Displays for Statistics 5401/8401

Lecture 31

November 18, 2005

Christopher Bingham, Instructor

612-625-1024, kb@umn.edu
372 Ford Hall

Class Web Page

http://www.stat.umn.edu/~kb/classes/5401

Estimating Factor Scores (continued)

Factor scores f_i are not directly *observable*, but *can* be estimated.

Slightly modified notation:

The vector of factor scores for case i is

$$\mathbf{f}_i = [f_{i1}, f_{i2}, \ldots, f_{im}]', \ i = 1,\ldots,N.$$

The (unobservable) N by m matrix of factor scores for all N cases is

$$\mathbf{F} = \begin{bmatrix} \mathbf{f}_1' \\ \cdots \\ \mathbf{f}_N' \end{bmatrix}, \quad \mathbf{f}_i' \text{ is row i of } \mathbf{F}, \ i=1,\ldots,N.$$

The factor analysis model for case i is

$$\mathbf{x}_i = \boldsymbol{\mu} + \mathbf{L}\mathbf{f}_i + \boldsymbol{\varepsilon}_i, \ \mathbf{L} = [\boldsymbol{\ell}_1,\ldots,\boldsymbol{\ell}_m], \ i = 1,\ldots,N$$

$$V[\boldsymbol{\varepsilon}_i] = \boldsymbol{\Psi} = \text{diag}[\psi_1,\psi_2,\ldots,\psi_p]$$

The full data matrix is

$$\underset{N\times p}{\mathbf{X}} = [\mathbf{x}_1, \ldots, \mathbf{x}_N]' = \underset{N\times p}{\mathbf{1}_N\boldsymbol{\mu}'} + \underset{N\times m}{\mathbf{F}}\ \underset{m\times p}{\mathbf{L}'} + \underset{N\times p}{\boldsymbol{\varepsilon}}$$

Estimates of $\mathbf{f}_i$ and $\mathbf{F}$ are notated $\hat{\mathbf{f}}_i$ and $\hat{\mathbf{F}}$.

For **Principal Components (PC) "factor analysis"**, factors *are* observable when parameters are known since

$$\mathbf{f} = [z_1/\sqrt{\lambda_1}, z_2/\sqrt{\lambda_2}, \ldots, z_m/\sqrt{\lambda_m}]',$$

where $z_j = \mathbf{v}_j'(\mathbf{x} - \boldsymbol{\mu})$, j = 1,...,m, are principal components.

Here λ_j and $\mathbf{v}_j$ are eigenvalue and eigenvector of $\boldsymbol{\Sigma}$ or $\boldsymbol{\rho}$. For correlation PC's, replace $\mathbf{x} - \boldsymbol{\mu}$ by $\tilde{\mathbf{x}}$, with $\tilde{x}_k = (x_k-\mu_k)/\sqrt{\sigma_{kk}}$.

You estimate $\mathbf{f}_i$ by

$$\hat{\mathbf{f}}_i = [\hat{z}_{i1}/\sqrt{\hat\lambda_1}, z_{i2}/\sqrt{\hat\lambda_2}, \ldots, \hat{z}_{im}/\sqrt{\hat\lambda_m}]',$$

where $\hat{z}_{ij} = \hat{\mathbf{v}}_j'(\mathbf{x}_i - \overline{\mathbf{x}})$ or $\hat{z}_{ij} = \hat{\mathbf{v}}_j'\tilde{\mathbf{x}}_i$, $\tilde{x}_{ki} = (x_{ki} - \overline{x}_k)/\sqrt{s_{kk}}$.

The estimated matrix of factor scores is

$$\hat{\mathbf{F}} = \tilde{\mathbf{X}}\,[\hat\lambda_1^{-1/2}\hat{\mathbf{v}}_1, \hat\lambda_2^{-1/2}\hat{\mathbf{v}}_2, \ldots, \hat\lambda_m^{-1/2}\hat{\mathbf{v}}_m]$$

where $\tilde{\mathbf{X}} = \mathbf{X} - \mathbf{1}_N\overline{\mathbf{x}}'$.

These are unrotated scores.

