Correspondence Analysis

The **Correspondence Analysis** procedure creates a map of the rows and columns in a two-way contingency table for the purpose of providing insights into the relationships amongst the categories of the row and/or column variables. Often, no more than two or three dimensions are needed to display most of the variability or "inertia" in the table. An important part of the output is a correspondence map on which the distance between two categories is a measure of their similarity.

Sample StatFolio: *correspondence.sgp*

Sample Data:

The file *funding.sgd* contains data that describe the level of funding awarded to 796 scientific researchers (from Greenacre, 2007). The researchers are divided into 10 scientific disciplines, which form the rows of the table shown below:

	A	*B*	*C*	*D*	*E*
Geology	3	19	39	14	19
Biochemistry	1	2	13	1	12
Chemistry	6	25	49	21	29
Zoology	3	15	41	35	26
Physics	10	22	47	9	26
Engineering	3	11	25	15	34
Microbiology	1	6	14	5	11
Botany	0	12	34	17	23
Statistics	2	5	11	4	7
Mathematics	2	11	37	8	20

The columns identify the level of funding awarded to each researcher, from A (most highly funded) to D (least funded), with E representing no funding at all. In addition to the rows and columns that will be used to generate the map, additional *supplementary* rows and columns may be added to the table for plotting purposes only. For example, the table below adds two supplementary rows and one supplementary column:

	A	*B*	*C*	*D*	*E*	*Y*
Geology	3	19	39	14	19	0
Biochemistry	1	2	13	1	12	1
Chemistry	6	25	49	21	29	0
Zoology	3	15	41	35	26	0
Physics	10	22	47	9	26	1
Engineering	3	11	25	15	34	1
Microbiology	1	6	14	5	11	1
Botany	0	12	34	17	23	1
Statistics	2	5	11	4	7	0
Mathematics	2	11	37	8	20	2
Museums	4	12	11	9	7	
Math Sciences	4	16	48	12	27	

© 2009 by StatPoint Technologies, Inc.

Data Input

The procedure is designed to handle two types of data:

1. Data that has been tabulated into a two-way table such as that shown above.

2. Data that has not yet been tabulated.

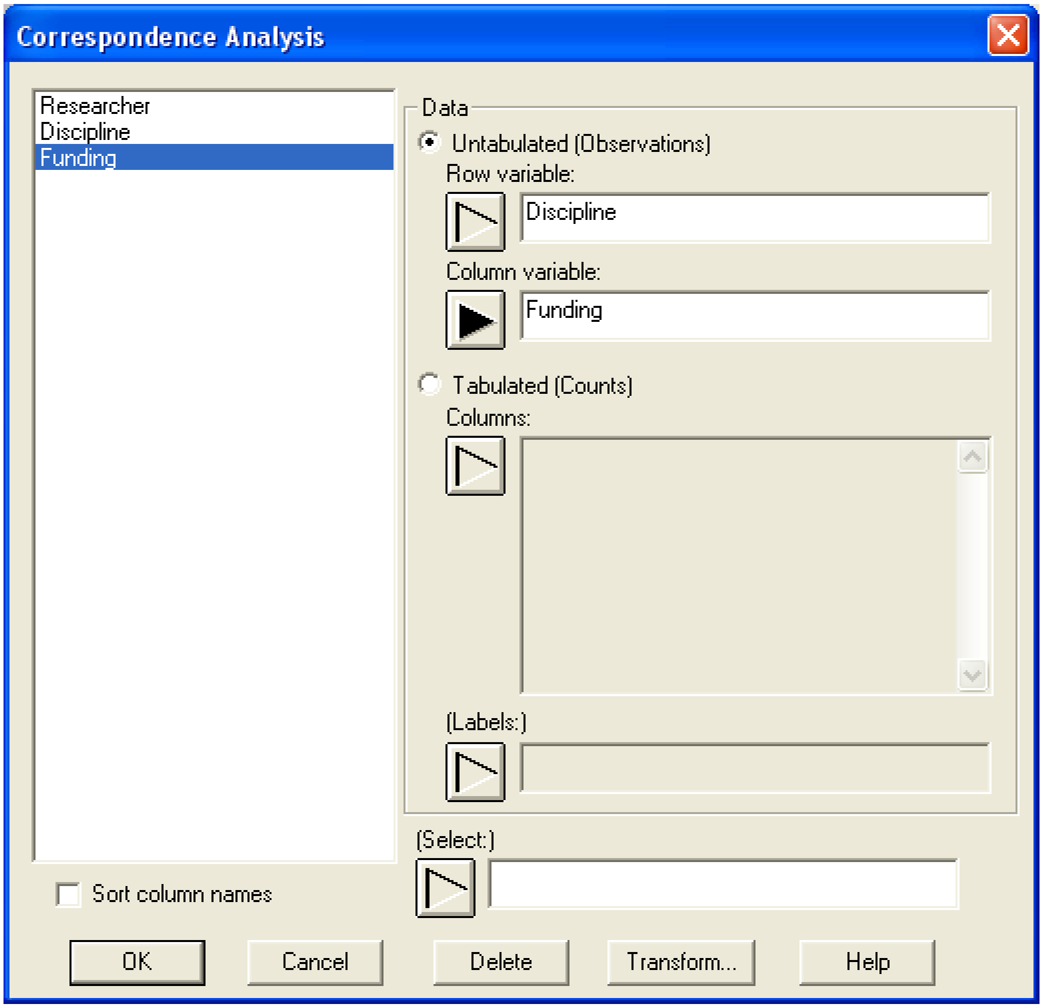
For example, the data file or the above example might contain 792 rows, one for each researcher:

<untitled>

	Researcher	Discipline	Funding
1	Wanda Muller	Chemistry	B
2	Lucy Baskett	Physics	A
3	Tiara Stoddard	Zoology	Y
4	Grant Demuth	Statistics	A
5	Jose Fisher	Biochemistry	C
6	Brooke Moen	Physics	C
7	Phyliss Anderson	Mathematics	B
8	Briscoe Basinger	Zoology	A
9	Anthony Gonzales	Zoology	D
10	Moreen Keilbach	Engineering	C
11	Morton Tilton	Physics	B
12	Yong Li	Chemistry	B
13	Quintin Cox	Statistics	A

In that case, select the *Untabulated (Observations)* radio button on the data input dialog box and specify the columns of the data file as shown below:

© 2009 by StatPoint Technologies, Inc.

