Coordinate representation of compositional tables

Kamila Fačevicová

Department of Mathematical Analysis and Applications of Mathematics,
Palacký University Olomouc

How to study relationship between factors?

Contingency tables

- discrete nature of cells
- multinomial distribution
- *n* independent observations

Compositional tables

- continuous nature of cells
- multivariate continuous distribution
- 1 observation

CZE	HT	MT	LT
VA	$1.87814 \cdot 10^{10}$	$1.50362 \cdot 10^{10}$	$1.03281 \cdot 10^{10}$
I	$7.46011 \cdot 10^{10}$	$4.67950 \cdot 10^{10}$	$2.95888 \cdot 10^{10}$

Definition of compositional tables

- Special case of $I \cdot J$-part compositional data

$$\mathbf{x} = \mathcal{C}\begin{pmatrix} x_{11} & \cdots & x_{1J} \\ \vdots & \ddots & \vdots \\ x_{I1} & \cdots & x_{IJ} \end{pmatrix}, \quad \text{where} \quad x_{ij} > 0, \forall i, j.$$

- Closure operation

$$\mathcal{C}(\mathbf{x}) = \begin{pmatrix} \frac{\kappa \cdot x_{11}}{\sum_{i,j=1}^{I,J} x_{ij}} & \cdots & \frac{\kappa \cdot x_{1J}}{\sum_{i,j=1}^{I,J} x_{ij}} \\ \vdots & \ddots & \vdots \\ \frac{\kappa \cdot x_{I1}}{\sum_{i,j=1}^{I,J} x_{ij}} & \cdots & \frac{\kappa \cdot x_{IJ}}{\sum_{i,j=1}^{I,J} x_{ij}} \end{pmatrix}.$$

- The sample space is subspace of $I \cdot J$-part simplex with dimension $I \cdot J - 1$

$$\mathcal{S}^{IJ} = \{\mathbf{x} = (x_{11}, \ldots, x_{1J}, \ldots, x_{IJ}) | x_{ij} > 0, \text{ i=1, 2, ..., I, j=1, 2, ..., J}; \sum_{i,j=1}^{I,J} x_{ij} = \kappa\}.$$

Basic operations with compositional tables

- Perturbation:

$$\mathbf{x} \oplus \mathbf{y} = \mathcal{C} \begin{pmatrix} x_{11}y_{11} & \cdots & x_{1J}y_{1J} \\ \vdots & \ddots & \vdots \\ x_{I1}y_{I1} & \cdots & x_{IJ}y_{IJ} \end{pmatrix}$$

- Power transformation:

$$\alpha \odot \mathbf{x} = \mathcal{C} \begin{pmatrix} x_{11}^{\alpha} & \cdots & x_{1J}^{\alpha} \\ \vdots & \ddots & \vdots \\ x_{I1}^{\alpha} & \cdots & x_{IJ}^{\alpha} \end{pmatrix}$$

- Inner product:

$$\langle \mathbf{x}, \mathbf{y} \rangle_a = \frac{1}{2IJ} \sum_{i,j} \sum_{k,l} \ln \frac{x_{ij}}{x_{kl}} \ln \frac{y_{ij}}{y_{kl}}$$

- Norm and distance:

$$\|\mathbf{x}\|_a = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle_a} \quad \text{a} \quad d_a(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} \ominus \mathbf{y}\|_a$$

Decomposition of the compositional table

Projections of compositional table $\mathbf{x}$:

- $\mathrm{row}_i(\mathbf{x})$ - relations between parts of the i-th row.
- $\mathrm{col}_j(\mathbf{x})$ - relations between parts of the j-th column.
- $\mathrm{row}^{\perp}(\mathbf{x})$ - relations between parts of different rows.
- $\mathrm{col}^{\perp}(\mathbf{x})$ - relations between parts of the different columns.

