AgroStat 2016

# MatrixCorrelation

Kristian Hovde Liland[1,2], Tormod Næs[1], Ulf Geir Indahl[2]

1) Nofima – Norwegian Institute of Food, Fisheries and Aquaculture research
2) NMBU – Norwegian University of Life Sciences

MatrixCorrelation is an R package for comparing data sets. It contains many methods, but emphasizes the use of the Similarity of Matrices Index (SMI). This is a new method that compares stable subspaces from coupled data matrices. Visualization for explorative data analysis and statistical testing of equality are included.

## Matrix comparisons in the package

Ramsey's matrix correlation measures:
- `r1`: inner product correlation
- `r2`: orientation-independent inner product correlation
- `r3`: spectra-independent inner product correlations (including orientation)
- `r4`: Spectra-Independent inner product correlations

RV type measures:
- `RV`: Original RV, much used in sensometrics
- `RV2`: Smilde's variant (less affected by high dimensionality)
- `Rvadj`: Maye's variant (not affected by high dimensionality)

Matrix subspace correlations:
- `GCD`: Yanai's GCD Measure
- `SMI`: Similarity of Matrices Index

Cross-validated Principal Component Analysis (PCA) for complexity assessment:
- PCAcv

## Method comparison

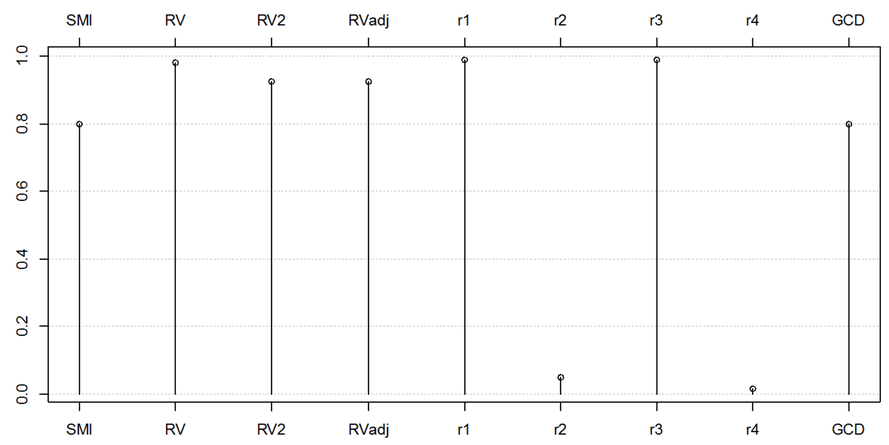
For simple comparison of various matrix correlation measures, the convenience function `allCorrelations` is included.

Sample a normal matrix and centre it:
```
X1  <- scale( matrix( rnorm(100*300), 100,300),
 scale = FALSE)
```

Create a copy where the 3rd spectral component (SVD) has been removed:
```
X2  <- usv$u[,-3] %*% diag(usv$d[-3]) %*% t(usv$v[,-3])
usv <- svd(X1)
```

Display the result for 5,5 components:
```
allCorrelations(X1,X2, ncomp1 = 5, ncomp2 = 5)
```

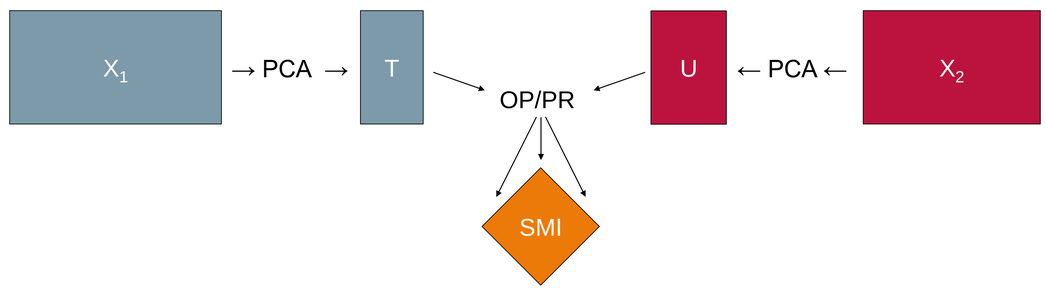
```
 SMI    RV   RV2 RVadj    r1    r2    r3    r4   GCD
0.800 0.980 0.924 0.926 0.988 0.050 0.989 0.017 0.800
```