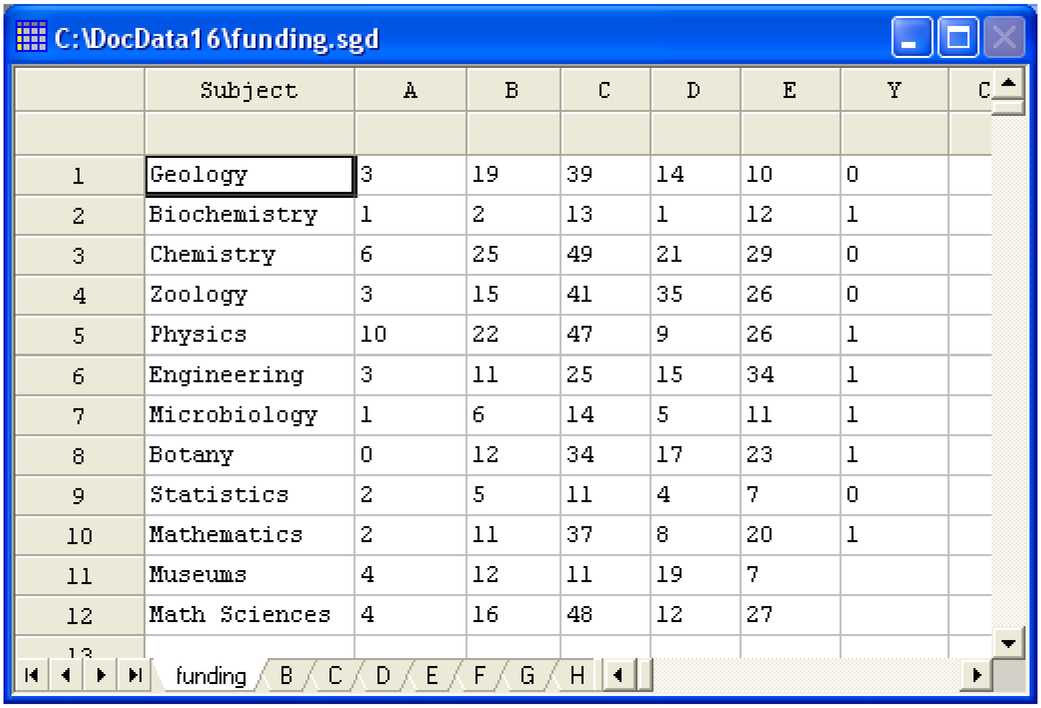


- **Row variable:** name of numeric or non-numeric column identifying the row category for each individual.

- **Column variable:** name of numeric or non-numeric column identifying the column category for each individual.

- **Select:** subset selection.

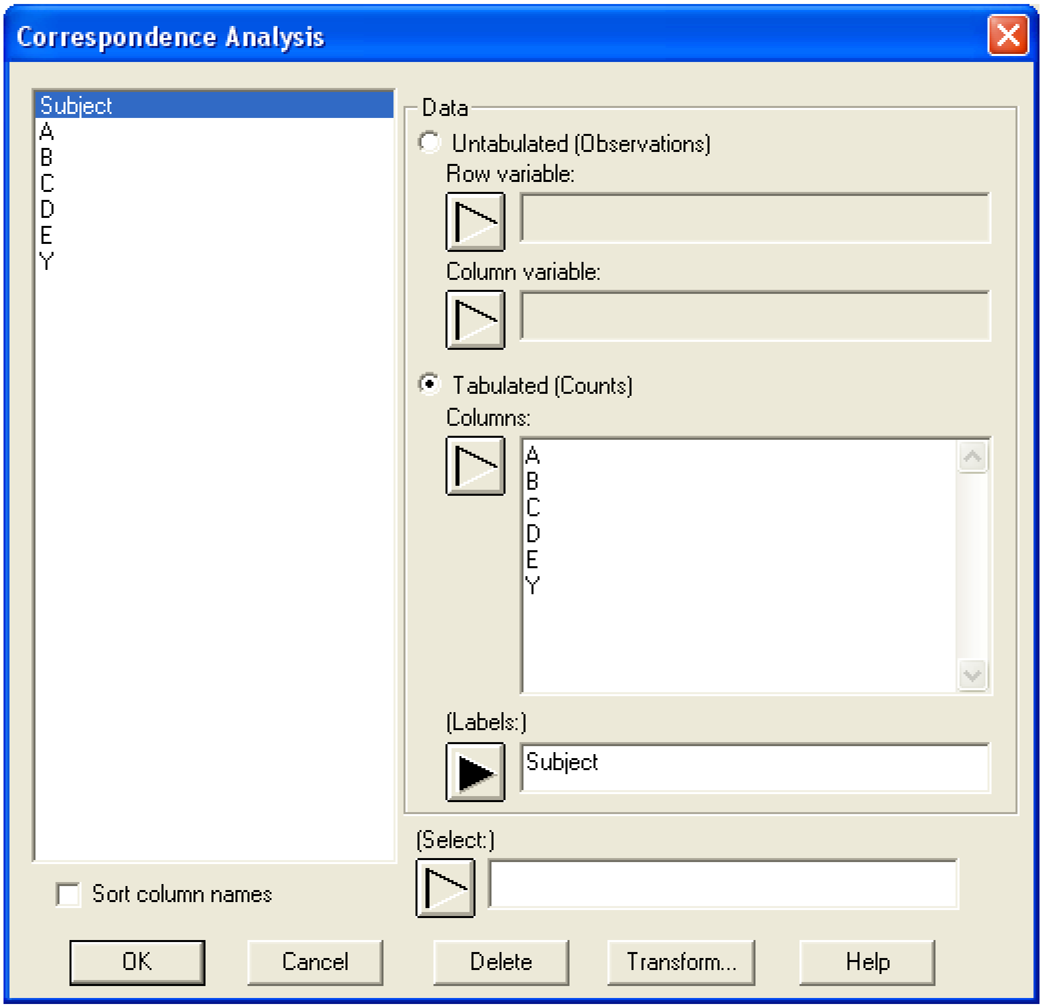
On the other hand, the data file might contain tabulated counts and be laid out in a format similar to the two-way table:


C:\DocData16\funding.sgd

	Subject	A	B	C	D	E	Y
1	Geology	3	19	39	14	10	0
2	Biochemistry	1	2	13	1	12	1
3	Chemistry	6	25	49	21	29	0
4	Zoology	3	15	41	35	26	0
5	Physics	10	22	47	9	26	1
6	Engineering	3	11	25	15	34	1
7	Microbiology	1	6	14	5	11	1
8	Botany	0	12	34	17	23	1
9	Statistics	2	5	11	4	7	0
10	Mathematics	2	11	37	8	20	1
11	Museums	4	12	11	19	7	
12	Math Sciences	4	16	48	12	27	

In that case, select the *Tabulated (Counts)* radio button and specify the names of the columns as shown below:

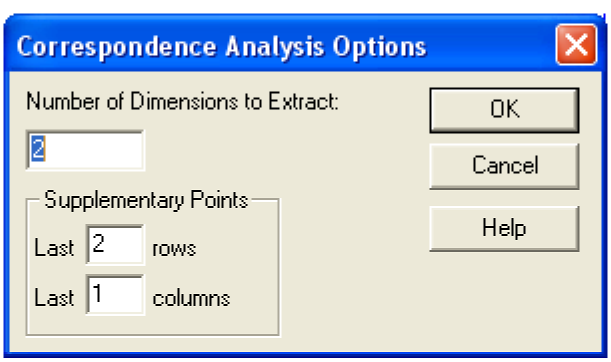
© 2009 by StatPoint Technologies, Inc.