Original table $\mathbf{x}$ could be decomposed as

$$\mathbf{x} = \mathrm{row}^{\perp}(\mathbf{x}) \oplus \left(\bigoplus_{i=1}^{I} \mathrm{row}_i(\mathbf{x}) \right)$$

or

$$\mathbf{x} = \mathrm{col}^{\perp}(\mathbf{x}) \oplus \left(\bigoplus_{j=1}^{J} \mathrm{col}_j(\mathbf{x}) \right).$$

Decomposition of the compositional table

Compositional table could be decomposed onto its independent and interactive parts.

$$\mathbf{x} = \mathbf{x}_{ind} \oplus \mathbf{x}_{int}.$$

Independence table:

$$\mathbf{x}_{ind} = \mathrm{row}^{\perp}(\mathbf{x}) \oplus \mathrm{col}^{\perp}(\mathbf{x}) = \left(x_{ij}^{ind} = \left(\prod_{k=1}^{I} \prod_{l=1}^{J} x_{kj} x_{il} \right)^{\frac{1}{IJ}} \right)_{i,j=1}^{I,J}.$$

Interaction table:

$$\mathbf{x}_{int} = \mathbf{x} \ominus \mathbf{x}_{ind} = \left(x_{ij}^{int} = \left(\prod_{k=1}^{I} \prod_{l=1}^{J} \frac{x_{ij}}{x_{kj} x_{il}} \right)^{\frac{1}{IJ}} \right)_{i,j=1}^{I,J}.$$

Relationship between factors

If row and column factors are independent

- whole information about $\mathbf{x}$ carries the independence table,
- interaction table is the neutral element,
- all coordinates of $\mathbf{x}_{int}$ are 0.

Expression of the interaction table in coordinates

Coordinates of the D-part composition constitute $D-1$-dimensional real vector

$$\mathbf{z} = h(\mathbf{x}) = (\langle \mathbf{x}, \mathbf{e}_1 \rangle_a, \dots, \langle \mathbf{x}, \mathbf{e}_{D-1} \rangle_a) = (z_1, z_2, \dots, z_{D-1}),$$

where $\mathbf{e}_i = \mathcal{C}(e_{i1}, \dots, e_{i,D})$, $i = 1, \dots, D-1$ constitute the orthonormal basis of the D-part simplex.

Following properties hold

$$h(\alpha \odot \mathbf{x}_1 \oplus \beta \odot \mathbf{x}_2) = \alpha \cdot \mathbf{z}_1 + \beta \cdot \mathbf{z}_2, \quad \langle \mathbf{x}_1, \mathbf{x}_2 \rangle_a = \langle \mathbf{z}_1, \mathbf{z}_2 \rangle.$$

Expression of the interaction table in coordinates

$$z_{rs}^{int} = \frac{1}{\sqrt{r \cdot s \cdot (r-1) \cdot (s-1)}} \ln \prod_{i=1}^{r-1} \prod_{j=1}^{s-1} \frac{x_{ij} x_{rs}}{x_{is} x_{rj}}$$

or

$$\sqrt{\frac{(r-1) \cdot (s-1)}{r \cdot s}} \ln \frac{\sqrt[(r-1)(s-1)]{x_{11} x_{12} \cdots x_{1,s-1} \cdots x_{r-1,1} \cdots x_{r-1,s-1}} \cdot x_{rs}}{\sqrt[s-1]{x_{r1} \cdots x_{r,s-1}} \cdot \sqrt[r-1]{x_{1s} \cdots x_{r-1,s}}}$$

for $r = 2, 3, \ldots, I$ and $s = 2, 3, \ldots, J$.

- $(I-1)(J-1)$ nonzero coordinates,
- related to odds ratio.