For PC-based factor analysis, the estimated loading matrix is

$$\hat{\mathbf{L}} = [\sqrt{\hat\lambda_1}\hat{\mathbf{v}}_1, \sqrt{\hat\lambda_2}\hat{\mathbf{v}}_2, \ldots, \sqrt{\hat\lambda_m}\hat{\mathbf{v}}_m]$$

Then $\hat{\mathbf{F}} = \tilde{\mathbf{X}}\hat{\mathbf{L}}\hat{\boldsymbol{\Lambda}}_m^{-1}$ where $\hat{\boldsymbol{\Lambda}}_m = \text{diag}[\hat\lambda_1, \hat\lambda_2, \ldots, \hat\lambda_m] = \hat{\mathbf{L}}'\hat{\mathbf{L}}$ because the eigenvectors $\hat{\mathbf{v}}_1, \ldots, \hat{\mathbf{v}}_m$ are orthonormal. Thus $\hat{\mathbf{F}} = \tilde{\mathbf{X}}\hat{\mathbf{L}}(\hat{\mathbf{L}}'\hat{\mathbf{L}})^{-1}$.

When $\hat{\mathbf{L}}_{rot} = \hat{\mathbf{L}}\mathbf{H}$, where $\mathbf{H}'\mathbf{H} = \mathbf{I}_m$, are orthogonally rotated loadings, then $\hat{\mathbf{L}} = \hat{\mathbf{L}}_{rot}\mathbf{H}'$. The rotated estimated factors matrix is

$$\begin{aligned}\hat{\mathbf{F}}_{rot} &= \hat{\mathbf{F}}\mathbf{H} = \tilde{\mathbf{X}}\hat{\mathbf{L}}(\hat{\mathbf{L}}'\hat{\mathbf{L}})^{-1}\mathbf{H} \\ &= \tilde{\mathbf{X}}\hat{\mathbf{L}}_{rot}\mathbf{H}'(\mathbf{H}\hat{\mathbf{L}}_{rot}'\hat{\mathbf{L}}_{rot}\mathbf{H}')^{-1}\mathbf{H} = \tilde{\mathbf{X}}\hat{\mathbf{L}}_{rot}(\hat{\mathbf{L}}_{rot}'\hat{\mathbf{L}}_{rot})^{-1}\end{aligned}$$

In general, estimated factors from PC-based factor analysis are $\tilde{\mathbf{X}}\hat{\boldsymbol{\beta}}_{pc}$, where $\hat{\boldsymbol{\beta}}_{pc} = \hat{\mathbf{L}}(\hat{\mathbf{L}}'\hat{\mathbf{L}})^{-1}$, $\hat{\mathbf{L}}$ = estimated loading matrix

Continuing with the artificial data set:

```
Cmd> eigs <- eigen(r); eigs$values
(1)      2.9773     0.81302     0.65535     0.29061     0.26371

Cmd> Lhat_pc <- eigs$vectors[,run(m)] *\
           sqrt(eigs$values[run(m)]')

Cmd> Lhat_pc # unrotated loading matrix
             (1)         (2)
Y1      0.65674     0.23525
Y2      0.55496     0.76752
Y3     -0.88329     0.16769
Y4     -0.86356     0.17873
Y5      0.84385    -0.32942

Cmd> Lhat_pc' %*% Lhat_pc # diagonal matrix of m eigenvalues
              (1)         (2)
(1)       2.9773 -8.1715e-17
(2) -8.1715e-17     0.81302

Cmd> scores_pc <- \
     standardize(y) %*% Lhat_pc %*% solve(Lhat_pc' %*% Lhat_pc)

Cmd> head(scores_pc,3) # unrotated estimated factor scores
              (1)         (2)
(1)     -0.71819    -0.67613
(2)     -0.83434    -0.92947
(3)     -0.82209      1.2269

Cmd> Lhat_pc_rot <- \
        rotation(Lhat_pc,method:"quartimax",kaiser:T)

Cmd> Lhat_pc_rot # rotated factor loadings
             (1)         (2)
Y1      0.56238     0.41277
Y2      0.31304     0.89391
Y3     -0.89444   -0.091162
Y4     -0.87867   -0.074957
Y5      0.90275   -0.075098

Cmd> scores_pcrot <- standardize(y) %*% Lhat_pc_rot %*% \
          solve(Lhat_pc_rot' %*% Lhat_pc_rot)

Cmd> head(scores_pcrot,3) # rotated estimated factor scores
              (1)         (2)
(1)     -0.49556    -0.85286
(2)     -0.53463     -1.1288
(3)      -1.1378     0.94151
```


Regression Method for estimating f

This estimates $\mathbf{f}$ as the conditional expectation $E[\mathbf{f} \mid \mathbf{x}]$ of $\mathbf{f}$ given $\mathbf{x}$.