- **Counts:** two or more numeric columns containing the counts in the table.
- **Labels:** column containing identifiers for each row of the table.
- **Select:** subset selection.

Analysis Options

The *Analysis Options* dialog box is shown below:



© 2009 by StatPoint Technologies, Inc.

- **Number of Dimensions to Extract:** the number of dimensions that you wish to extract from the table. You must specify a number between 2 and one less than the smaller of the number of rows and the number of columns. Often, 2 or 3 dimensions is sufficient to explain most of the differences amongst the categories.

- **Supplementary Points:** the number of rows and/or columns that you wish to exclude from the calculations. These rows and columns will be plotted on the correspondence map but will not be used to determine the scaling.

Analysis Summary

The *Analysis Summary* displays the data in the form of a contingency table:

Correspondence Analysis

Contingency Table

	A	B	C	D	E	TOTAL
Geology	3	19	39	14	10	85
Biochemistry	1	2	13	1	12	29
Chemistry	6	25	49	21	29	130
Zoology	3	15	41	35	26	120
Physics	10	22	47	9	26	114
Engineering	3	11	25	15	34	88
Microbiology	1	6	14	5	11	37
Botany	0	12	34	17	23	86
Statistics	2	5	11	4	7	29
Mathematics	2	11	37	8	20	78
TOTAL	31	128	310	129	198	796

Supplemental Rows

	A	B	C	D	E
Museums	4	12	11	19	7
Math Sciences	4	16	48	12	27

Supplemental Columns

	Y
Geology	0
Biochemistry	1
Chemistry	0
Zoology	0
Physics	1
Engineering	1
Microbiology	1
Botany	1
Statistics	0
Mathematics	1

Any supplemental rows or columns will be displayed separately from the counts used to calculate the scaling.

Row and Column Profiles

An important concept in correspondence analysis is that of a row or column *profile*. A profile is simply a row or column of a contingency table, divided by its total. The row and column profiles for the sample data are shown below:

Row and Column Profiles

Row Profiles

	A	B	C	D	E	MASS
Geology	0.035	0.224	0.459	0.165	0.118	0.107
Biochemistry	0.034	0.069	0.448	0.034	0.414	0.036
Chemistry	0.046	0.192	0.377	0.162	0.223	0.163
Zoology	0.025	0.125	0.342	0.292	0.217	0.151
Physics	0.088	0.193	0.412	0.079	0.228	0.143
Engineering	0.034	0.125	0.284	0.170	0.386	0.111
Microbiology	0.027	0.162	0.378	0.135	0.297	0.046
Botany	0.000	0.140	0.395	0.198	0.267	0.108
Statistics	0.069	0.172	0.379	0.138	0.241	0.036
Mathematics	0.026	0.141	0.474	0.103	0.256	0.098
MASS	0.039	0.161	0.389	0.162	0.249	

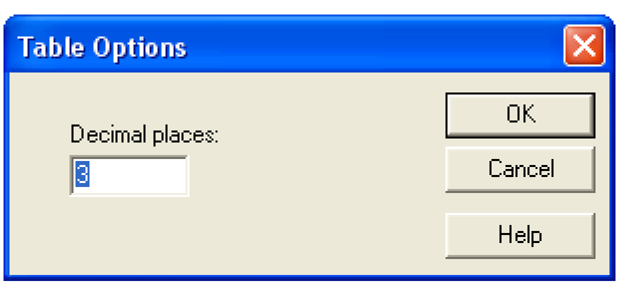
Column Profiles

	A	B	C	D	E	MASS
Geology	0.097	0.148	0.126	0.109	0.051	0.107
Biochemistry	0.032	0.016	0.042	0.008	0.061	0.036
Chemistry	0.194	0.195	0.158	0.163	0.146	0.163
Zoology	0.097	0.117	0.132	0.271	0.131	0.151
Physics	0.323	0.172	0.152	0.070	0.131	0.143
Engineering	0.097	0.086	0.081	0.116	0.172	0.111
Microbiology	0.032	0.047	0.045	0.039	0.056	0.046
Botany	0.000	0.094	0.110	0.132	0.116	0.108
Statistics	0.065	0.039	0.035	0.031	0.035	0.036
Mathematics	0.065	0.086	0.119	0.062	0.101	0.098
MASS	0.039	0.161	0.389	0.162	0.249	

For example, the top left cell *(Geology – A)* represents approximately 3.5% of the *Geology* row and approximately 9.7% of the *A* column.

The table also displays the total *mass* associated with each row and column. Mass is a measure of the total count in a row or column divided by the total count in the entire table. For example, about 10.7% of the researchers in the table are in *Geology*.

Pane Options

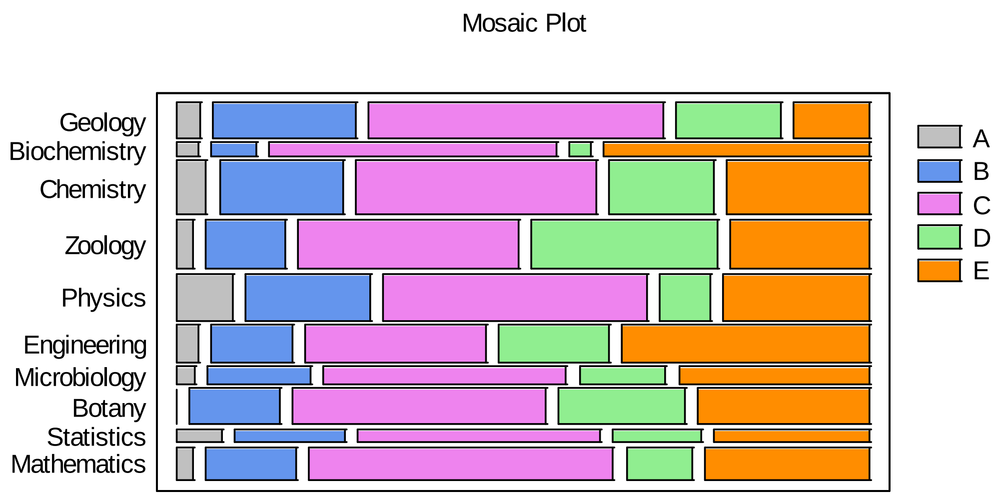


Specify the number of decimal places for displaying the profiles.