Expression of the interaction table in coordinates

$$z_{22}^{int} = \frac{1}{2} \ln \frac{x_{11}x_{22}}{x_{12}x_{21}}$$

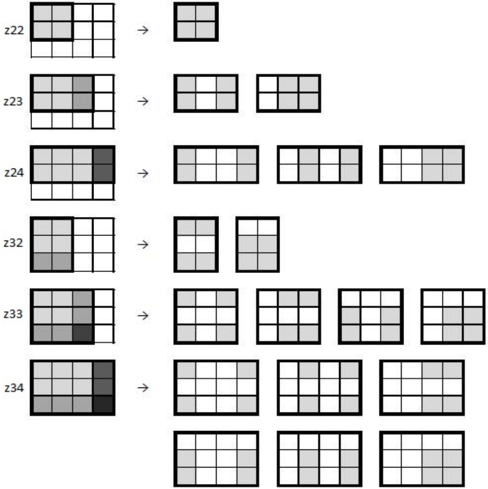
$$z_{23}^{int} = \frac{1}{\sqrt{12}} \ln \frac{x_{11}x_{12}x_{23}^2}{x_{21}x_{22}x_{13}^2}$$

$$z_{24}^{int} = \frac{1}{\sqrt{24}} \ln \frac{x_{11}x_{12}x_{13}x_{24}^3}{x_{21}x_{22}x_{23}x_{14}^3}$$

$$z_{32}^{int} = \frac{1}{\sqrt{12}} \ln \frac{x_{11}x_{21}x_{32}^2}{x_{31}^2x_{12}x_{22}}$$

$$z_{33}^{int} = \frac{1}{\sqrt{36}} \ln \frac{x_{11}x_{12}x_{21}x_{22}x_{33}^4}{x_{31}^2x_{32}^2x_{13}^2x_{23}^2}$$

$$z_{34}^{int} = \frac{1}{\sqrt{72}} \ln \frac{x_{11}x_{12}x_{13}x_{21}x_{22}x_{23}x_{34}^6}{x_{31}^2x_{32}^2x_{33}^2x_{14}^3x_{24}^3}$$



Generalized formula

$$\frac{1}{\sqrt{(I-a)(J-b)IJab}}\cdot$$

$$\ln\frac{x_{11}^{ab}\cdots x_{1,J-b}^{ab}\cdots x_{I-a,1}^{ab}\cdots x_{I-a,J-b}^{ab}\cdot x_{I-a+1,J-b+1}^{(I-a)(J-b)}\cdots x_{I-a+1,J}^{(I-a)(J-b)}\cdots x_{I,J-b+1}^{(I-a)(J-b)}\cdots x_{I,J}^{(I-a)(J-b)}}{x_{1,J-b+1}^{a(J-b)}\cdots x_{1,J}^{a(J-b)}\cdots x_{I-a,J-b+1}^{a(J-b)}\cdots x_{I-a,J}^{a(J-b)}\cdot x_{I-a+1,1}^{b(I-a)}\cdots x_{I-a+1,J-b}^{b(I-a)}\cdots x_{I,1}^{b(I-a)}\cdots x_{I,J-b}^{b(I-a)}},$$

or

$$\sqrt{\frac{(I-a)(J-b)ab}{IJ}}\cdot$$

$$\ln\frac{\sqrt[(I-a)(J-b)]{x_{11}\cdots x_{1,J-b}\cdots x_{I-a,1}\cdots x_{I-a,J-b}}\cdot\sqrt[ab]{x_{I-a+1,J-b+1}\cdots x_{I-a+1,J}\cdots x_{I,J-b+1}\cdots x_{I,J}}}{\sqrt[(I-a)b]{x_{1,J-b+1}\cdots x_{1,J}\cdots x_{I-a,J-b+1}\cdots x_{I-a,J}}\cdot\sqrt[a(J-b)]{x_{I-a+1,1}\cdots x_{I-a+1,J-b}\cdots x_{I,1}\cdots x_{I,J-b}}}.$$

