Statistics for the Sciences

Correspondence Analysis

Xuemao Zhang
East Stroudsburg University

January 18, 2025

Outline

- Idea of CA
- Example
- Lab

Idea of CA

- Correspondence analysis (CA) is a statistical visualization method for picturing the associations between the levels of a two-way contingency table.
  - CA is for the relationship between **two categorical variables**.
  - The term `correspondance` denotes a `system of associations` between the levels of the two sets.
- CA is used to check in a two-way contingency table whether certain levels of one categorical variable are associated with some levels of another.
- CA is a geometric technique for displaying the rows and columns of a two-way contingency table as points in a low-dimensional space, such that the positions of the row and column points are consistent with their associations in the table.

Idea of CA

- The form of a contingency table for two categorical variables

Two-Way Contingency Table

Row	Column 1	2	...	c	Marginals
1	O_{11}	O_{12}	...	O_{1c}	$R_{1\cdot}$
2	O_{21}	O_{22}	...	O_{2c}	$R_{2\cdot}$
.	.	.		.	.
.	.	.		.	.
.	.	.		.	.
r	O_{r1}	O_{r2}	...	O_{rc}	$R_{r\cdot}$
Marginals	$C_{\cdot 1}$	$C_{\cdot 2}$	...	$C_{\cdot c}$	n

Idea of CA

- Divide each cell count by the total number of observations n, we get the matrix of joint cell probabilities.
  - In the following table, $p_{ij} = Oij/n, i = 1, \ldots, r, j = 1, \ldots, n$.
  - They are actually estimates the cell probabilities.

Two-Way Table of Proportions

	Column				Marginals
row	1	2	...	c	
1	p_{11}	p_{12}	...	p_{1c}	$p_{1\cdot}$
2	p_{21}	p_{22}	...	p_{2c}	$p_{2\cdot}$
.	.	.		.	.
.	.	.		.	.
.	.	.		.	.
r	p_{r1}	p_{r2}	...	p_{rc}	$p_{r\cdot}$
Marginals	$p_{\cdot 1}$	$p_{\cdot 2}$	...	$p_{\cdot c}$	1

Idea of CA

- Consider a 3×3 contingency table

```
##      Col1 Col2 Col3
## Row1    3    2    4
## Row2    3    6    6
## Row3   10    5    9
```

- The matrix **P** of joint cell probabilities

```
## [1] "matrix of joint cell probabilities P="
```

```
##           Col1       Col2       Col3
## Row1 0.0625000 0.04166667 0.08333333
## Row2 0.0625000 0.12500000 0.12500000
## Row3 0.2083333 0.10416667 0.18750000
```

Idea of CA

- Calculate the row margin **r**
  - called row masses

```
##          [,1]
## Row1 0.1875
## Row2 0.3125
## Row3 0.5000
```

- Calculate the column margin **c**
  - called column masses

```
##              [,1]
## Col1 0.3333333
## Col2 0.2708333
## Col3 0.3958333
```

Idea of CA

```
##              Col1       Col2       Col3 Rtotal
## Row1   0.0625000 0.04166667 0.08333333 0.1875
## Row2   0.0625000 0.12500000 0.12500000 0.3125
## Row3   0.2083333 0.10416667 0.18750000 0.5000
## Ctotal 0.3333333 0.27083333 0.39583333 1.0000
```

Idea of CA

- Put the row masses in a diagonal matrix $\mathbf{D_r}$, $p_{i\cdot}, i = 1, 2, 3$

```
##        [,1]   [,2] [,3]
## [1,] 0.1875 0.0000  0.0
## [2,] 0.0000 0.3125  0.0
## [3,] 0.0000 0.0000  0.5
```

- Put the column masses in a diagonal matrix $\mathbf{D_c}$, $p_{\cdot j}, j = 1, 2, 3$

```
##           [,1]      [,2]      [,3]
## [1,] 0.3333333 0.0000000 0.0000000
## [2,] 0.0000000 0.2708333 0.0000000
## [3,] 0.0000000 0.0000000 0.3958333
```

Idea of CA

- The matrix of row conditional probabilities
  - Each row of **P** divided by row masses

```
##           Col1      Col2      Col3
## [1,] 0.3333333 0.2222222 0.4444444
## [2,] 0.2000000 0.4000000 0.4000000
## [3,] 0.4166667 0.2083333 0.3750000
```

- The matrix of column conditional probabilities
  - Each column of **P** divided by column masses

```
##        [,1]      [,2]      [,3]
## Row1 0.1875 0.1538462 0.2105263
## Row2 0.1875 0.4615385 0.3157895
## Row3 0.6250 0.3846154 0.4736842
```

Idea of CA

- The expected proportions $\mathbf{E} = \mathbf{rc^T}$
  - the proportions we expect in the case there is no relationship between the two categorical variables
  - that is, $p_{ij} = p_{i\cdot} \times p_{\cdot j}$

```
##           Col1       Col2       Col3
## Row1 0.0625000 0.05078125 0.07421875
## Row2 0.1041667 0.08463542 0.12369792
## Row3 0.1666667 0.13541667 0.19791667
```

Idea of CA

- The residuals $\mathbf{R} = \mathbf{P} - \mathbf{E}$ quantify the difference between the observed data and the data we would expect under the assumption that there is no relationship between the row and column categories of the table

```
##             Col1         Col2         Col3
## Row1  0.00000000 -0.009114583  0.009114583
## Row2 -0.04166667  0.040364583  0.001302083
## Row3  0.04166667 -0.031250000 -0.010416667
```

Idea of CA

- Dividing the residuals by the expected values, we obtain the matrix $\mathbf{I_r}$ of indexed residuals
  - the further the value from the table, the larger the observed proportion relative to the expected proportion

```
##       Col1       Col2        Col3
## Row1  0.00 -0.1794872  0.12280702
## Row2 -0.40  0.4769231  0.01052632
## Row3  0.25 -0.2307692 -0.05263158
```

Idea of CA

- Finally, calculate the matrix of standardized residuals $\mathbf{Z}$, obtained by multiplying each element of $\mathbf{I_r}$ by each element of $\sqrt{\mathbf{E}}$

```
##             [,1]        [,2]        [,3]
## [1,]  0.0000000 -0.04044689  0.03345646
## [2,] -0.1290994  0.13874726  0.00370218
## [3,]  0.1020621 -0.08492078 -0.02341465
```

Idea of CA

- The standardized residuals matrix $\mathbf{Z}$ is used in correspondence analysis to visualize and interpret the relationships between row and column categories in a contingency table.
- Then a mathematical tool Singular Value Decomposition (SVD) will be used to Decompose $\mathbf{Z}$
  - The squared singular values correspond to the eigenvalues of $\mathbf{Z}$
  - In the reduced dimensional space, the left singular vectors represent the row coordinates, and
  - the right singular vectors represent the column coordinates.
  - thenn alculate the row and column scores
- The eigenvalues represent the amount of variance explained by each dimension in the reduced space.
  - Larger eigenvalues indicate dimensions that explain more variance.
- The row and column scores provide the coordinates for plotting the rows and columns in a biplot. These scores show the relationship between row and column categories.