© 2009 by StatPoint Technologies, Inc.

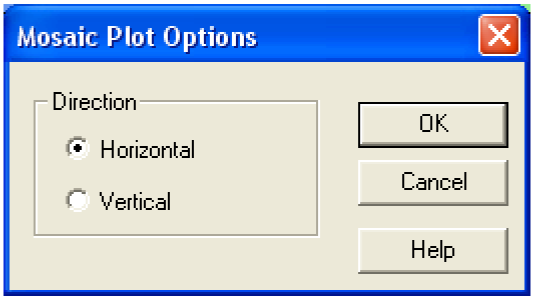
Mosaic Plot

An interesting way to plot the data in a contingency table is through a *Mosaic Plot*:



In this plot, the height of each row is proportional to its row total. The width of the bars within each row is proportional to the frequency of each cell *within that row* (its row profile). This results in bars whose *area* is proportional to the frequency in a particular cell.

Pane Options



- **Direction:** the orientation of the bars. Switching to a vertical plot will plot column profiles rather than row profiles.

Expected Frequencies and Residuals

If the row and column classifications in a contingency table are independent, the expected count in a cell is the total count in the table, multiplied by the row mass, multiplied by the column mass. The *Expected Frequencies and Residuals* pane displays these expected values, together with the differences between the actual counts and those expectations:

Expected Frequencies and Residuals

Expected Frequencies

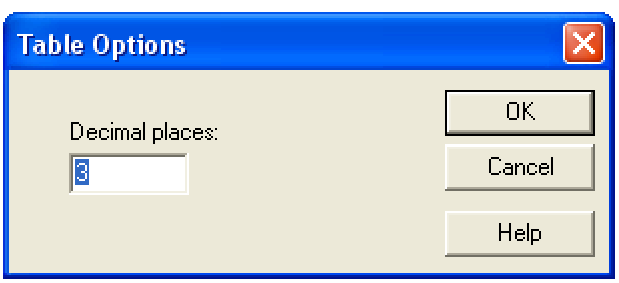
	A	*B*	*C*	*D*	*E*
Geology	3.31	13.67	33.10	13.78	21.14
Biochemistry	1.13	4.66	11.29	4.70	7.21
Chemistry	5.06	20.90	50.63	21.07	32.34
Zoology	4.67	19.30	46.73	19.45	29.85
Physics	4.44	18.33	44.40	18.47	28.36
Engineering	3.43	14.15	34.27	14.26	21.89
Microbiology	1.44	5.95	14.41	6.00	9.20
Botany	3.35	13.83	33.49	13.94	21.39
Statistics	1.13	4.66	11.29	4.70	7.21
Mathematics	3.04	12.54	30.38	12.64	19.40

Residuals (Observed - Expected)

	A	*B*	*C*	*D*	*E*
Geology	-0.31	5.33	5.90	0.22	-11.14
Biochemistry	-0.13	-2.66	1.71	-3.70	4.79
Chemistry	0.94	4.10	-1.63	-0.07	-3.34
Zoology	-1.67	-4.30	-5.73	15.55	-3.85
Physics	5.56	3.67	2.60	-9.47	-2.36
Engineering	-0.43	-3.15	-9.27	0.74	12.11
Microbiology	-0.44	0.05	-0.41	-1.00	1.80
Botany	-3.35	-1.83	0.51	3.06	1.61
Statistics	0.87	0.34	-0.29	-0.70	-0.21
Mathematics	-1.04	-1.54	6.62	-4.64	0.60

For example, the expected count in the upper left corner of the table if *Discipline* and *Funding* are independent is 3.31. Subtracting 3.31 from the observed value of 3 leaves a residual equal to -0.31.

Pane Options



Specify the number of decimal places for displaying the frequencies and residuals.

Chi-Square Distances

A well-known statistic for measuring how much the counts in a contingency table differ from that expected if the row and column classifications were independent is given by the *chi-square statistic*. If e_{ij} represents the expected count in the cell for row *i*, column *j* and n_{ij} represents the actual count, then the chi-square statistic is calculated from:

$$X^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(n_{ij}-e_{ij})^2}{e_{ij}} \tag{1}$$

The table below shows the contributions of each cell to the chi-square statistic:

Chi-Square Distances

	A	B	C	D	E	TOTAL
Geology	0.029	2.080	1.050	0.004	5.873	9.036
Biochemistry	0.015	1.521	0.258	2.913	3.176	7.882
Chemistry	0.173	0.802	0.052	0.000	0.344	1.373
Zoology	0.599	0.957	0.703	12.438	0.496	15.194
Physics	6.964	0.734	0.153	4.859	0.196	12.906
Engineering	0.053	0.702	2.508	0.038	6.700	10.001
Microbiology	0.135	0.000	0.012	0.166	0.351	0.663
Botany	3.349	0.242	0.008	0.673	0.121	4.393
Statistics	0.671	0.024	0.008	0.104	0.006	0.814
Mathematics	0.354	0.190	1.444	1.704	0.018	3.710
TOTAL	12.343	7.252	6.196	22.899	17.282	65.972

The cell with the largest contribution to the total chi-squared statistic is displayed in red. The combined chi-squared contribution of all cells in a row or column (shown under *TOTAL*) can also be thought of as the distance between an observed row or column profile and the centroid of the all the row or column profiles, where the centroid is the profile expected under the assumption of independence.

Relative Inertias

If the chi-square distances are divided by the total count in the contingency table, then the resulting quantities are called *inertia*. If the inertia in a cell is further divided by the total inertia in the entire table, the result is called a *relative inertia*. Summing over a row or column yields the relative inertia of that row or column. The table below shows the relative inertias of each cell, row and column in the contingency table for the sample data:

Relative Inertias

	A	B	C	D	E	TOTAL
Geology	0.000	0.032	0.016	0.000	0.089	0.137
Biochemistry	0.000	0.023	0.004	0.044	0.048	0.119
Chemistry	0.003	0.012	0.001	0.000	0.005	0.021
Zoology	0.009	0.015	0.011	0.189	0.008	0.230
Physics	0.106	0.011	0.002	0.074	0.003	0.196
Engineering	0.001	0.011	0.038	0.001	0.102	0.152
Microbiology	0.002	0.000	0.000	0.003	0.005	0.010
Botany	0.051	0.004	0.000	0.010	0.002	0.067
Statistics	0.010	0.000	0.000	0.002	0.000	0.012
Mathematics	0.005	0.003	0.022	0.026	0.000	0.056
TOTAL	0.187	0.110	0.094	0.347	0.262	1.000

© 2009 by StatPoint Technologies, Inc.