Because

$$\mathbf{x} = [x_1, \ldots, x_p]' = \boldsymbol{\mu} + \mathbf{L}\mathbf{f} + [\varepsilon_1,\ldots,\varepsilon_p]',$$

when $V[\mathbf{f}] = \mathbf{I}_m$ (orthogonal factors),

- $\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}' + \boldsymbol{\Psi}$
- the joint variance matrix of $\mathbf{x}$ and $\mathbf{f}$ is

$$\underset{(p+m)\times(p+m)}{V\begin{bmatrix}\mathbf{x}\\ \mathbf{f}\end{bmatrix}} = \begin{bmatrix}\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}'+\boldsymbol{\Psi} & \mathbf{L}\\ \mathbf{L}' & \mathbf{I}_m\end{bmatrix}, \quad \begin{matrix}p\\ m\end{matrix}$$
$$\qquad\qquad\qquad\qquad p \qquad\qquad\quad m$$

When $\mathbf{x}$ and $\mathbf{f}$ are jointly <u>multivariate normal</u>, the <u>conditional expectation</u> is

$$E[\mathbf{f} \mid \mathbf{x}] = \boldsymbol{\beta}_{reg}'(\mathbf{x} - \boldsymbol{\mu}), \text{ with}$$

$$\underset{p\times m}{\boldsymbol{\beta}_{reg}} = \boldsymbol{\Sigma}^{-1}\mathrm{Cov}[\mathbf{x}, \mathbf{f}] = (\mathbf{L}\mathbf{L}' + \boldsymbol{\Psi})^{-1}\mathbf{L}$$

Then $\hat{\mathbf{f}} \equiv \boldsymbol{\beta}_{reg}'(\mathbf{x} - \boldsymbol{\mu}) = \mathbf{L}'(\mathbf{L}\mathbf{L}' + \boldsymbol{\Psi})^{-1}(\mathbf{x} - \boldsymbol{\mu})$ satisfies $E[\hat{\mathbf{f}} - \mathbf{f} \mid \mathbf{x}] = \mathbf{0}$ and $V[\hat{\mathbf{f}} - \mathbf{f} \mid \mathbf{x}]$ is as small a possible.

$\boldsymbol{\beta}_{reg}$ is the matrix of coefficients for the multivariate linear regression of $\mathbf{f}$ on $\mathbf{x}$.

The error in estimating $\mathbf{f}$,

$$\mathbf{f} - \hat{\mathbf{f}}_{reg} \equiv \mathbf{f} - \boldsymbol{\beta}_{reg}'(\mathbf{x} - \boldsymbol{\mu}) \neq \mathbf{0}.$$

will *not* be $\mathbf{0}$ even when $\boldsymbol{\beta}$ and $\boldsymbol{\mu}$ are known *exactly*. This is what is meant by $\mathbf{f}$ being "unobservable".

A "plug in" estimate for $\boldsymbol{\beta}_{reg}$ is

$$\hat{\boldsymbol{\beta}}_{reg} = \hat{\boldsymbol{\Sigma}}^{-1}\hat{\mathbf{L}} = (\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\boldsymbol{\Psi}})^{-1}\hat{\mathbf{L}}.$$