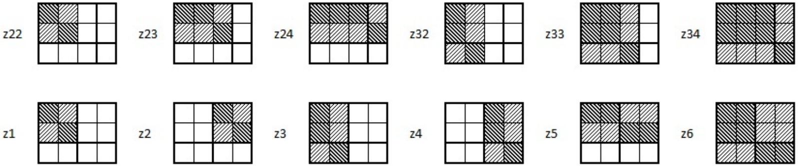
Comparison of methods of expression in coordinates

$$z_{22} = \frac{1}{2}\ln\frac{x_{11}x_{22}}{x_{12}x_{21}} \qquad z_{23} = \frac{1}{2\sqrt{3}}\ln\frac{x_{11}x_{12}x_{23}^2}{x_{13}^2x_{21}x_{22}} \qquad z_{24} = \frac{1}{2\sqrt{6}}\ln\frac{x_{11}x_{12}x_{13}x_{24}^3}{x_{14}^3x_{21}x_{22}x_{23}}$$

$$z_{32} = \frac{1}{2\sqrt{3}}\ln\frac{x_{11}x_{21}x_{32}^2}{x_{12}x_{22}x_{31}^2} \qquad z_{33} = \frac{1}{6}\ln\frac{x_{11}x_{12}x_{21}x_{22}x_{33}^4}{x_{13}^2x_{23}^2x_{31}^2x_{32}^2} \qquad z_{34} = \frac{1}{6\sqrt{2}}\ln\frac{x_{11}x_{12}x_{13}x_{21}x_{22}x_{23}x_{34}^6}{x_{14}^3x_{24}^3x_{31}^2x_{32}^2x_{33}^2}$$

$$z_1 = \frac{1}{2}\ln\frac{x_{11}x_{22}}{x_{12}x_{21}} \qquad z_2 = \frac{1}{2}\ln\frac{x_{13}x_{24}}{x_{14}x_{23}} \qquad z_3 = \frac{1}{2\sqrt{3}}\ln\frac{x_{11}x_{21}x_{32}^2}{x_{12}x_{22}x_{31}^2}$$

$$z_4 = \frac{1}{2\sqrt{3}}\ln\frac{x_{13}x_{23}x_{34}^2}{x_{14}x_{24}x_{33}^2} \qquad z_5 = \frac{1}{2\sqrt{2}}\ln\frac{x_{11}x_{12}x_{23}x_{24}}{x_{13}x_{14}x_{21}x_{22}} \qquad z_6 = \frac{1}{2\sqrt{6}}\ln\frac{x_{11}x_{12}x_{21}x_{22}x_{33}^2x_{34}^2}{x_{13}x_{14}x_{23}x_{24}x_{31}^2x_{32}^2}$$



Example - Data

- 77 2×3 compositional tables
- distribution of the manufacturing production in a given country for 2008
- Resource efficiency - Value added, Input
- Technology intensity - Low, Medium, High

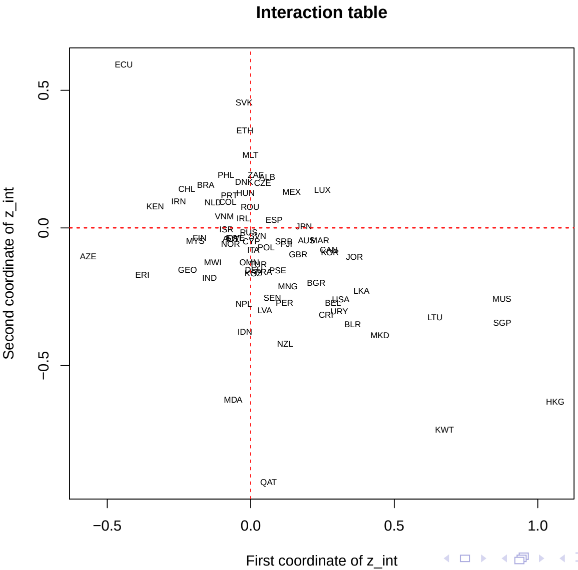
Mauritius	LT	MT	HT
VA	$3.31406 \cdot 10^{10}$	$6.89430 \cdot 10^{9}$	$8.6170 \cdot 10^{8}$
I	$5.75729 \cdot 10^{10}$	$6.89525 \cdot 10^{10}$	$2.2993 \cdot 10^{9}$
VA	0.19526	0.04062	0.00508
I	0.33922	0.40627	0.01355