Note that the cell shown in red contributes about 18.9% of the total inertia in the contingency table.

Inertia and Chi-Square Decomposition

As in a principal components analysis, if the profiles for a contingency table were plotted in a multi-dimensional space, they would rarely cover that space in a uniform manner. Rather, the profiles are often situated geometrically so that most of the variability amongst them is oriented along a small number of dimensions. A correspondence analysis calculates a set of orthogonal dimensions such that the first dimension explains as much inertia as possible, the second explains the next most, etc.

The *Inertia and Chi-Square Decomposition* displays important information about those dimensions. The output for the sample data is shown below:

Inertia and Chi-Square Decomposition

	Singular		*Chi-*		*Cumulative*	
Dimension	*Value*	*Inertia*	*Square*	*Percentage*	*Percentage*	*Histogram*
1	0.1978	0.0391	31.1368	47.1973	47.1973	***************
2	0.1743	0.0304	24.1831	36.6569	83.8542	***********
3	0.1043	0.0109	8.6519	13.1146	96.9688	****
4	0.0501	0.0025	1.9997	3.0312	100.0000	*
TOTAL		0.0829	65.971			

This table displays the following information for each dimension:

- **Singular Value** - the square roots of the eigenvalues of a square, symmetric matrix calculated from the differences between observed and expected profiles.

- **Inertia** - the eigenvalues of that matrix. The largest the inertia for a particular dimension, the more variability amongst the profiles that dimension represents.

- **Chi-Square** – the contribution of a particular dimension to the chi-squared statistic.

- **Percentage** - the percentage of the total inertia or total chi-square statistic represented by each dimension.

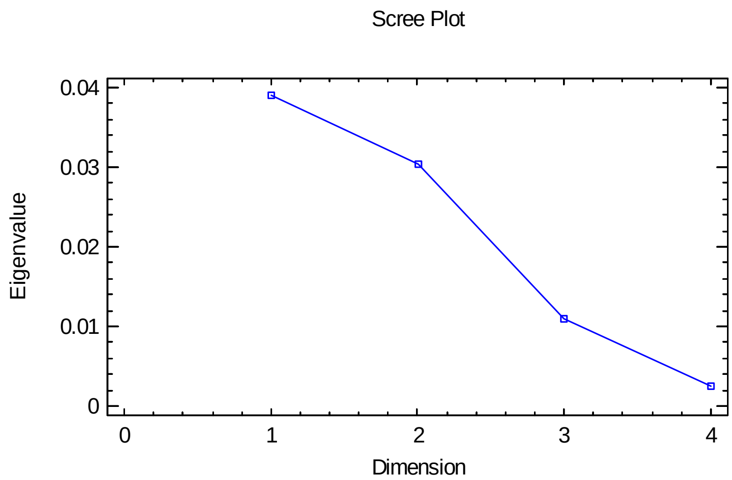
- **Cumulative Percentage** - the percentage of total inertia represented by a selected dimension and those extracted earlier. Often, two or three dimensions are sufficient to represent a large percentage of the total.

- **Histogram** – a graphical representation of each percentage.

In the current example, the first two dimensions account for almost 84% of the total inertia in the contingency table.

Scree Plot

A useful plot in helping to determine how many dimensions are necessary to adequately represent the contingency table is the *Scree Plot*, which plots the eigenvalues in decreasing order:



Often, there will be several large values, after which the remainder of the values will be relatively small.

© 2009 by StatPoint Technologies, Inc.

Row and Column Contributions

Once the principal dimensions have been calculated, the coordinates of the rows and columns in those dimensions can be examined. The *Row and Column Contributions* pane for the sample data is shown below:

Row and Column Contributions

Row Contributions

						Dim #1			*Dim #2*	
		Quality	*Mass*	*Inertia*	*Coord*	*Corr*	*Contr*	*Coord*	*Corr*	*Contr*
1	Geology	0.916	0.107	0.137	-0.076	0.055	0.016	-0.303	0.861	0.322
2	Biochemistry	0.881	0.036	0.119	-0.180	0.119	0.030	0.455	0.762	0.248
3	Chemistry	0.644	0.163	0.021	-0.038	0.134	0.006	-0.073	0.510	0.029
4	Zoology	0.929	0.151	0.230	0.327	0.846	0.413	-0.102	0.083	0.052
5	Physics	0.886	0.143	0.196	-0.316	0.880	0.365	-0.027	0.006	0.003
6	Engineering	0.870	0.111	0.152	0.117	0.121	0.039	0.292	0.749	0.310
7	Microbiology	0.680	0.046	0.010	-0.013	0.009	0.000	0.110	0.671	0.018
8	Botany	0.654	0.108	0.067	0.179	0.625	0.088	0.039	0.029	0.005
9	Statistics	0.561	0.036	0.012	-0.125	0.554	0.014	-0.014	0.007	0.000
10	Mathematics	0.319	0.098	0.056	-0.107	0.240	0.029	0.061	0.079	0.012

Supplemental Rows

						Dim #1			*Dim #2*	
		Quality	*Mass*	*Inertia*	*Coord*	*Corr*	*Contr*	*Coord*	*Corr*	*Contr*
11	Museums	0.556	0.067	0.353	0.314	0.225	0.168	-0.381	0.331	0.318
12	Math Sciences	0.559	0.134	0.041	-0.112	0.493	0.043	0.041	0.066	0.007

Column Contributions

						Dim #1			*Dim #2*	
		Quality	*Mass*	*Inertia*	*Coord*	*Corr*	*Contr*	*Coord*	*Corr*	*Contr*
1	A	0.587	0.039	0.187	-0.478	0.574	0.228	-0.072	0.013	0.007
2	B	0.816	0.161	0.110	-0.127	0.286	0.067	-0.173	0.531	0.159
3	C	0.465	0.389	0.094	-0.083	0.341	0.068	-0.050	0.124	0.032
4	D	0.968	0.162	0.347	0.390	0.859	0.632	-0.139	0.109	0.103
5	E	0.990	0.249	0.262	0.032	0.012	0.006	0.292	0.978	0.699

Supplemental Columns

						Dim #1			*Dim #2*	
		Quality	*Mass*	*Inertia*	*Coord*	*Corr*	*Contr*	*Coord*	*Corr*	*Contr*
6	Y	0.807	0.008	0.122	-0.269	0.068	0.014	0.888	0.739	0.196

The output displays:

- **Quality** – a measure of how well a row or column category can be represented by the number of dimensions that have been extracted relative to its representation using all dimensions. The closer quality is to 1, the better the representation.