The matrix of estimated factor scores is

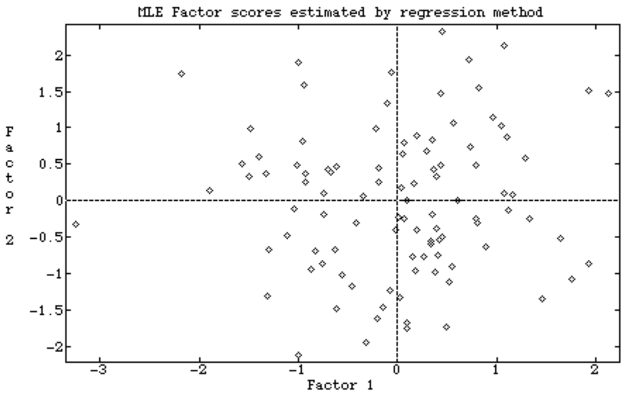
$$\underset{N\times m}{\hat{\mathbf{F}}_{reg}} = \tilde{\mathbf{X}}\hat{\boldsymbol{\beta}}_{reg} = \tilde{\mathbf{X}}\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{L}'$$

$$= \tilde{\mathbf{X}}(\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\boldsymbol{\Psi}})^{-1}\mathbf{L}', \quad \tilde{\mathbf{X}} = \mathbf{X} - \mathbf{1}_N\overline{\mathbf{x}}'$$

Because $\overline{\mathbf{x}}$ is subtracted from each row of $\mathbf{X}$, the sample mean $\overline{\mathbf{f}}$ of the estimated scores is $\mathbf{0}$.


```
Cmd> facanal(r,m,method:"mle",rotation:"quartimax")
Convergence in 26 iterations by criterion 2
estimated uniquenesses:
           Y1           Y2           Y3           Y4           Y5
      0.72178   1.5031e-06       0.2457      0.30368      0.29364
quartimax rotated estimated loadings:
      Factor 1    Factor 2
Y1     0.51052      0.1326
Y2     0.39154     0.92016
Y3    -0.86747   -0.042402
Y4     -0.8327   -0.054116
Y5     0.83799   -0.064357
minimized mle criterion:
(1)   0.0035949

Cmd> rhohat_mle <- LOADINGS %*% LOADINGS' + dmat(PSI)

Cmd> betahat_reg <- solve(rhohat_mle, LOADINGS);betahat_reg
      Factor 1    Factor 2
Y1    0.069041   -0.029377      Coefficients to compute
Y2    0.028727      1.0745      estimated rotated factor
Y3    -0.37938     0.16143      scores
Y4     -0.2926      0.1245
Y5     0.32341    -0.13761

Cmd> scores_reg <- standardize(y) %*% betahat_reg

Cmd> list(scores_reg)
scores          REAL    100    2     (labels)

Cmd> scores[run(10),] # estimated rotated scores for cases 1-10
(1)      -0.56509     -1.0167
(2)      -0.42147    -0.30271
(3)      -0.96222     0.82245
(4)       -2.1753      1.7447
(5)        1.3287    -0.23539
(6)      0.021364     -1.3216
(7)        1.0515       1.036
(8)      -0.10595      1.3347
(9)       -0.6242    -0.65937
(10)       1.9332      1.5097
```

```
Cmd> plot(scores_reg[,1],scores_reg[,2],symbols:"\11",\
     xlab:"Factor 1", ylab:"Factor 2",\
     title:"MLE Factor scores estimated by regression method")
```



Because the sample correlation matrix $\mathbf{R}$ is another estimate for $\boldsymbol{\rho}$, an alternate estimate for $\boldsymbol{\beta}_{reg} = \boldsymbol{\rho}^{-1}\mathbf{L}$ is

$$\tilde{\boldsymbol{\beta}}_{reg} \equiv \mathbf{R}^{-1}\hat{\mathbf{L}}$$

using the unrestricted estimate $\mathbf{R}$ for $\boldsymbol{\rho}$ instead of the factor analytic estimate $\hat{\boldsymbol{\rho}} = \hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\boldsymbol{\Psi}}$.

