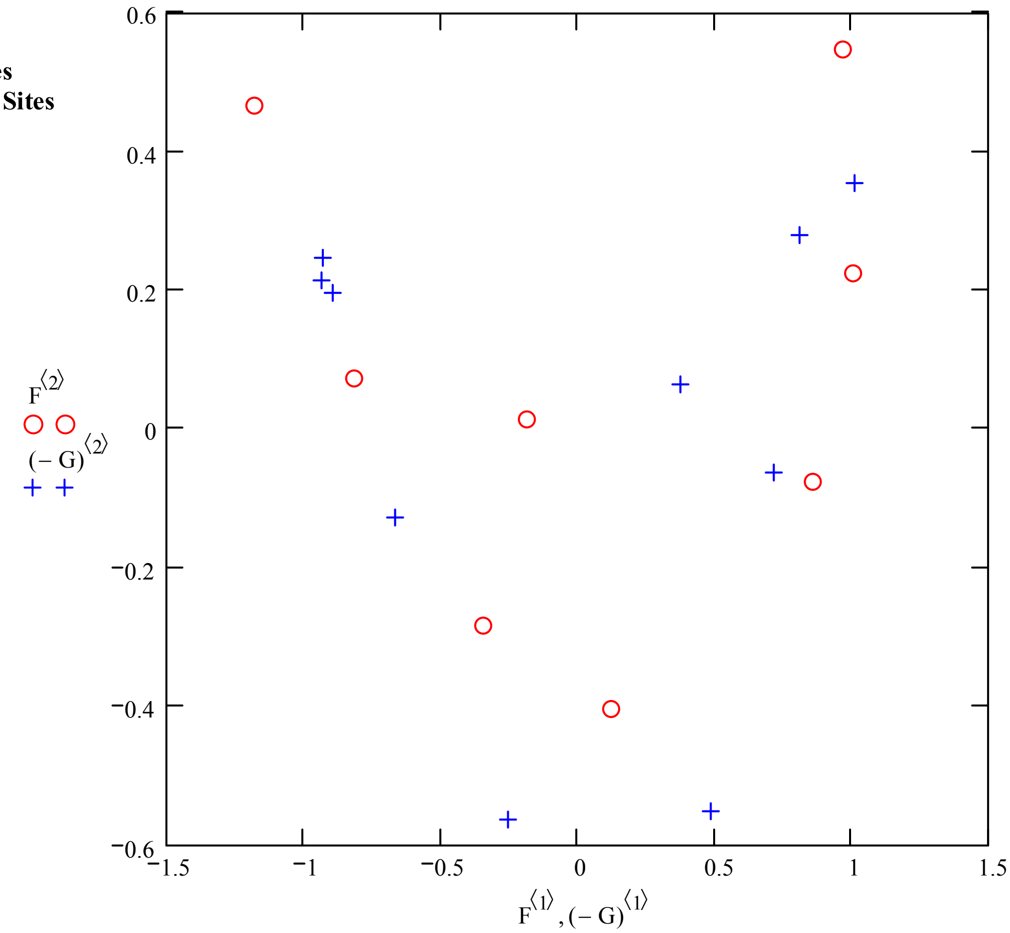
$$G = \begin{pmatrix}
0.9342 & -0.2149 & 0.0544 & -0.2828 & -0.0201 & -0.0724 & 0.087 & 1.6213\times10^{-9} \\
0.929 & -0.2476 & 0.0093 & -0.305 & -0.0675 & 0.0436 & -0.054 & 4.3292\times10^{-9} \\
0.8932 & -0.1969 & -0.3821 & 0.4257 & 0.086 & 0.0274 & 0.0716 & 3.6435\times10^{-9} \\
0.6653 & 0.127 & -0.0438 & 0.1706 & 0.0002 & 0.0168 & -0.1251 & 1.6225\times10^{-9} \\
0.2533 & 0.5628 & 0.4068 & 0.1064 & -0.0766 & -0.0338 & 0.015 & 3.6818\times10^{-9} \\
-0.4871 & 0.5508 & -0.4085 & -0.1886 & -0.0338 & 0.0874 & 0.0547 & 2.5121\times10^{-9} \\
-0.3761 & -0.065 & 0.3788 & 0.0172 & 0.1757 & 0.0365 & 0.0263 & 2.7269\times10^{-9} \\
-0.7176 & 0.0623 & -0.2843 & -0.1043 & 0.1105 & -0.1382 & -0.0487 & 3.4367\times10^{-9} \\
-1.012 & -0.3553 & 0.0348 & -0.0818 & -0.0216 & 0.0836 & -0.0298 & 2.8699\times10^{-9} \\
-0.8113 & -0.2803 & 0.0289 & 0.2259 & -0.1967 & -0.0421 & 0.0245 & 2.6995\times10^{-9}
\end{pmatrix}$$

Biplot of Rows and Columns:

F (0) = ROWS of M = Trees
G (+) = COLUMNS of M = Sites

Note: values of G are plotted as -G for comparison with R's output below.