Example - coordinates

Coordinate	Analytical form	Sample mean	Sample st. dev.
z_{22}^{int}	$\frac{1}{2}\ln\frac{x_{11}x_{22}}{x_{12}x_{21}}$	0.0758	0.2747
z_{23}^{int}	$\frac{1}{\sqrt{12}}\ln\frac{x_{11}x_{12}x_{23}^2}{x_{21}x_{22}x_{13}^2}$	−0.0908	0.2455

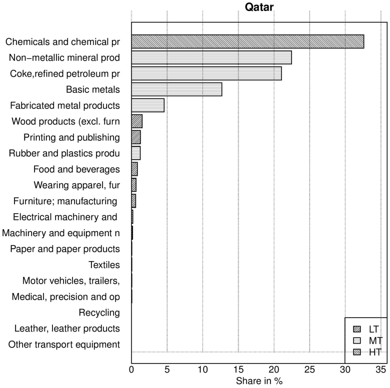
- $z_{22}^{int} = 0.0758 \rightarrow \frac{\frac{\text{VALow}}{\text{InputLow}}}{\frac{\text{VAMed.}}{\text{InputMed}}} = 1.16 \rightarrow$ Value added to Input ratio is slightly higher for low technology.
- $z_{23}^{int} = -0.0908 \rightarrow \frac{\frac{\text{VALow+Med.}}{\text{InputLow+Med.}}}{\frac{\text{VAHigh}}{\text{InputHigh}}} = 0.855 \rightarrow$ Value added to Input ratio is higher for high technology.

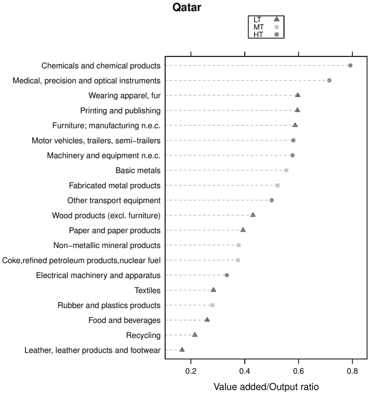
Comparison of countries



Example - Qatar

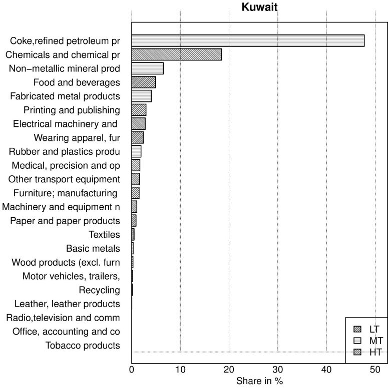
High absolute value of the second coordinate.

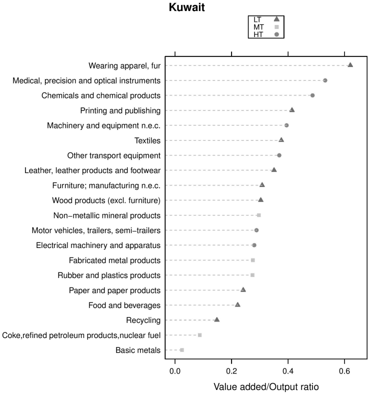




Example - Kuwait

High absolute value of both coordinates.





Conclusions

- Decomposition of compositional tables
- Expression of tables in coordinates
- Generalization of the method
- Analysis of independence between resource efficiency and technology intensity

References

- Aitchison J (1986) *The statistical analysis of compositional data.* Chapman and Hall, London.
- Egozcue JJ, Díaz-Barrero JL, Pawlowsky-Glahn V (2008) *Compositional analysis of bivariate discrete probabilities.* In Daunis-i-Estadella J, Martín-Fernández JA, eds, Proceedings of CODAWORK'08, The 3rd Compositional Data Analysis Workshop. University of Girona, Spain.
- Fačevicová K, Hron K, Todorov V, Guo D, Templ M (2014) *Logratio approach to statistical analysis of 2x2 compositional tables.* Journal of Applied Statistics, 41 (2014), 5, 944-958.
- Fačevicová K, Hron K, Todorov V, Templ M (2014) *Compositional tables analysis in coordinates.* Statistical Science, submitted.