When $\hat{\boldsymbol{\Psi}}$ and $\hat{\mathbf{L}}$ are fully converged *maximum likelihood* (ML) estimates,

$$\underset{p\times m}{\tilde{\boldsymbol{\beta}}_{reg}} = \underset{p\times p}{\mathbf{R}^{-1}}\underset{p\times m}{\hat{\mathbf{L}}} = \hat{\boldsymbol{\beta}}_{reg} = \hat{\boldsymbol{\rho}}^{-1}\hat{\mathbf{L}} = (\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\boldsymbol{\Psi}})^{-1}\hat{\mathbf{L}}$$

so that $\tilde{\mathbf{F}} = \hat{\mathbf{F}}_{reg}$.

```
Cmd> solve(r,LOADINGS)
       Factor 1    Factor 2
Y1    0.069041   -0.029377
Y2    0.028727      1.0745
Y3    -0.37938     0.16143      Same as before
Y4     -0.2926      0.1245
Y5     0.32341    -0.13761
```


Weighted least squares method

This estimates vectors $\hat{\mathbf{f}}$ of factor scores in such a way that the vector $\hat{\boldsymbol{\varepsilon}} = (\mathbf{x} - \overline{\mathbf{x}}) - \mathbf{L}\hat{\mathbf{f}}$ of estimated *unique* factor scores is as small as possible. This may make sense in a context where the unique factors ε_i are considered as errors.

What is minimized is a *weighted* sum of squares of estimated unique factor scores, with weights for the i^{th} unique factor score proportional to $\hat{\psi}_i^{-1}$.

The solution is *weighted least squares* estimated coefficients

$$\hat{\boldsymbol{\beta}}_{LS} = \hat{\boldsymbol{\Psi}}^{-1}\hat{\mathbf{L}}\hat{\boldsymbol{\Delta}}^{-1} = \hat{\boldsymbol{\beta}}_{reg}(\mathbf{I}_m + \hat{\boldsymbol{\Delta}}^{-1})$$

$$\hat{\boldsymbol{\Delta}} = \hat{\mathbf{L}}'\hat{\boldsymbol{\Psi}}^{-1}\hat{\mathbf{L}}$$

$$\hat{\mathbf{f}}_{LS} = (\mathbf{x} - \overline{\mathbf{x}})\hat{\boldsymbol{\beta}}_{LS}$$

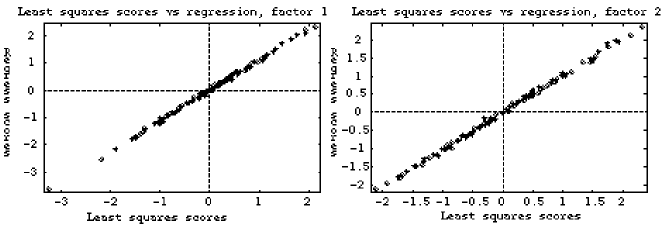
When all $\hat{\psi}_i$ are small, $\hat{\boldsymbol{\Delta}}$ is large, $\hat{\boldsymbol{\Delta}}^{-1}$ is small and $\hat{\boldsymbol{\beta}}_{LS} \approx \hat{\boldsymbol{\beta}}_{reg}$ so that both types of factor scores are essentially the same.


```
Cmd> deltahat <- LOADINGS' %*% dmat(1/PSI) %*% LOADINGS

Cmd> solve(delta)
         Factor 1  Factor 2
Factor 1 0.003507 -0.005517
Factor 2 -0.005517  0.009226

Cmd> betahat_ls <- betahat_reg %*% (dmat(2,1)+solve(deltahat))

Cmd> betahat_ls # coeffs for computing LS factor estimates
       Factor 1    Factor 2
Y1    0.079322   -0.033752
Y2   -0.025299      1.0975
Y3    -0.43588     0.18547
Y4    -0.33617     0.14304
Y5     0.37157    -0.15811

Cmd> scores_ls <- standardize(y) %*% betahat_ls

Cmd> head(scores_ls[run(10),],5) # Weighted LS scores
          Factor 1    Factor 2
(1)        -0.5818     -1.0096
(2)       -0.45837    -0.28701
(3)        -1.1277     0.89285
(4)        -2.5432      1.9012
(5)         1.5088    -0.31204
```

These are almost the same as the regression matrix scores.