- **Mass** – the proportion of observations in each row and column.

- **Inertia** – the relative inertia of each row and column category. The values sum to 1 for those categories that were used in the calculations (for rows and columns separately). For supplemental categories, their size relative to the other categories is of interest.

© 2009 by StatPoint Technologies, Inc.

- **Coord**. – the principal coordinate or location of the category along the indicated dimension. These values may be plotted using either the *Correspondence Map* or the coordinate plots.

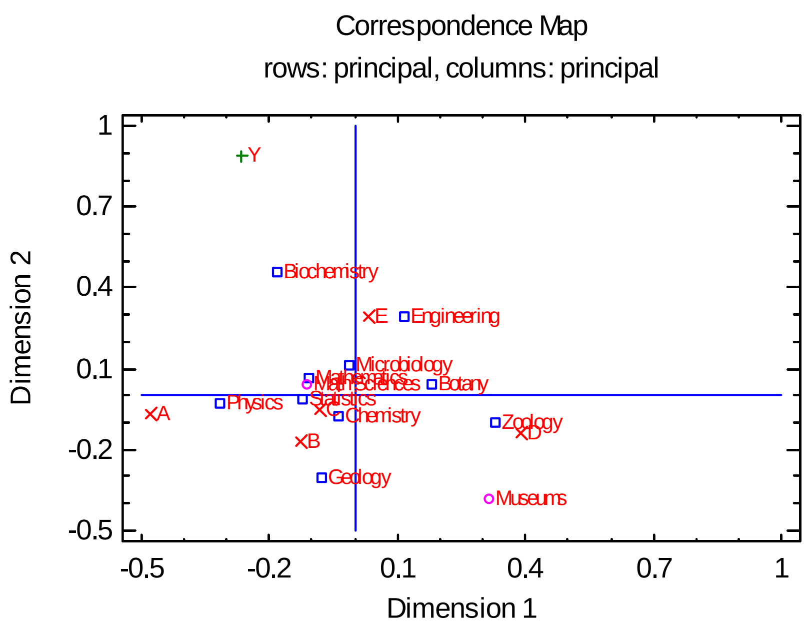
- **Corr**. – the correlation between a selected category and the axis defining a given dimension.

- **Contr**. – the relative contribution of an individual category to the inertia of a selected dimension. The higher the contribution, the more that category contributes to variability along that dimension.

For example, *Zoology* and *Physics* together contribute almost 80% of the column inertia along dimension #1. On the other hand, the quality of the representation for *Mathematics* based upon the two dimensions is relatively low.

2D Correspondence Map

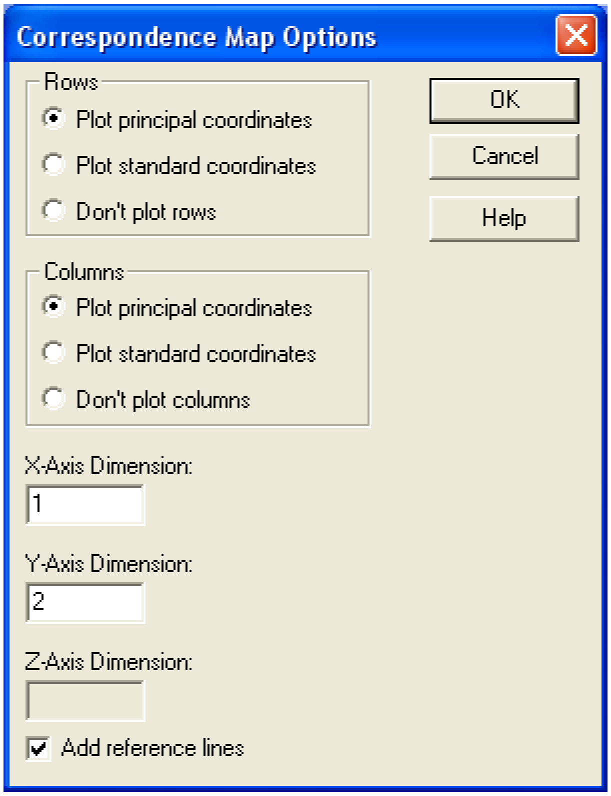
The coordinates of the rows and/or columns may be plotted for any 2 dimensions by selecting the *2D Correspondence Map*. By default, the coordinates for the first two dimensions are displayed:



Different point symbols are used for the rows, columns, and supplementary points.

The distance between any two rows or between any two columns is indicative of their similarity. For example, note that the points labeled *Mathematics*, *Math Sciences*, and *Statistics* are all very close together.

Pane Options



- **Rows –** the type of row coordinates to plot, if any. The weighted average of the squared principal coordinates equals the eigenvalue of that dimension, while the weighted average of the squared standard coordinates equals 1.

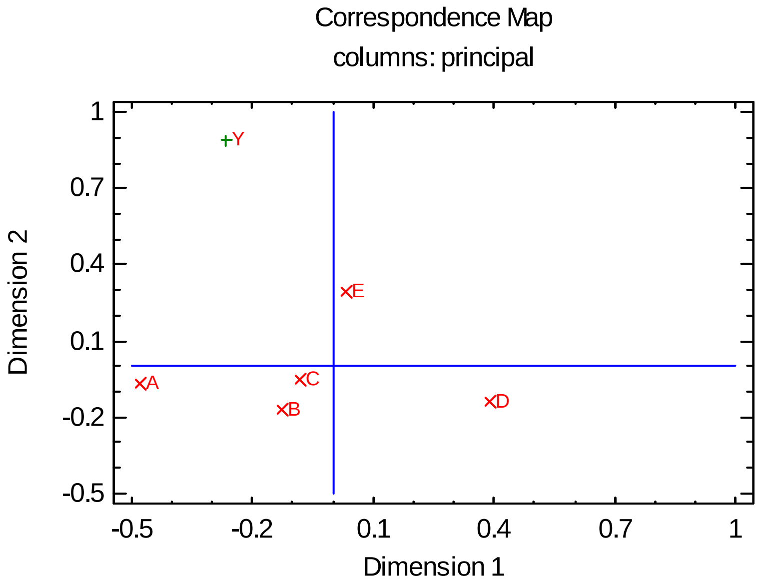
- **Columns –** the type of column coordinates to plot, if any.

- **X-Axis Dimension –** the dimension to plot along the horizontal axis.

- **Y-Axis Dimension –** the dimension to plot along the vertical axis.

- **Add reference lines –** whether to add vertical and horizontal lines through the origin.