As in PCA directions of eigenvectors are not unique.



Prototype in R:

```
#CORRESPONDENCE ANALYSIS
M=read.table("c:/MultivariateDATA/T12-10.DAT")
M
attach(M)
P=prop.table(M) # CREATES CORRESPONDENCE MATRIX P
P
```

```
> P
           V1          V2         V3         V4          V5         V6
1 0.030821918 0.027397260 0.01027397 0.01712329 0.020547945 0.00000000
2 0.027397260 0.030821918 0.02739726 0.02397260 0.000000000 0.00000000
3 0.017123288 0.013698630 0.03082192 0.03082192 0.023972603 0.02397260
4 0.010273973 0.013698630 0.00000000 0.02054795 0.030821918 0.02739726
5 0.006849315 0.006849315 0.01369863 0.01712329 0.020547945 0.00000000
6 0.000000000 0.000000000 0.00000000 0.00000000 0.006849315 0.02397260
7 0.000000000 0.000000000 0.00000000 0.00000000 0.000000000 0.00000000
8 0.000000000 0.000000000 0.00000000 0.00000000 0.000000000 0.01712329
          V7         V8          V9         V10
1 0.01712329 0.00000000 0.000000000 0.000000000
2 0.00000000 0.00000000 0.000000000 0.000000000
3 0.01369863 0.02054795 0.000000000 0.006849315
4 0.02397260 0.02054795 0.013698630 0.010273973
5 0.01712329 0.00000000 0.006849315 0.017123288
6 0.02054795 0.02054795 0.023972603 0.020547945
7 0.02397260 0.01369863 0.020547945 0.017123288
8 0.01369863 0.02739726 0.027397260 0.030821918
```

^ correspondence matrix P

```
library(ca)  # INSTALL LIBRARY {ca} from CRAN
?ca
CA <- ca(P)
summary(CA)
```

inertias >

mass = row and column sums of correspondence matrix P X 1000

(k=1,k=2) = values of matrix F & G for principal axes 1 & 2 are given for each row & column.

Other values in this table have not yet been identified.

```
> summary(CA)
Principal inertias (eigenvalues):

 dim    value      %   cum%   scree plot
 1      0.536732  69.7  69.7  *************************
 2      0.096182  12.5  82.2  ****
 3      0.072097   9.4  91.6  ***
 4      0.045547   5.9  97.5  **
 5      0.011066   1.4  98.9
 6      0.004539   0.6  99.5
 7      0.003882   0.5 100.0
        -------- -----
 Total: 0.770046 100.0
Rows:
    name   mass  qlt  inr     k=1  cor ctr    k=2  cor ctr
1 |    1 |  123  752  142 |  -815 746 152 |   73    6   7 |
2 |    2 |  110  930  246 | -1179 804 284 |  467  126 249 |
3 |    3 |  182  634   74 |  -344 378  40 | -283  256 151 |
4 |    4 |  171  785   50 |   124  67   5 | -403  717 290 |
5 |    5 |  106   87   54 |  -184  86   7 |   14    1   0 |
6 |    6 |  116  936  120 |   860 929 160 |  -75    7   7 |
7 |    7 |   75  882  138 |   970 669 132 |  548  213 235 |
8 |    8 |  116  918  175 |  1007 874 220 |  225   44  61 |
Columns:
     name   mass  qlt  inr    k=1  cor ctr    k=2  cor ctr
1  |   V1 |   92  905  122 | -934 860 150 |  215   46  44 |
2  |   V2 |   92  900  123 | -929 841 149 |  248   60  59 |
3  |   V3 |   82  711  126 | -893 678 122 |  197   33  33 |
4  |   V4 |  110  907   72 | -665 875  90 | -127   32  18 |
5  |   V5 |  103  674   75 | -253 114  12 | -563  561 338 |
6  |   V6 |   92  716   91 |  487 314  41 | -551  402 292 |
7  |   V7 |  130  452   54 |  376 439  34 |   65   13   6 |
8  |   V8 |  103  805   86 |  718 799  99 |  -62    6   4 |
9  |   V9 |   92  986  140 | 1012 878 176 |  355  108 121 |
10 |  V10 |  103  888  111 |  811 793 126 |  280   95  84 |
```

```
?plot.ca
plot(CA, arrows = c(FALSE, TRUE))
```