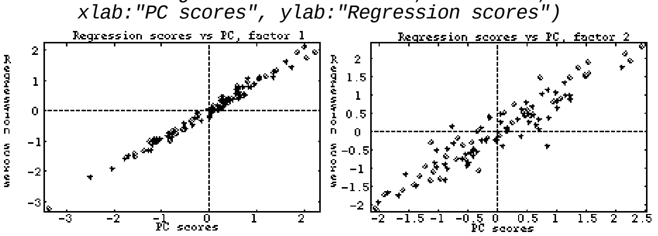
You can see how similar the scores are by plotting Regression scores vs least squares scores.

```
Cmd> plot(scores[,1],scores_ls[,1],symbols:"\1",\
     title:"Least squares scores vs regression, factor 1",\
     xlab:"Least squares scores", ylab:"Regress scores")

Cmd> plot(scores[,2],scores_ls[,2],symbols:"\1",\
     title:"Least squares scores vs regression, factor 2",\
     xlab:"Least squares scores", ylab:"Regress scores")
```



Regression vs PC scores (rotated):

```
Cmd> plot(scores_pcrot[,1],scores_regs[,1],symbols:"\1",\
        title:"Regression scores vs PC, factor 1",\
        xlab:"PC scores", ylab:"Regression scores")

Cmd> plot(scores_pcrot[,2],scores_regs[,2],symbols:"\1",\
        title:"Regression scores vs PC, factor 2",\
        xlab:"PC scores", ylab:"Regression scores")
```



Correlation between two sets of variables

Suppose

$$\mathbf{x}^{(1)} = [x_1^{(1)},\ldots,x_p^{(1)}]' \text{ and } \mathbf{x}^{(2)} = [x_1^{(2)},\ldots,x_q^{(2)}]'$$

are two sets of measurements on the *same* subject or case.

Typically $\mathbf{x}^{(1)}$ and $\mathbf{x}^{(2)}$ each represent a natural grouping of variables.

- $\mathbf{x}^{(1)}$ might consist of *demographic* variables while $\mathbf{x}^{(2)}$ consists of results of *medical tests*.

Because $\mathbf{x}^{(1)}$ and $\mathbf{x}^{(2)}$ are variables associated with the *same* subject, you must presume that they are correlated.

Q How do you test the hypothesis

$$H_0\text{: } \mathbf{x}^{(1)} \text{ and } \mathbf{x}^{(2)} \text{ are uncorrelated?}$$

Q How should you describe any association between $\mathbf{x}^{(1)}$ and $\mathbf{x}^{(2)}$?

Combine $\mathbf{x}^{(1)}$ and $\mathbf{x}^{(2)}$ in a single vector

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}^{(1)} \\ \mathbf{x}^{(2)} \end{bmatrix} \begin{matrix} p \\ q \end{matrix}$$

, a length p + q multivariate observation with p+q by p+q variance matrix $\mathbf{\Sigma} = V[\mathbf{x}]$ and correlation matrix $\boldsymbol{\rho} = \text{Corr}[\mathbf{x}]$.

Partition $\mathbf{\Sigma}$ and $\boldsymbol{\rho}$ in the natural way.

$$\mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{bmatrix}\begin{matrix} p \\ q \end{matrix}, \quad \boldsymbol{\rho} = \begin{bmatrix} \boldsymbol{\rho}_{11} & \boldsymbol{\rho}_{12} \\ \boldsymbol{\rho}_{21} & \boldsymbol{\rho}_{22} \end{bmatrix}\begin{matrix} p \\ q \end{matrix},$$

$$\begin{matrix} p & q \end{matrix} \qquad\qquad \begin{matrix} p & q \end{matrix}$$

- $\mathbf{\Sigma}_{11} = V[\mathbf{x}^{(1)}] = [\sigma_{ij}^{11}]$ $(p \times p)$
  $\boldsymbol{\rho}_{11} = \text{Corr}[\mathbf{x}^{(1)}] = [\rho_{ij}^{11}]$ $(p \times p)$
- $\mathbf{\Sigma}_{12} = [\sigma_{ij}^{12}] = \mathbf{\Sigma}_{21}'$ $(p \times q)$
  $\boldsymbol{\rho}_{12} = [\rho_{ij}^{12}] = \boldsymbol{\rho}_{21}'$ $(p \times q)$
- $\mathbf{\Sigma}_{22} = V[\mathbf{x}^{(2)}] = [\sigma_{ij}^{22}]$ $(q \times q)$
  $\boldsymbol{\rho}_{22} = \text{Corr }[\mathbf{x}^{(2)}] = [\rho_{ij}^{22}]$ $(q \times q)$.