One Application of CA

- The application of CA in ecology is radically different.
- For example, Abundance data
  - Abundance data are gathered from various physical samples (e.g., areas or volumes) where different species are identified and counted.
  - This results in a matrix is not a statistical contingency table, as the counts are not derived from random sampling.
- Since ecological samples (e.g., fixed-size quadrats in botany or fixed volumes in marine research) are often uniform in size, an important question arises:
  - Should abundances be profiled relative to the total abundance within each sample?
- CA can help us visualize relationships between species and environmental factors.

Example

- (`lemminvert2.csv`): Lemmens et al. (2015) did a detailed study of various biotic communities in artificial ponds in Belgium. They sampled 28 ponds that represented different types of management, a combination of fish farming strategies (no fish, farming young fish, low intensity management, no management), and drainage frequencies ($>$ 10 years ago, occasional, annual). They also quantified taxon abundances for fish, zooplankton, and macro-invertebrates (different families and species within some groups) and covers of submerged, floating, and emergent vegetation.
  - The macroinvertebrate dataset only included 23 ponds and we will use these data to illustrate CA by examining the ordination of the macroinvertebrate community (abundances of families).

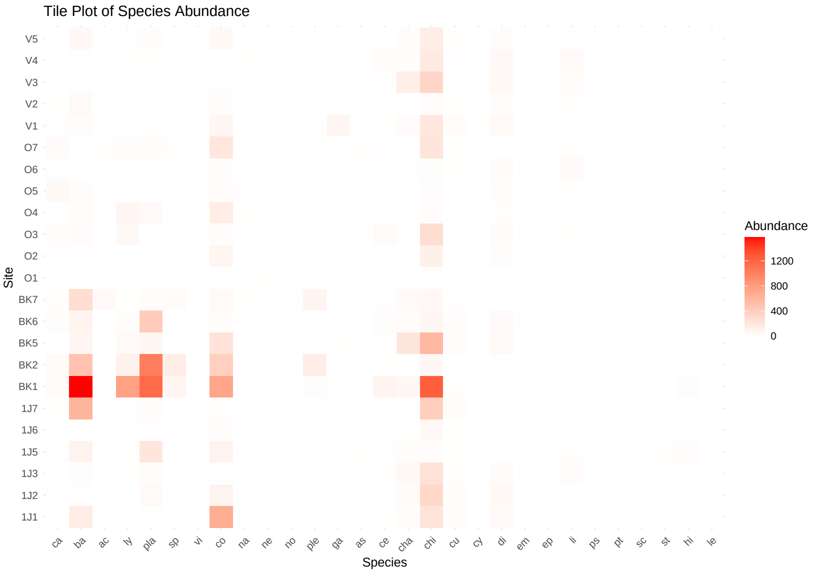
Example

- Remove the first 2nd and 3rd columns, and check some statistics

```
##    site ca   ba ac  ly  pla  sp vi  co na ne no ple ga as ce cha  chi cu cy di
## 1   1J1  0  158  0   3    5   0  0 671  0  0  0   0  0  0  4  23  222 19  0 40
## 2   1J2  6   13  0   0   50   0  0  95  0  0  0   0  0  0  7  37  322 17  0 52
## 3   1J3  0   20  0   3   16   0  0   8  0  0  0   0  1  0  9  60  240 12  0 30
## 4   1J5  5   98  0  10  206   2  0  98  0  0  0   0  2  4  5  20   26  4  0  2
## 5   1J6  0    3  0   9   14   0  0  23  0  0  0   0  1  0  0   3   60  3  0  5
## 6   1J7  3  605  0   7   19   0  0   9  0  0  0   0  0  0  7   5  394 19  0  0
## 7   BK1 31 1571  0 763 1164  80  0 730  0  0  3  20  4  0 88  68 1236 12  0  4
## 8   BK2 36  494  0 109 1035 147  0 389  2  0  1 141  0  0  3   0   62  0  0  6
## 9   BK5  0   73  1  46   64   0  0 239  2  0  0   5  4  0  6 204  569 24  0 47
## 10  BK6 16   90  0  15  421   0  0  25  0  0  0   0  1  1 20  22   69 21  0 38
## 11  BK7 12  282 42   4   25  36  0  46  3  0  7  91  1  1  6  38   54  8  0  0
## 12   O1  0    4  0   0    0   0  0   2  0  3  0   0  0  2  0   0   12  0  0  0
## 13   O2 13   13  0   0    1   0  0  83  0  0  1   0  0  0  8   2  124  0  0 19
## 14   O3 15   31  0  51    6   1  0  24  0  0  0   0  0  0 34   8  274  5  0 29
## 15   O4  0   28  2  72   39   2  0 150  3  0  0   1  0  0  0   1   21  1  0  9
## 16   O5 53   18  0   6    2   0  0  22  5  0  6   0  0  4  1   0   16  0  0 19
## 17   O6  6    3  0  10    0   0  0  18  0  6  0   0  0  0  0   5   19  4  0 35
## 18   O7 30    0 10  20   18   9  1 197  0  2  0   0  0  9 14   0  215  3  0  2
## 19   V1  6   16  0   5    9   0  0  81  0  0  0   0 74  1  3  29  196 26  3 39
## 20   V2  3   39  0   0    0   0  0  15  0  0  0   0  0  1  3   2   22  4  0 26
## 21   V3  0    6  0   2    0   1  0   6  0  0  0   0  1  0 14 131  342  0  0 57
## 22   V4  0    6  0   4    3   0  0   2  4  0  0   0  0  0 27  16  176 14  0 56
## 23   V5  0   58  0   7   18   2  0  51  1  0  0   0  0  0  0  18  147  9  0 21
##    em ep li ps pt sc st hi le
## 1   0  0  8  0  0  0  1  0  1
## 2   0  0  1  0  2  0  0  0  0
## 3   0  0 25  0  0  0  0  1  0
```