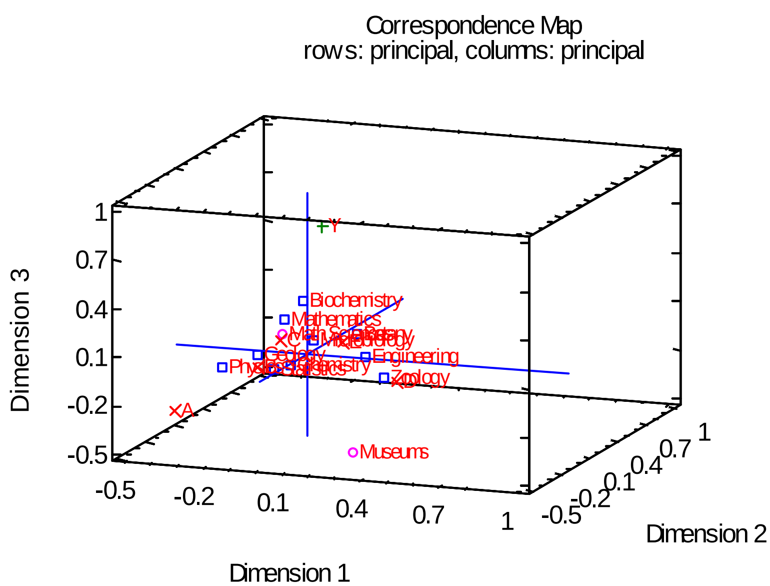
The plot below shows only the columns:



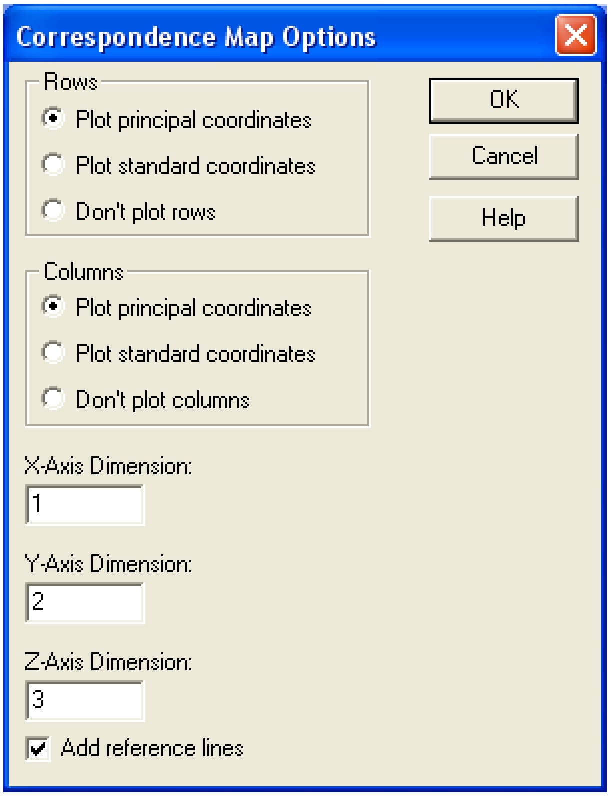
Note that the funding levels *A*, *B*, *C* and *D* line up in sequence along dimension 1. Dimension 2 appears to contrast those 4 levels with *E* (no funding). The supplemental point *Y*, for which there is very little data, is far removed from the other points.

3D Correspondence Map

The coordinates of the rows and/or columns may be plotted for any 3 dimensions by selecting the *3D Correspondence Map*. By default, the coordinates for the first three dimensions are displayed:



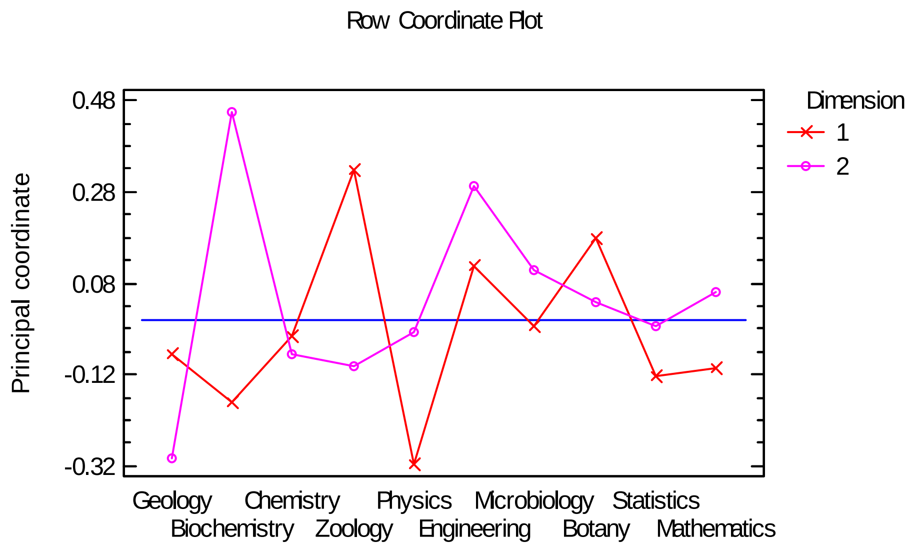
Pane Options



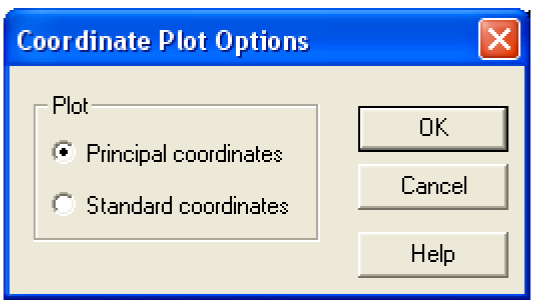
- **Rows** – the type of row coordinates to plot, if any.
- **Columns** – the type of column coordinates to plot, if any.
- **X-Axis Dimension** – the dimension to plot along the X axis.
- **Y-Axis Dimension** – the dimension to plot along the Y axis.
- **Z-Axis Dimension** – the dimension to plot along the Z axis.
- **Add reference lines** – whether to add reference lines through the origin.