Notation:

$$\rho_{ij}^{k\ell} \equiv \text{corr}[x_i^{(k)},x_j^{(\ell)}]$$

- $k = 1, 2$ and $\ell = 1, 2$ index the sets of variables
- i and j index variables within a set.

Examples:

$\rho_{22}^{12} = \text{corr}[x_2^{(1)},x_2^{(2)}]$
$\rho_{23}^{11} = \text{corr}[x_2^{(1)},x_3^{(1)}]$
$\rho_{ii}^{11} = 1,\ i = 1,2,\ldots,\ p$
$\rho_{ii}^{22} = 1,\ i = 1,2,\ldots,q$

$\rho_{ii}^{12} = \text{corr}[x_i^{(1)},x_i^{(2)}] \neq 1.$

$\mathbf{x}^{(1)}$ and $\mathbf{x}^{(2)}$ are uncorrelated if and only if all p×q correlations $\rho_{ij}^{12} = 0$, that is if the null hypothesis

$$H_0: \rho_{ij}^{12} = 0,\ i = 1, \ldots, p,\ j = 1, \ldots, q$$

is true.

In terms of matrices,

$$H_0: \mathbf{\Sigma}_{12} = \boldsymbol{\rho}_{12} = \mathbf{0}.$$

When $\mathbf{x}$ is $N_{p+q}(\boldsymbol{\mu}, \mathbf{\Sigma})$, $\boldsymbol{\rho}_{12} = \mathbf{0}$ is equivalent to

$$\widetilde{H}_0: \mathbf{x}^{(1)} \text{ and } \mathbf{x}^{(2)} \text{ are independent}$$

Usually $\widetilde{H}_0$ is the real hypothesis of interest rather than $\boldsymbol{\rho}_{12} = \mathbf{0}$, but it's almost impossible to test without assuming multivariate normality.

There are other ways to state H_0:

H_0 in terms of regression coefficients

- $\boldsymbol{\beta}_{2.1} \equiv \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}$ = q by p matrix of (true) multivariate regression coefficients of $\mathbf{x}^{(2)}$ on $\mathbf{x}^{(1)}$ ($E[\mathbf{x}^{(2)} | \mathbf{x}^{(1)}] = \boldsymbol{\mu}_2 + \boldsymbol{\beta}_{2.1}'(\mathbf{x}^{(1)} - \boldsymbol{\mu}_1)$)
- $\boldsymbol{\beta}_{1.2} \equiv \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}$ = p by q matrix of (true) multivariate regression coefficients of $\mathbf{x}^{(1)}$ on $\mathbf{x}^{(2)}$ ($E[\mathbf{x}^{(1)} | \mathbf{x}^{(2)}] = \boldsymbol{\mu}_1 + \boldsymbol{\beta}_{1.2}'(\mathbf{x}^{(2)} - \boldsymbol{\mu}_2)$)

  $H_0: \boldsymbol{\rho}_{12} = \mathbf{0}$ is equivalent to either of $H_0: \boldsymbol{\beta}_{2.1} = \mathbf{0}$ or $H_0: \boldsymbol{\beta}_{1.2} = \mathbf{0}$

$\boldsymbol{\beta}_{1.2}$ and $\boldsymbol{\beta}_{2.1}$ are related by identities:

- $\boldsymbol{\beta}_{1.2} = \mathbf{\Sigma}_{11}\boldsymbol{\beta}_{2.1}'\mathbf{\Sigma}_{22}^{-1}$
- $\boldsymbol{\beta}_{2.1} = \mathbf{\Sigma}_{22}\boldsymbol{\beta}_{1.2}'\mathbf{\Sigma}_{11}^{-1}$

This generalizes the bivariate regression identity (p = q = 1)

$$\beta_{x.y} = (\sigma_x^2/\sigma_y^2)\beta_{y.x}$$

- When you think of $\mathbf{x}^{(2)}$ as depending on $\mathbf{x}^{(1)}$, $\boldsymbol{\beta}_{2.1}$ is often a good way to summarize association between $\mathbf{x}^{(1)}$ and $\mathbf{x}^{(2)}$.
- When you think of $\mathbf{x}^{(1)}$ as depending on $\mathbf{x}^{(2)}$, $\boldsymbol{\beta}_{1.2}$ is often a good way to summarize association between $\mathbf{x}^{(1)}$ and $\mathbf{x}^{(2)}$.