Example

- Get a tile plot to check relative abundance of each species within each site



Example

- Regard variables `site` as row names/labels

```
##    ca   ba ac  ly  pla  sp vi  co na ne no ple ga as ce cha  chi cu cy di em ep
## 1   0  158  0   3    5   0  0 671  0  0  0   0  0  0  4  23  222 19  0 40  0  0
## 2   6   13  0   0   50   0  0  95  0  0  0   0  0  0  7  37  322 17  0 52  0  0
## 3   0   20  0   3   16   0  0   8  0  0  0   0  1  0  9  60  240 12  0 30  0  0
## 4   5   98  0  10  206   2  0  98  0  0  0   0  2  4  5  20   26  4  0  2  0  0
## 5   0    3  0   9   14   0  0  23  0  0  0   0  1  0  0   3   60  3  0  5  1  1
## 6   3  605  0   7   19   0  0   9  0  0  0   0  0  0  7   5  394 19  0  0  0  0
## 7  31 1571  0 763 1164  80  0 730  0  0  3  20  4  0 88  68 1236 12  0  4  0  0
## 8  36  494  0 109 1035 147  0 389  2  0  1 141  0  0  3   0   62  0  0  6  0  0
## 9   0   73  1  46   64   0  0 239  2  0  0   5  4  0  6 204  569 24  0 47  0  0
## 10 16   90  0  15  421   0  0  25  0  0  0   0  1  1 20  22   69 21  0 38  5  0
## 11 12  282 42   4   25  36  0  46  3  0  7  91  1  1  6  38   54  8  0  0  0  0
## 12  0    4  0   0    0   0  0   2  0  3  0   0  0  2  0   0   12  0  0  0  0  0
## 13 13   13  0   0    1   0  0  83  0  0  1   0  0  0  8   2  124  0  0 19  0  0
## 14 15   31  0  51    6   1  0  24  0  0  0   0  0  0 34   8  274  5  0 29  0  0
## 15  0   28  2  72   39   2  0 150  3  0  0   1  0  0  0   1   21  1  0  9  0  0
## 16 53   18  0   6    2   0  0  22  5  0  6   0  0  4  1   0   16  0  0 19  0  0
## 17  6    3  0  10    0   0  0  18  0  6  0   0  0  0  0   5   19  4  0 35  0  0
## 18 30    0 10  20   18   9  1 197  0  2  0   0  0  9 14   0  215  3  0  2  0  0
## 19  6   16  0   5    9   0  0  81  0  0  0   0 74  1  3  29  196 26  3 39  0  0
## 20  3   39  0   0    0   0  0  15  0  0  0   0  0  1  3   2   22  4  0 26  0  0
## 21  0    6  0   2    0   1  0   6  0  0  0   0  1  0 14 131  342  0  0 57  0  0
## 22  0    6  0   4    3   0  0   2  4  0  0   0  0  0 27  16  176 14  0 56  0  0
## 23  0   58  0   7   18   2  0  51  1  0  0   0  0  0  0  18  147  9  0 21  0  0
##    li ps pt sc st hi le
## 1   8  0  0  0  1  0  1
## 2   1  0  2  0  0  0  0
## 3  25  0  0  0  0  1  0
```

Example

```
##
##  Principal inertias (eigenvalues):
##            1        2        3        4        5        6        7
## Value      0.364603 0.156342 0.147145 0.135003 0.124771 0.098519 0.080732
## Percentage 27.91%   11.97%   11.26%   10.33%   9.55%    7.54%    6.18%
##            8        9        10       11       12       13      14
## Value      0.046041 0.040712 0.031388 0.026204 0.018642 0.01379 0.006885
## Percentage 3.52%    3.12%    2.4%     2.01%    1.43%    1.06%   0.53%
##            15       16       17       18       19       20       21
## Value      0.005144 0.004454 0.003261 0.001128 0.000886 0.000536 0.000155
## Percentage 0.39%    0.34%    0.25%    0.09%    0.07%    0.04%    0.01%
##            22
## Value      9.8e-05
## Percentage 0.01%
##
##
##  Rows:
##               1J1       1J2       1J3       1J5       1J6       1J7       BK1
## Mass    0.061626  0.032120  0.022676  0.028172  0.006723  0.056984  0.309519
## ChiDist 1.201693  0.933132  1.171392  1.148363  1.722272  1.105904  0.468482
## Inertia 0.088992  0.027968  0.031116  0.037152  0.019942  0.069693  0.067932
## Dim. 1  0.396955  1.179828  1.641302 -0.779788  0.789258 -0.044049 -0.485453
## Dim. 2  0.106377 -0.129481 -0.215181  1.095410 -0.109035 -1.811533 -0.508792
##               BK2       BK5       BK6       BK7       O1       O2        O3
## Mass     0.129975  0.068563  0.041351  0.035268 0.001334 0.014193  0.025878
## ChiDist  1.001979  0.895062  1.218067  1.798732 5.509043 0.966582  1.024400
## Inertia  0.130490  0.054928  0.061352  0.114109 0.040483 0.013260  0.027156
## Dim. 1  -1.426088  1.082130 -0.791462 -0.897415 1.532024 0.998728  1.050517
## Dim. 2   0.743469 -0.592846  1.540257 -0.609251 3.187125 0.369797 -0.398976
##                O4       O5       O6       O7       V1       V2        V3
## Mass     0.017661 0.009124 0.007683 0.028972 0.026891 0.006776  0.031480
## ChiDist  1.149777 3.381547 3.207381 1.112261 2.268350 1.423040  1.483498
## Inertia  0.023347 0.104330 0.079040 0.035842 0.138368 0.013722  0.069280
## Dim. 1  -0.329851 0.691269 2.177059 0.638978 1.709381 1.016049  1.978599
## Dim. 2   0.299910 5.343375 4.392634 0.739371 0.633688 1.111008 -0.339015
##               V4        V5
## Mass    0.018781  0.018248
## ChiDist 1.645441  0.623136
## Inertia 0.050850  0.007086
## Dim. 1  1.927954  0.817559
```