*Comparison of various matrix correlation measures applied to the simulated dataset.*

## Similarity of Matrices Index

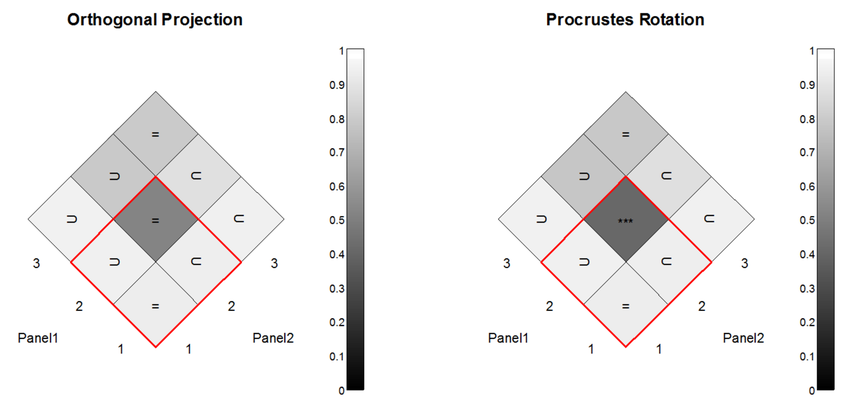
A two-step process starts with extraction of stable subspaces using PCA or some other method yielding two orthonormal bases. These bases are compared using Orthogonal Projection (OP / ordinary least squares) or Procrustes Rotation (PR).




The result is a similarity measure that can be adjusted to various data sets and contexts and which includes explorative plotting and permutation based testing of matrix subspace equality.

```
sens.OP <- SMI(Panel1, Panel2, ncomp1 = 3, ncomp2 = 3)
sens.PR <- SMI(Panel1, Panel2, ncomp1 = 3, ncomp2 = 3,
                               projection = "Procrustes")
RV(Panel1, Panel2)
= 0.93

plot(sens.OP, main = "Orthogonal Projection",
                               replicates = rep(1:6,each=3))
plot(sens.PR, main = "Procrustes Rotation",
                               replicates = rep(1:6,each=3))
```




*Sensory assessment of candy. $SMI_{OP}$ and $SMI_{PR}$ for all combinations of up to three components from assessor Panels 1 and 2. The red square displays the subspace region usually interpreted in sensory analyses. "=", "⊂" and "⊃" shows that $H_0$ is not rejected (equal/included matrix subspaces). Stars indicate rejection of $H_0$ at the significance levels: *** = P<0.001, ** = P<0.01, * = P<0.05.*

## References

**SMI:** Indahl, UG; Næs, T; Liland KH (2016). "A similarity index for comparing coupled matrices" (submitted).

**RV:** Robert, P.; Escoufier, Y. (1976). "A Unifying Tool for Linear Multivariate Statistical Methods: The RV-Coefficient". *Appl. Stat.* 25 (3): 257-265.

**RV2:** Smilde, AK; Kiers, HA; Bijlsma, S; Rubingh, CM; van Erk, MJ (2009). "Matrix correlations for high-dimensional data: the modified RV-coefficient". *Bioinformatics* 25(3): 401-5.

**Adjusted RV:** Maye, CD; Lorent, J; Horgan, GW. (2011). "Exploratory analysis of multiple omics datasets using the adjusted RV coefficient". *Stat. Appl. Genet. Mol. Biol*. 10(14).

**Sensory:** Tomic, O., Luciano, G., Nilsen, A., Hyldig, G., Lorensen, K., Næs, T. (2010). "Analysing sensory panel performance in a proficiency test using the PanelCheck software" *Europ. Food Res. Tech*. 230. 3, 497-511