Plot of Row Coordinates

This pane displays the row coordinates for each extracted dimension:



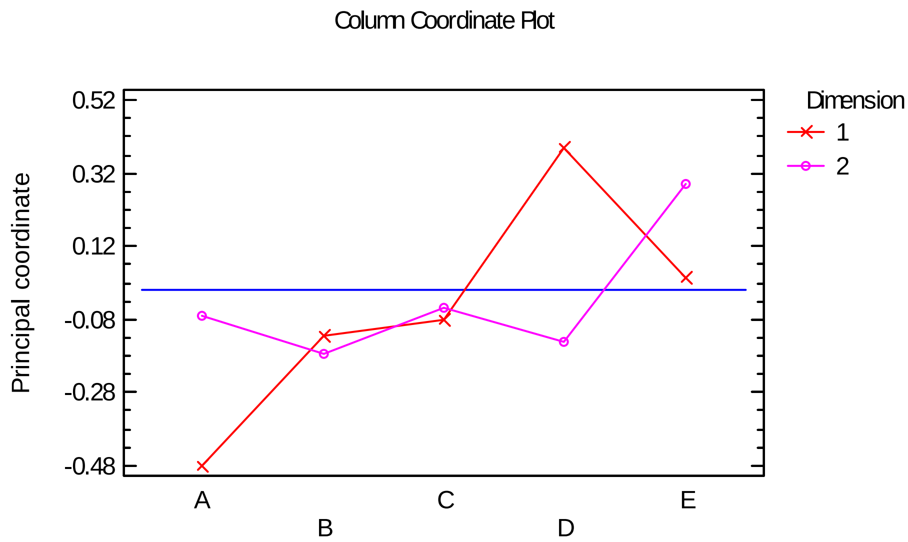
Pane Options



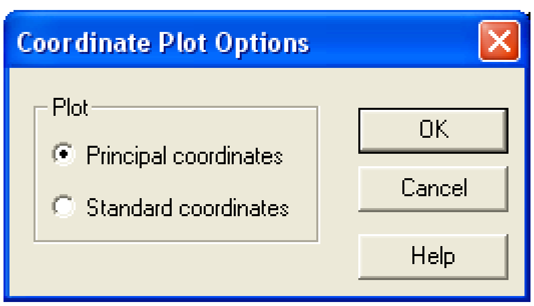
- **Plot** – the type of coordinates to display.

Plot of Column Coordinates

This pane displays the column coordinates for each extracted dimension:



Pane Options



- **Plot** – the type of coordinates to display.

Save Results

You may save the following results to the datasheet:

1. **Point Labels –** labels identifying each row and column.
2. **Point Quality –** the quality of the representation for each row and column.
3. **Point Mass –** the mass of each row and column.
4. **Point Inertia –** the relative inertia of each row and column.
5. **Principal Coordinates –** the principal coordinates for each row and column, for each extracted dimension.
6. **Standard Coordinates –** the standard coordinates for each row and column, for each extracted dimension.
7. **Point Correlations –** the correlations for each row and column, for each extracted dimension.
8. **Point Contributions -** the contributions of each row and column, for each extracted dimension.

© 2009 by StatPoint Technologies, Inc.

Calculations

r = number of rows in contingency table
c = number of columns
v = number of extracted dimensions.
n_{ij} = cell count in row i, column j

Totals

Row i total: $$n_{i.} = \sum_{j=1}^{c} n_{ij} \tag{2}$$

Column j total: $$n_{.j} = \sum_{i=1}^{r} n_{ij} \tag{3}$$

Table total: $$n_{..} = \sum_{i=1}^{r} \sum_{j=1}^{c} n_{ij} \tag{4}$$

Profiles

Profile for row i: $$p_{ij} = \frac{n_{ij}}{n_{i.}} \tag{5}$$

Profile for column j: $$q_{ij} = \frac{n_{ij}}{n_{.j}} \tag{6}$$

Masses

Mass for row i: $$m_{i.} = \frac{n_{i.}}{n_{..}} \tag{7}$$

Mass for column j: $$m_{.j} = \frac{n_{.j}}{n_{..}} \tag{8}$$

Expected Frequencies

Expected frequency for cell (i,j): $$e_{ij} = n_{..} m_{i.} m_{.j} \tag{9}$$

Residuals

Residual for cell (i,j): $$r_{ij} = n_{ij} - e_{ij} \tag{10}$$

© 2009 by StatPoint Technologies, Inc.

Chi-Square Distance

For cell (*i*,*j*): $d_{ij} = \frac{r_{ij}^2}{e_{ij}}$ (11)

Total: $X^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} d_{ij}$ (12)

Inertia

For cell (*i*,*j*): $in_{ij} = \frac{d_{ij}}{n_{..}}$ (13)

Total: $in_{..} = \frac{X^2}{n_{..}}$ (14)

Relative Inertia

For cell (*i*,*j*): $rin_{ij} = \frac{in_{ij}}{in_{..}}$ (15)

For row *i*: $rin_{i.} = \frac{in_{i.}}{in_{..}}$ (16)

For column *j*: $rin_{.j} = \frac{in_{.j}}{in_{..}}$ (17)

Eigenanalysis

Construct *r* by *c* matrix *T* with element:

$$t_{ij} = \frac{\frac{n_{ij}}{n_{..}} - m_{i.} m_{.j}}{\sqrt{m_{i.} m_{.j}}} \quad (18)$$

Calculate eigenvalues *eval* and eigenvectors *evect* of matrix

S = T'T (19)

Principal Coordinates

For row i, dimension k: $$rpc_{ik} = \frac{\sum_{j=1}^{c} t_{ij} evect_{jk}}{\sqrt{m_{i.}}} \quad (20)$$

For column j, dimension k: $$cpc_{jk} = \frac{evect_{jk} \sqrt{eval_k}}{\sqrt{m_{.j}}} \quad (21)$$

Standard Coordinates

For row i, dimension k: $$rsc_{ik} = \frac{rpc_{ik}}{\sqrt{eval_k}} \quad (22)$$

For column j, dimension k: $$\mathrm{csc}_{jk} = \frac{cpc_{jk}}{\sqrt{eval_k}} \quad (23)$$

Quality

For row i: $$q_{i.} = \frac{\sum_{k=1}^{v} rpc_{ik}^2}{\sum_{k} rpc_{ik}^2} \quad (24)$$

For column j: $$q_{.j} = \frac{\sum_{k=1}^{v} cpc_{jk}^2 / rin_{.j}}{\sum_{k} cpc_{jk}^2 / rin_{.j}} \quad (25)$$

Correlation

For row i, dimension k: $$rcorr_{ik} = \frac{rpc_{ik}^2}{\sum_{k} rpc_{ik}^2} \quad (26)$$

For column j, dimension k: $$ccorr_{jk} = \frac{cpc_{jk}^2}{\sum_{k} cpc_{jk}^2} \quad (27)$$

Contribution

For row i, dimension k: $$rcontr_i = \frac{rpc_{ik}^2 m_{i.}}{eval_k} \quad (28)$$

For column j, dimension k: $$ccontr_j = \frac{cpc_{jk}^2 m_{.j}}{eval_k} \quad (29)$$

© 2009 by StatPoint Technologies, Inc.