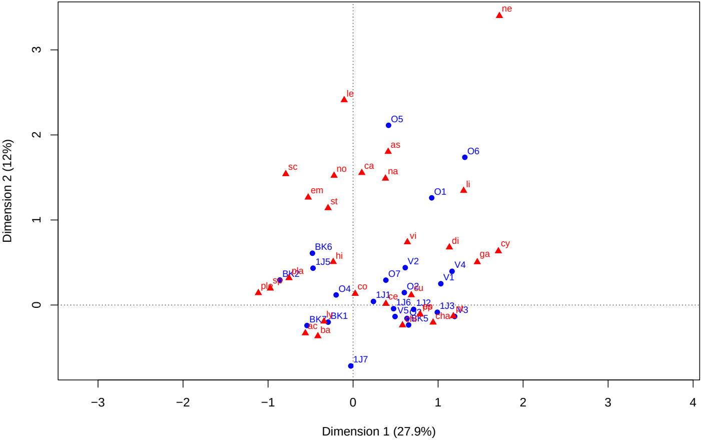
Example

- Proportion of variation explained

```
##
## Principal inertias (eigenvalues):
##
##  dim    value      %   cum%   scree plot
##  1      0.364603  27.9  27.9  *******
##  2      0.156342  12.0  39.9  ***
##  3      0.147145  11.3  51.1  ***
##  4      0.135003  10.3  61.5  ***
##  5      0.124771   9.6  71.0  **
##  6      0.098519   7.5  78.6  **
##  7      0.080732   6.2  84.7  **
##  8      0.046041   3.5  88.3  *
##  9      0.040712   3.1  91.4  *
##  10     0.031388   2.4  93.8  *
##  11     0.026204   2.0  95.8  *
##  12     0.018642   1.4  97.2
##  13     0.013790   1.1  98.3
##  14     0.006885   0.5  98.8
##  15     0.005144   0.4  99.2
##  16     0.004454   0.3  99.5
##  17     0.003261   0.2  99.8
##  18     0.001128   0.1  99.9
##  19     0.000886   0.1  99.9
##  20     0.000536   0.0 100.0
##  21     0.000155   0.0 100.0
##  22     9.8e-050   0.0 100.0
##         -------- -----
##  Total: 1.306439 100.0
```

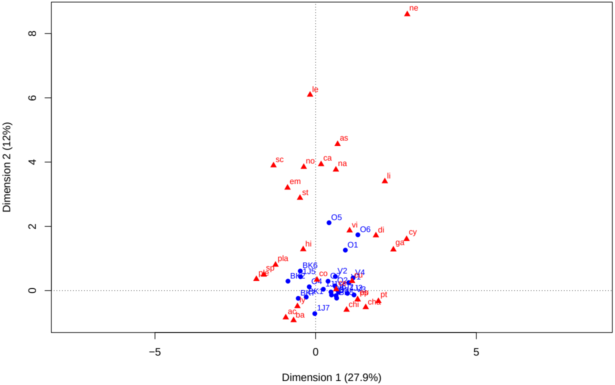
Example

- Visualization of relationships among categories
- row/col points `near` each other are positively associated
- The red points are for columns



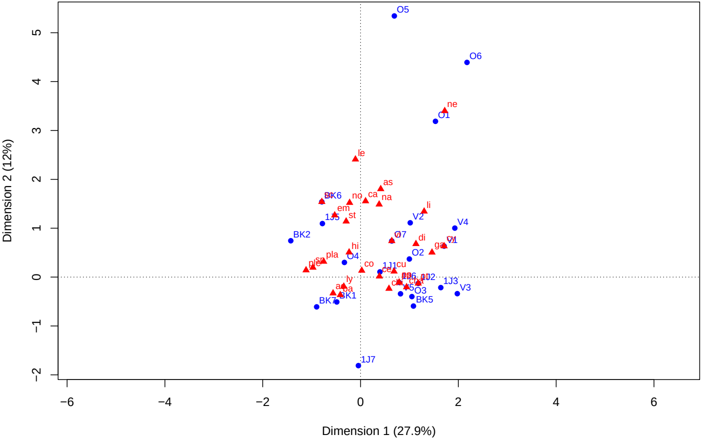
Example

- In standard CA, the distances between row and column coordinates are exaggerated, and there isn't a straightforward interpretation of relationships between row and column categories.
- To interpret the relationships between row coordinates, we use **row principal normalization**
  - Points that are closer together in the plot indicate rows (sites) that have similar species compositions.



Example

- To interpret the relationships between column coordinates, we use **column principal normalization**
  - Points that are closer together in the plot indicate columns (species) that have similar distributions across the rows (sites).



Lab

- CA is for a contingency table. In SPSS so we need to reshape the data from

	var_1	var_2	$\cdots$	var_{29}
row1	x_{11}	x_{12}	$\cdots$	$x_{1,29}$
row2	x_{21}	x_{22}	$\cdots$	$x_{2,29}$
$\vdots$	$\vdots$	$\ldots$	$\vdots$	$\vdots$
row23	$x_{23,1}$	$x_{23,2}$	$\cdots$	$x_{23,29}$

Lab

- to

Row	Column	Count
row1	var_1	x_{11}
$\vdots$	$\vdots$	$\vdots$
row1	var_{29}	$x_{1,29}$
row2	var_1	x_{21}
$\vdots$	$\vdots$	$\vdots$
row2	var_{29}	$x_{2,29}$
$\vdots$	$\vdots$	$\vdots$
row23	var_1	$x_{23,1}$
$\vdots$	$\vdots$	$\vdots$
row23	var_{29}	$x_{23,29}$

Lab

- We import the data with measures of all numerical variables as `Scale`

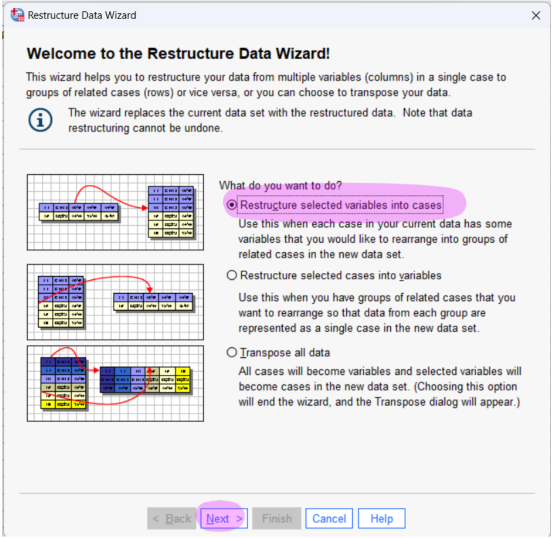
File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help

