The use of compositional tables in plant research

Data: measurements of N variables on I genotypes of a plant under $J = 2$ conditions (control and stress).

Proposed approach to data analysis:

1) The effect of genotype is assessed by focusing only on the control data – PCA is performed on log-transformed data (matrix of size $I \times N$).

2) The role of both factors and their interaction is examined by transferring the problem into the analysis of N compositional tables of type $I \times 2$ (see Table 1).

Table 1: Illustration a compositional table of type $I \times 2$.

Variable n	Stress	Control
Genotype 1	x_{11}	x_{12}
Genotype 2	x_{21}	x_{22}
$\vdots$	$\vdots$	$\vdots$
Genotype I	x_{I1}	x_{I2}

Compositional tables can be considered as a continuous counterpart to contingency tables; they carry relative information about relationships between two factors; within the framework of the logratio methodology, orthogonal decomposition into independent and interaction tables is possible; interpretable coordinate representation for independent and interaction tables can be constructed, which enable further statistical processing.

3) Construction of coordinates of compositional tables:

- Coordinates of *independence tables* (I coordinate systems consisting of I coordinates for each of the N tables):
  $z_i^{\mathrm{r}(l)} = \sqrt{\frac{2(I-i)}{I-i+1}} \ln \frac{g(x_{i1}^{(l)}, x_{i1}^{(l)})}{g(x_{i+1,1}^{(l)}, \ldots, x_{I1}^{(l)}, x_{i+1,2}^{(l)}, \ldots, x_{I2}^{(l)})}, \quad i = 1, \ldots, I-1,\ l = 1, \ldots, I;$
  $z_1^{\mathrm{c}} = \sqrt{\frac{I}{2}} \ln \frac{g(x_{11}, x_{21}, \ldots, x_{I1})}{g(x_{12}, x_{22}, \ldots, x_{I2})},$
  where $g()$ stands for geometric mean a and the superscript l refers to the $l-$th row permuted to the pivot (first) position within the whole table.

- Coordinates of *interaction tables* (I coordinate systems consisting of $I-1$ coordinates for each of the N tables):
$$z_i^{\mathrm{OR}(l)} = \sqrt{\frac{I-i}{2(I-i+1)}} \ln \frac{x_{i1}^{(l)}/x_{i2}^{(l)}}{g\left(x_{i+1,1}^{(l)}/x_{i+1,2}^{(l)},\ldots,x_{I1}^{(l)}/x_{I2}^{(l)}\right)}, \quad i = 1,\ldots,I-1, \quad l = 1,\ldots,I.$$

4) Interest in the first coordinates from each of the systems; their interpretation:

- $z_1^{\mathrm{r}(l)}$ – stands for relative contribution (dominance) of the average value of the given variable in genotype l (across both conditions) with respect to its the average value in the remaining genotypes (across both conditions); these coordinates are not of primary interest (analysis of the control data is more informative for assessing the role of genotype by itself) but some interesting information can still be revealed when the coordinates of independence table are examined together.
- z_1^{c} – relative contribution of the average value of the given variable in the stress condition (across all the genotypes) with respect to its average value in the control condition (across all the genotypes); this coordinate enables us to identify which variables change the most under the stress condition – the variables with the lowest (resp. highest) value of z_1^{c} are those which overall decreased (resp. increased) the most.
- $z_1^{\mathrm{OR}(l)}$ – dominance of the value of the ratio stress vs. control in the given variable in genotype l with respect to the average value of those ratios across the remaining genotypes; these coordinates enable us to identify which ratios stress vs. control deviate the most when comparing individual genotypes to the remaining genotypes, they also help us to determine which genotype has overall relatively higher (resp. lower) ratios stress vs. control.

5) PCA for independence and interaction tables:

- PCA for independence tables: performing PCA in each of the I coordinate systems – for data matrix of size $N \times I$ whose rows are formed by values in coordinates $z_1^{\mathrm{r}(l)},\ldots,z_{I-1}^{\mathrm{r}(l)},z_1^{\mathrm{c}}$ for each of the N variables, $l = 1,\ldots,I$; scores are the same in each system due to orthonormality of coordinates, so are the loadings corresponding to coordinate z_1^{c}; for the construction of the biplot loadings corresponding to the first coordinate $z_1^{\mathrm{r}(l)}$ are extracted from each of the system $l = 1,\ldots,I$.
- PCA for interaction tables: performing PCA in each of the I coordinate systems – for data matrix of size $N \times (I-1)$ whose rows are formed

by values of coordinates $z_1^{\mathrm{OR}(l)}, \ldots, z_{I-1}^{\mathrm{OR}(l)}$ for each of the N variables, $l = 1, \ldots, I$; scores are the same in each system; for the construction of the biplot loadings corresponding to the first coordinate $z_1^{\mathrm{OR}(l)}$ are extracted from each of the system $l = 1, \ldots, I$.