Camiz, S. & Pillar, V.D.
Identifying the Informational/Signal Dimension in Principal Component
Analysis.
Mathematics

Script with the options chosen in software *MULTIV* (available at
http://ecoqua.ecologia.ufrgs.br/MULTIV.html) for generating simulated data and testing
power and type I error with them. The case depicted here is eigenstructure #6 specified in
Table 1 (see main text) with noise level = 0.02. Selected options and input are in *italics*.
Note that the special option "U data generator with known correlation or eigen structure" is
hidden in the Main Menu of the released versions of the software, yet it is active, as it is
shown below.

```
MULTIV
for multivariate analysis, randomization tests and bootstrapping      v.3.72b
--------------------------------------------------------------------------------
Developed by V.Pillar
Departamento de Ecologia, Universidade Federal do Rio Grande do Sul
Porto Alegre, RS, 91540-000, Brazil          e-mail: vpillar@ufrgs.br
--------------------------------------------------------------------------------

Analysis status: Data not available.


--------------------------------------------------------------------------------
MAIN MENU
      * N   specify data or open existing session
        V   descriptive attributes
        T   transform data
        R   resemblance measures
        G   specify groups of sampling units
        D   scatter diagrams
        O   ordination
        C   cluster analysis
        P   randomization tests comparing groups of sampling units
        A   randomization tests comparing variables
        F   path analysis
        K   relations between two sets of variables
        Y   analysis of dynamics
      * E   preferences
        S   save session
      * Z   run script
      * X   exit
--------------------------------------------------------------------------------
Enter option (valid options with *): U


GENERATING DATA plus ORDINATION AXES TESTS:
By V.Pillar. Version 08Mar2017.

Wish to specify random number initializer (s) or let it be automatic (a)? s/a a
The random number generator is a 'mt19937' generator
Read input from file (f) or from the keyboard (k)? f/k k


GENERATING DATA plus ORDINATION:
Version 08Mar2017
File name to store output: #6_Noise-0.02.txt
Number of sampling units in the sample: 30
Number of sampling units in the population: 1000
Number of variables (p): 9
Do you wish to specify the correlation matrix (c) or the eigenvalues (e)? c/e e
Number of eigenvalues larger than zero (rank r ,â§ p): 5
Enter the variances' weights along each one of the 5 dimensions:
0.3594412630.2212620510.1665877460.1362027690.116506171
Add noise to the generated sample? y/n y
Enter the noise variance (noise will be normally distributed with zero mean and this
specified variance): 0.02

Random numbers to be drawn from the following distribution:
  (1) Uniform distribution
  (2) Normal distribution
  (5) Lognormal distribution
  (7) Exponential distribution
  (8) Cubed exponential distribution
2


--------------------------------------------------------------------------------
```

```
DATA TRANSFORMATION
Options:
      (0)restore original data (no transformation)
      (1)specify transformations on original data
              Enter option: 0
--------------------------------------------------------------------------------
RESEMBLANCE MEASURES
Options:
      (1)comparing sampling units
      (2)comparing variables
              Enter option: 2

Type of resemblance measure (2)comparing variables:
      (1)product moment
      (2)absolute value function
      (3)Euclidean distance
      (4)mutual information
      (17)correlation
      (18)chord distance
      (21)second eigenvalue of pairwise correlation
      (22)covariance
              Enter option: 17
Ordination method:
      (1)principal coordinates analysis
      (2)principal components analysis
      (3)correspondence analysis
              Enter option: 2

Test criteria options:
      (1) Bootstrapped ordination (Pillar 1999)
      (2) RVDIM (Dray)
      (3) Eigenvalue (ter Braak 1988)
      (4) Pseudo F ratio (ter Braak 1988)
      (5) Average random eigenvalue (Peres-Neto et al. 2005)
      (6) Kaiser-Guttman (> 1)
      (7) Kaiser-Guttman (> 0.7)
      (8) Cangelosi & Goriely (2007) information measure
      (9) Cangelosi & Goriely (2007) information measure - 1
      (10) Bartlett's test
      (11) Broken Stick

Specify the test criteria to be considered (0/1):
Test criteria:
1 2 3 4 5 6 7 8 9 10 11
1 1 1 1 1 1 1 1 0  1 1
Number of tests (number of data sets generated): 1000
Specify the alpha probability for the tests: 0.05
Save intermediate results obtained for each generated data set? y/n n
Enter the maximum number of ordination principal axes to be monitored: 9
Initial sample size: 30
Enter number of iterations in resampling/permutation: 20
Save intermediate results? y/n  n
```