Search application

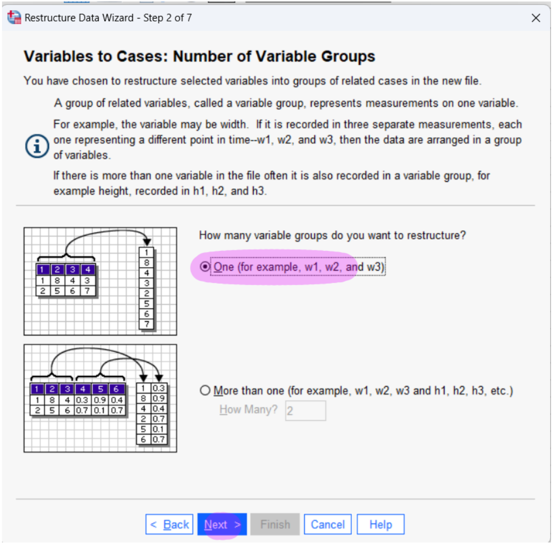
	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	site	String	3	0		None	None	7	Left	Nominal	Input
2	manag	String	2	0		None	None	4	Left	Nominal	Input
3	managsymb	Numeric	1	0		None	None	8	Right	Nominal	Input
4	caen	Numeric	2	0		None	None	8	Right	Scale	Input
5	baet	Numeric	4	0		None	None	8	Right	Scale	Input
6	acrol	Numeric	2	0		None	None	8	Right	Scale	Input
7	lymn	Numeric	3	0		None	None	8	Right	Scale	Input
8	plan	Numeric	4	0		None	None	8	Right	Scale	Input
9	sphaer	Numeric	3	0		None	None	8	Right	Scale	Input
10	vivip	Numeric	1	0		None	None	8	Right	Scale	Input
11	corix	Numeric	3	0		None	None	8	Right	Scale	Input
12	nauc	Numeric	1	0		None	None	8	Right	Scale	Input
13	nepid	Numeric	1	0		None	None	8	Right	Scale	Input
14	noto	Numeric	1	0		None	None	8	Right	Scale	Input
15	pleid	Numeric	3	0		None	None	8	Right	Scale	Input
16	gamm	Numeric	2	0		None	None	8	Right	Scale	Input
17	asell	Numeric	1	0		None	None	8	Right	Scale	Input
18	cerat	Numeric	2	0		None	None	8	Right	Scale	Input
19	chaob	Numeric	3	0		None	None	8	Right	Scale	Input
20	chiro	Numeric	4	0		None	None	8	Right	Scale	Input
21	culic	Numeric	2	0		None	None	8	Right	Scale	Input
22	cylind	Numeric	1	0		None	None	8	Right	Scale	Input
23	dixid	Numeric	2	0		None	None	8	Right	Scale	Input
24	empid	Numeric	1	0		None	None	8	Right	Scale	Input
25	ephyd	Numeric	1	0		None	None	8	Right	Scale	Input
26	limon	Numeric	2	0		None	None	8	Right	Scale	Input
27	psych	Numeric	1	0		None	None	8	Right	Scale	Input
28	ptych	Numeric	1	0		None	None	8	Right	Scale	Input
29	sciom	Numeric	1	0		None	None	8	Right	Scale	Input
30	strat	Numeric	2	0		None	None	8	Right	Scale	Input

Lab

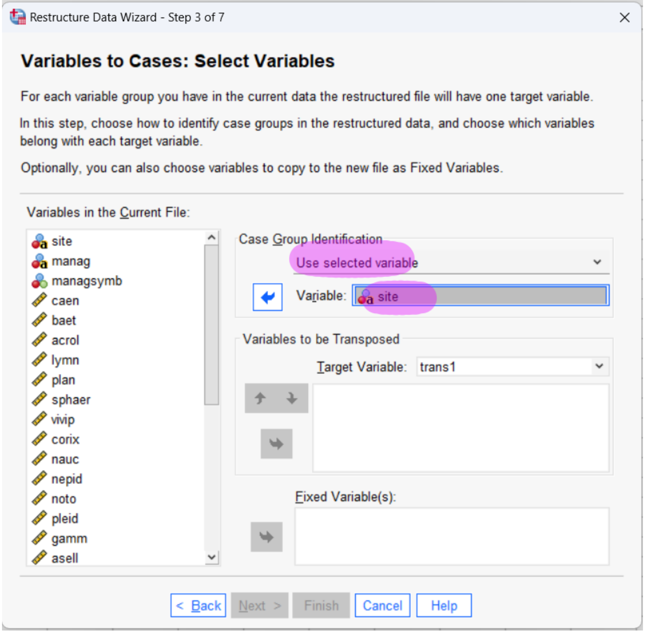
- Click on Data → Restructure...



Lab


Restructure Data Wizard - Step 2 of 7
Variables to Cases: Number of Variable Groups
You have chosen to restructure selected variables into groups of related cases in the new file.
A group of related variables, called a variable group, represents measurements on one variable.
For example, the variable may be width. If it is recorded in three separate measurements, each one representing a different point in time--w1, w2, and w3, then the data are arranged in a group of variables.
If there is more than one variable in the file often it is also recorded in a variable group, for example height, recorded in h1, h2, and h3.
How many variable groups do you want to restructure?
One (for example, w1, w2, and w3)
More than one (for example, w1, w2, w3 and h1, h2, h3, etc.)
How Many? 2
< Back
Next >
Finish
Cancel
Help

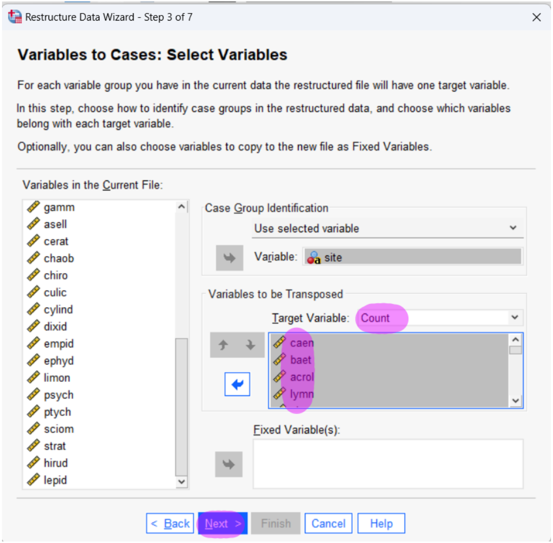
Lab



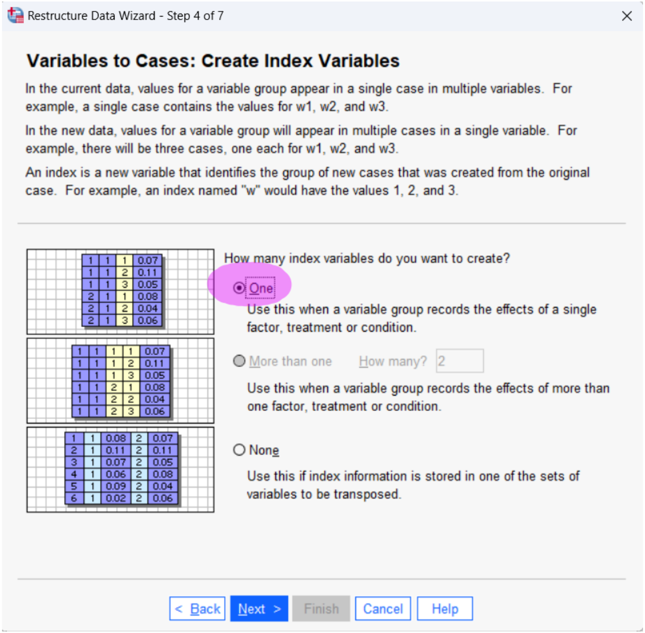
Restructure Data Wizard - Step 3 of 7
Variables to Cases: Select Variables
For each variable group you have in the current data the restructured file will have one target variable.
In this step, choose how to identify case groups in the restructured data, and choose which variables belong with each target variable.
Optionally, you can also choose variables to copy to the new file as Fixed Variables.
Variables in the Current File:
site
manag
managsymb
caen
baet
acrol
lymn
plan
sphaer
vivip
corix
nauc
nepid
noto
pleid
gamm
asell
Case Group Identification
Use selected variable
Variable:
site
Variables to be Transposed
Target Variable:
trans1
Fixed Variable(s):
< Back
Next >
Finish
Cancel
Help

