A Multivariate Randomization Test of Association Applied to Cognitive Test Results

Albert Ahumada and Bettina Beard
NASA Ames Research Center

Abstract

Randomization tests provide a conceptually simple, distribution-free way to implement significance testing. We have applied this method to the problem of evaluating the significance of the association among a number (k) of variables. The randomization method was the random re-ordering of k-1 of the variables. The criterion variable was the value of the largest eigenvalue of the correlation matrix.

Introduction

The experimental data for which the randomization test was devised were collected to measure possible changes in cognitive abilities and in ratings of cognitive test difficulty following a simulated space ascent in a vibration-augmented centrifuge (Adelstein, et al., 2009). The simulated space ascent study was done to evaluate the effects of G-load and vibration on display legibility (Beard, et al., 2009). The cognitive study was “piggy-backed” onto the legibility study. Each of 11 participants was given a 5 test battery before and after the "ride" were also asked to rate the difficulty/unpleasantness of each test on 5 dimensions. The problem was then to assess the statistical significance of the correlations among the ratings and the tests.

Multiple Correlation Correction

One approach would be to look at the largest absolute correlation and correct for the number of correlations being considered. A strict Bonferroni correction for *n* multiple significance tests at joint level α is α/n for each single test (Benjamini & Hochberg, 1995) . For a k = 5 test battery, there are k(k-1)/2 = 10 correlations to test. The simulation of Appendix 2 showed that the correction provided an accurate estimate of the significance of the maximum correlation for correlation matrices similar in size to ours constructed from independent Gaussian variables. However, the maximum correlation is relatively powerless against the alternative that the correlation is caused by a single common factor (Malevergne & Sornettea, 2004).

Largest Eigenvalue

Anderson (1958) shows that when all the correlations have the same value, the optimal test for whether there is a correlation is based on the largest eigenvalue or principal component of the correlation matrix. However, the distribution of the largest principal component is usually derived for asymptotic conditions (Johnstone, 2001) and for the case in which it comes from a covariance matrix rather than a correlation matrix (Harris, 2001).

The Randomization Test

Randomization tests have the advantages that they are distribution free and that the criterion statistic can be chosen to improve power against desired alternative hypotheses (Fisher, 1935; Siegel, 1956; Edgington, 1969). Given that our ratings are not normally distributed and the simplicity of using a randomization test, we developed the test shown in Appendix 1. It also serves as a randomization test for bi-variate correlation. Anderson (1958) shows that a likelihood ratio statistic for the hypothesis of independence against the general hypothesis is the product of the eigenvalues (i. e., the determinant of the covariance matrix). The function of Appendix 1 is easily modified to the use the product of 1 to k eigenvalues as the criterion statistic for the randomization test.

References

Adelstein, B. D., Beutter, B. R., Kaiser, M. K., McCann, R. S., Stone, L. S., Anderson, M. R., Renema, F., Paloski, W. H. (2009) Influence of Combined Whole-Body Vibration Plus G-Loading on Visual Performance, NASA/TM—2009-215386.

Anderson, T. W. (1958) An introduction to multivariate statistical analysis. Wiley and Sons, New York.

Beard, B. L., Holbrook, J., Geven-Sowul, C., Ahumada, A. J. Jr., (2009) Assessment of Cognitive Abilities in Simulated Space Ascent Environments, International Conference on Environmental Systems, Savannah, GA, 09ICES-0261.

Benjamini Y. & Hochberg Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society B, 57,1, 289-300.

Edgington, E. S. (1969) Statistical inference: The distribution-free approach. New York: McGraw-Hill Book Co.

Fisher, R. A. (1935) Mathematics of a Lady Tasting Tea, in Newman, J. R. (1956) The World of Mathematics, Vol. III, Part VIII, Statistics and the Design of Experiments (New York: Simon & Schuster, 1956), pp. 1514-1521.

Harris, R. J. (2001) A primer of multivariate statistics, Third Edition, Lawrence Erlbaum Associates, Mahwah, New Jersey [see page 352 paragraph 1].

Johnstone, I. M. (2001) On the distribution of the largest eigenvalue in principal component analysis, The Annals of Statistics 29 (2), 295-327 [see abstract].

Malevergne, Y. , Sornettea, D. (2004) Collective origin of the coexistence of random matrix theory noise and of factors in large sample correlation matrices, Physica A 331, 660 – 668.

Siegel, S. (1956) Nonparametric statistics. New York: McGraw-Hill Book Co.

Acknowledgments

Supported by the NASA Space Human Factors Engineering Project. Cynthia Null assisted with the literature search.

Appendix 1

A Matlab implementation of the randomization test.

```
function [pr, s, u, c] = randeig1cor(x, nr, sd, mn)
% largest eigenvalue statistic randomization test
%  based on the covariance matrix if mn =1
%  based on the correlation matrix if mn = sd = 1
% input x: N subjects by k scores matrix
% input nr: is the number of randomizations
% input sd: flag to normalize by standard deviations, 1=yes
% input mn: flag to normalize by means, 1=yes
% output pr: the proportion of randomizations whose first eigenvalue
%            exceeds or equals that of the sample
% output s: the sample eigenvalues
% output u: the sample eigenvectors
% output c: the sample covariance matrix
[N k]= size(x);
y = x;
if mn
    y = y - repmat(mean(y),[N 1]) ;
end
if sd
    y = y ./ repmat(std(y),[N 1]) ;
```

```
end
c = (y'*y)/(N-1) ;
[u s] = svd(c);
s1 = s(1);
pr = 0;
for ir = 1:nr
    for  ik = 2:k
        y(:,ik) = y(randperm(N),ik) ;
    end
    [ui si] = svd((y'*y)/(N-1));
    pr = pr + (si(1)>= s1) ;
end
pr = pr/nr;
% % test data
% x = [ [1 2  7  5  6 10]; [5 6 10 11 12 13]]' ;
% nr = 1000;
% % [pr s1 ];%  0.0170    1.9139
```

Appendix 2

A Matlab simulation of the Bonferroni method for adjusting the significance level to test the significance of the largest correlation in a correlation matrix.

```
% Test Bonferroni method for maximum correlation
N = 32; % so t df (N-2) will be in usual t table
N2 = N-2;
k = 5; % # of correlations is 5*4/2 = 10
p = [0.05 0.01] ; % significance levels
pb = p/(k*(k-1)/2) ; % Bonferroni level
tb = [3.029798 3.645959]; % tinv(pb,30) from Open Office Calc
% http://www.socr.ucla.edu/Applets.dir/T-table.html
% t = c/sqrt((1-c^2)/N2) => c = t/sqrt(t^2+N2)
cb = tb./sqrt(tb.^2+N2); % [0.4840 0.5541] Bonferroni criteria
nrep = 400000 ;
for ip = 1:2
 samp = zeros(nrep,1);
 for irep = 1:nrep
  c = corrcoef(randn(N,k))-diag(ones(k,1));
  samp(irep) = max(abs(c(:)));
 end
 psamp(ip) = sum(samp>cb(ip)) %  975 1030 1035 (three runs)
 samp = sort(samp);
 cbsamp(ip) = 0.5*(samp(nrep*(1-p(ip)))+samp(nrep*(1-p(ip))+1));
end
% Estimated appropriate test criteria
cbsamp  %  0.4823    0.5537
% Estimated actual significances
psamp/nrep  %  0.0482    0.0099
```