Lab

- Add all numerical variables to the Variables to be Transposed box, and name the Target Variable as Count

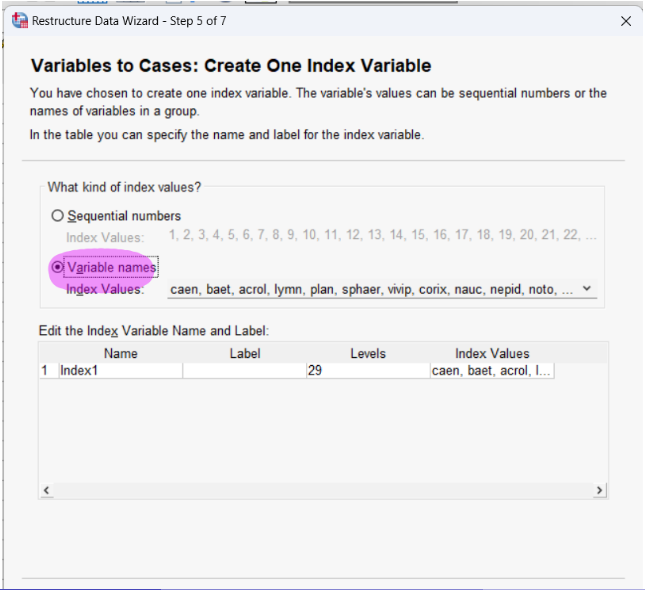


Lab


Restructure Data Wizard - Step 4 of 7
Variables to Cases: Create Index Variables
In the current data, values for a variable group appear in a single case in multiple variables. For example, a single case contains the values for w1, w2, and w3.
In the new data, values for a variable group will appear in multiple cases in a single variable. For example, there will be three cases, one each for w1, w2, and w3.
An index is a new variable that identifies the group of new cases that was created from the original case. For example, an index named "w" would have the values 1, 2, and 3.
How many index variables do you want to create?
One
Use this when a variable group records the effects of a single factor, treatment or condition.
More than one
How many?
2
Use this when a variable group records the effects of more than one factor, treatment or condition.
None
Use this if index information is stored in one of the sets of variables to be transposed.
< Back
Next >
Finish
Cancel
Help

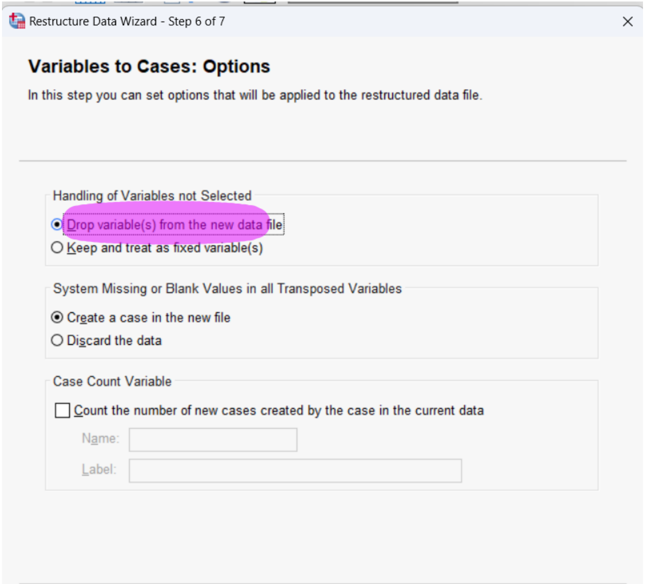
Lab

- Use variable names as index



Lab

- Drop the other two variables.



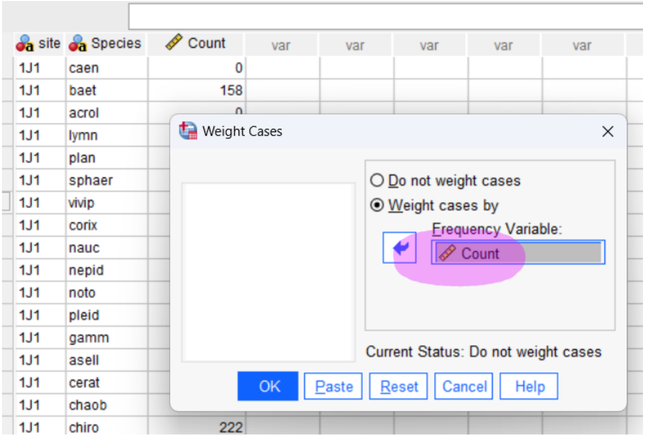
Lab

- The long data will be like this
  - Let's change the name of `Index1` to `Species`

	site	Index1	Count
1	1J1	caen	0
2	1J1	baet	158
3	1J1	acrol	0
4	1J1	lymn	3
5	1J1	plan	5
6	1J1	sphaer	0
7	1J1	vivip	0
8	1J1	corix	671
9	1J1	nauc	0
10	1J1	nepid	0
11	1J1	noto	0
12	1J1	pleid	0
13	1J1	gamm	0
14	1J1	asell	0
15	1J1	cerat	4
16	1J1	chaob	23

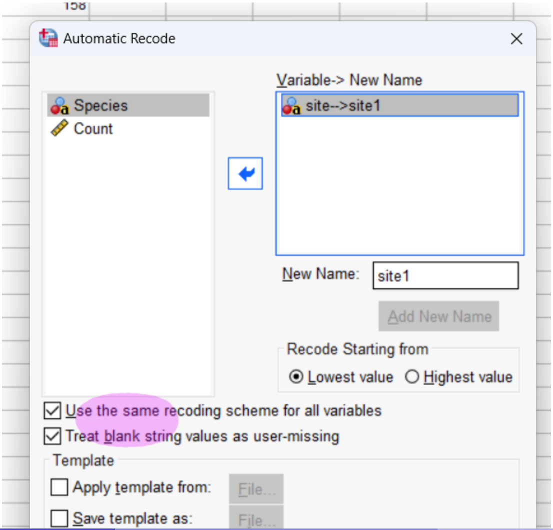
Lab

- We regard the values of `Count` as frequencies, so
  - Click on `Data` → `Weight Cases...`



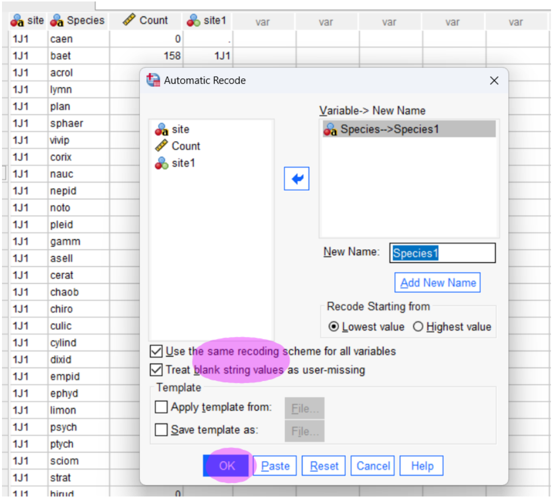
Lab

- To conduct CA in SPSS, we need the two categorical variables `Site` and `Species` numerical
  - Click on `Transform` → `Automatic Recode ...`



Lab

- Do the same thing to Species



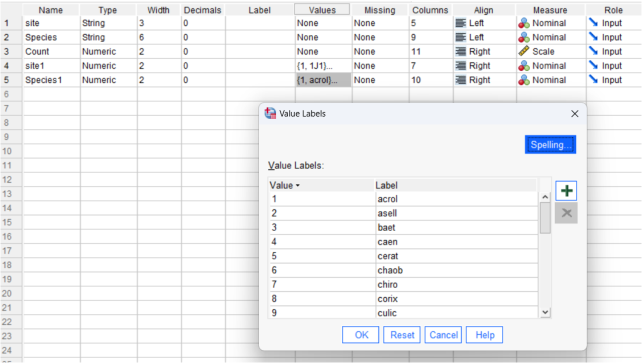
Lab

- The data now is like this

	site	Species	Count	site1	Species1	var
1	1J1	caen	0	.	.	
2	1J1	baet	158	1J1	baet	
3	1J1	acrol	0	.	.	
4	1J1	lymn	3	1J1	lymn	
5	1J1	plan	5	1J1	plan	
6	1J1	sphaer	0	.	.	
7	1J1	vivip	0	.	.	
8	1J1	corix	671	1J1	corix	
9	1J1	nauc	0	.	.	
10	1J1	nepid	0	.	.	
11	1J1	noto	0	.	.	
12	1J1	pleid	0	.	.	
13	1J1	gamm	0	.	.	
14	1J1	asell	0	.	.	
15	1J1	cerat	4	1J1	cerat	
16	1J1	chaob	23	1J1	chaob	
17	1J1	chiro	222	1J1	chiro	
18	1J1	culic	19	1J1	culic	
19	1J1	cylind	0	.	.	
20	1J1	dixid	40	1J1	dixid	
21	1J1	empid	0	.	.	
22	1J1	ephyd	0	.	.	

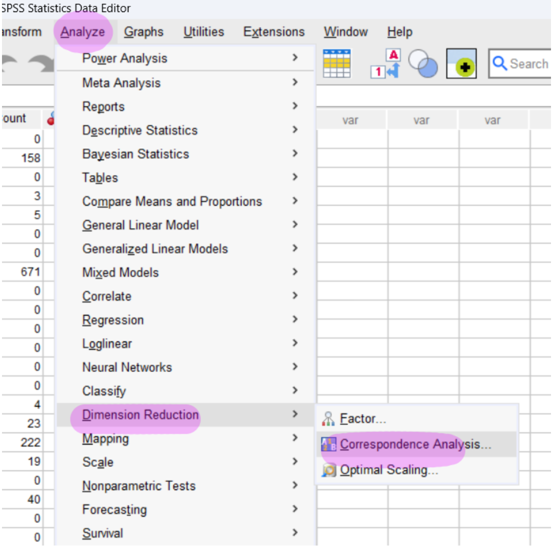
Lab

- The data now is like this



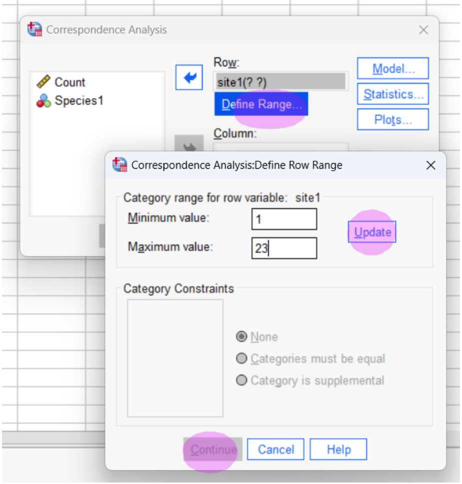
Lab

- Finally,


SPSS Statistics Data Editor
ansform
Analyze
Graphs
Utilities
Extensions
Window
Help
Search
ount
var
var
var
0
158
0
3
5
0
0
671
0
0
0
0
0
0
4
23
222
19
0
40
0
0
Power Analysis
Meta Analysis
Reports
Descriptive Statistics
Bayesian Statistics
Tables
Compare Means and Proportions
General Linear Model
Generalized Linear Models
Mixed Models
Correlate
Regression
Loglinear
Neural Networks
Classify
Dimension Reduction
Mapping
Scale
Nonparametric Tests
Forecasting
Survival
Factor...
Correspondence Analysis...
Optimal Scaling...

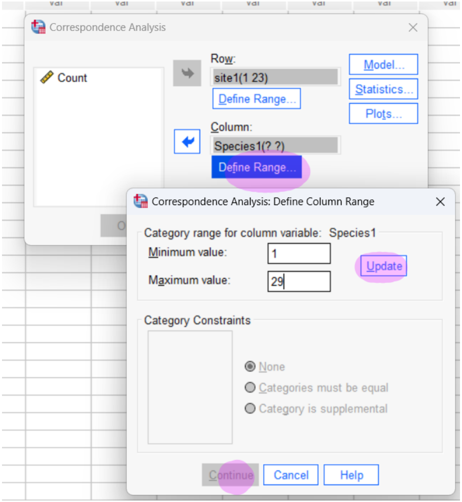
Lab

- Row variable

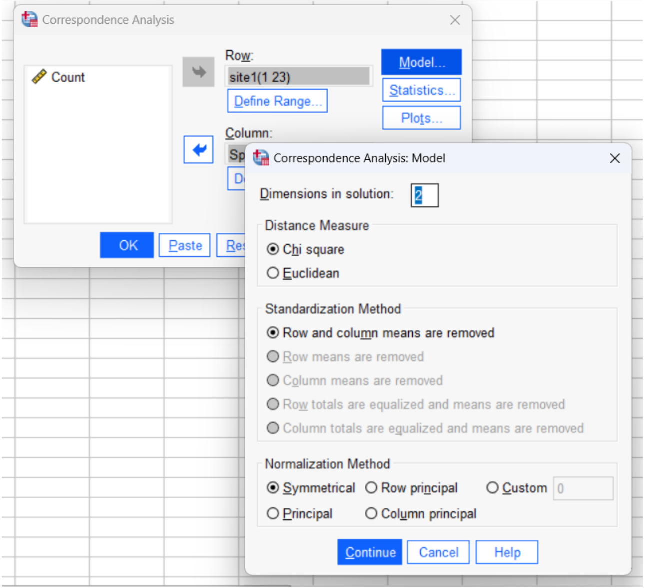


Lab

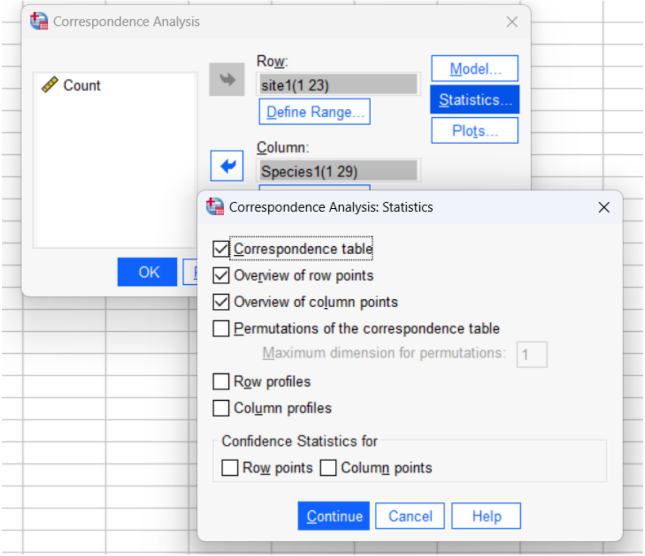
- Column variable



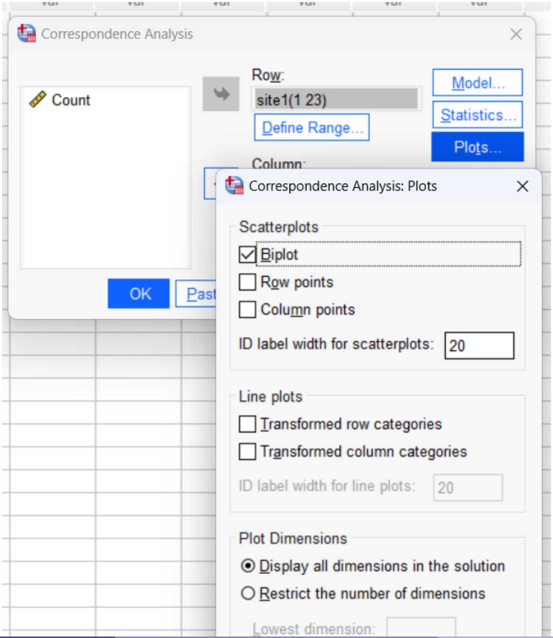
Lab


Correspondence Analysis
Count
Row:
site1(1 23)
Define Range...
Column:
Model...
Statistics...
Plots...
OK
Paste
Correspondence Analysis: Model
Dimensions in solution: 2
Distance Measure
Chi square
Euclidean
Standardization Method
Row and column means are removed
Row means are removed
Column means are removed
Row totals are equalized and means are removed
Column totals are equalized and means are removed
Normalization Method
Symmetrical
Row principal
Custom
0
Principal
Column principal
Continue
Cancel
Help

Lab

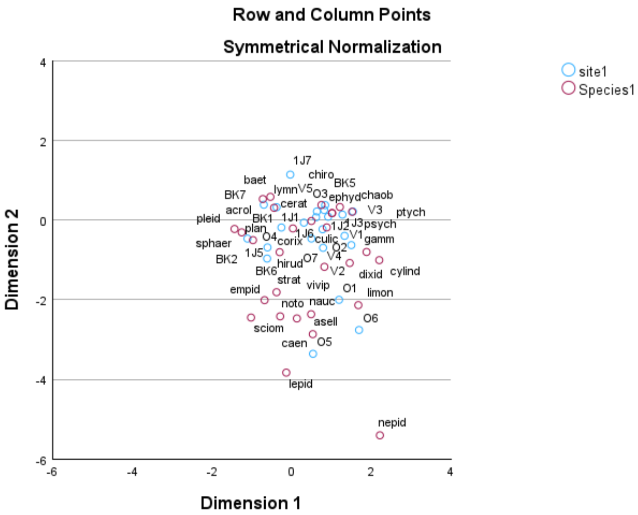

Correspondence Analysis
Count
Row:
site1(1 23)
Define Range...
Column:
Species1(1 29)
Model...
Statistics...
Plots...
OK
Correspondence Analysis: Statistics
Correspondence table
Overview of row points
Overview of column points
Permutations of the correspondence table
Maximum dimension for permutations:
1
Row profiles
Column profiles
Confidence Statistics for
Row points
Column points
Continue
Cancel
Help

Lab



Correspondence Analysis
Count
Row:
site1(1 23)
Define Range...
Model...
Statistics...
Plots...
Column:
OK
Past
Correspondence Analysis: Plots
Scatterplots
Biplot
Row points
Column points
ID label width for scatterplots:
20
Line plots
Transformed row categories
Transformed column categories
ID label width for line plots:
20
Plot Dimensions
Display all dimensions in the solution
Restrict the number of dimensions
Lowest dimension:

Lab


Row and Column Points
Symmetrical Normalization
site1
Species1
Dimension 1
Dimension 2
1J7
chiro
baet
BK5
lymn
V5
O3
ephyd
chaob
BK7
cerat
acrol
V3
ptych
pleid
BK1
1J1
1J3
psych
plan
1J2
1J6
V1
sphaer
O4
corix
culic
gamm
O2
1J5
V4
BK2
O7
hirud
BK6
V2
dixid
cylind
strat
vivip
O1
empid
limon
noto
nauc
O6
asell
sciom
caen
O5
lepid
nepid

License