$\boldsymbol{\beta}_{2.1}$ and $\boldsymbol{\beta}_{1.2}$ both treat $\mathbf{x}^{(1)}$ and $\mathbf{x}^{(2)}$ *assymetrically*.

When you think of $\mathbf{x}^{(1)}$ and $\mathbf{x}^{(2)}$ *symmetrically*, then you would usually prefer $\boldsymbol{\rho}_{12}$ to $\boldsymbol{\beta}_{2.1}$ or $\boldsymbol{\beta}_{1.2}$ as a summary of the dependence.

"Symmetric" means that swapping $\mathbf{x}^{(1)}$ and $\mathbf{x}^{(2)}$ will not effect how you view the relationship.

Data: Usually a random sample:

$$\mathbf{x}_i = [\mathbf{x}_i^{(1)\prime}, \mathbf{x}_i^{(2)\prime}]',\ i = 1,\ldots,n,$$

from a p+q dimensional population.

Consequence: *Both* $\mathbf{x}^{(1)}$ and $\mathbf{x}^{(2)}$ are random.

Suppose either $\mathbf{x}^{(1)}$ or $\mathbf{x}^{(2)}$ is *not* random.

- Population correlations between elements of $\mathbf{x}^{(1)}$ and elements of $\mathbf{x}^{(2)}$ *are not defined*.
- $\boldsymbol{\beta}_{1.2}$ ($\mathbf{x}^{(1)}$ random but not $\mathbf{x}^{(2)}$) or $\boldsymbol{\beta}_{2.1}$ ($\mathbf{x}^{(2)}$ random but not $\mathbf{x}^{(1)}$) may be defined.

In either of the following, be suspicious of any correlation-based analysis:

- Values of $\mathbf{x}^{(1)}$ and/or $\mathbf{x}^{(2)}$ are subject to manipulation or control
- Values of $\mathbf{x}^{(1)}$ and/or $\mathbf{x}^{(2)}$ are affected by a data selection procedure.

Either implies the sample is not random.

Tests of H_0: $\rho_{12} = \mathbf{0}$

Bonferronized $\{r_{ij}^{12}\}_{1 \le i \le p, 1 \le j \le q}$

This uses the pq sample correlations

$$r_{ij} \equiv r_{ij}^{12} = \widehat{\text{Corr}}[x_i^{(1)}, x_j^{(2)}] = s_{ij}/\sqrt{\{s_{ii}s_{jj}\}}$$

computed from **S** with f_e d.f.

A standard *bivariate* test statistic of

$$H_0^{(ij)}: \rho_{ij}^{12} = 0$$

is

$$t_{ij} = \sqrt{(f_e-1)}r_{ij}/\sqrt{(1 - r_{ij}^2)}$$

whose null distribution is Student's t on $f_e - 1$ degrees of freedom.

$f_e = n - 1$ for $\hat{\boldsymbol{\rho}}$ from a random sample.

$f_e = n - g$ for $\hat{\boldsymbol{\rho}}$ from a pooled estimate $\mathbf{S} = (n-g)^{-1}\mathbf{E}$ from a MANOVA with g-groups.

You reject $H_0^{(ij)}$ when

$$|t_{ij}| > t_{f_e-1}(\alpha/2)$$

Since you can recover r_{ij} from t_{ij} as

$$r_{ij} = t_{ij}/\sqrt{\{f_e - 1 + t_{ij}^2\}}$$

you can reject $H_0^{(ij)}$ when

$$|r_{ij}| > t_{f_e-1}(\alpha/2)/\sqrt{\{f_e - 1 + t_{f_e-1}(\alpha/2)^2\}}.$$

Assumptions required for Student's t

1. Either $\{x_{i1}^{(1)}, x_{i2}^{(1)}, ..., x_{in}^{(1)}\}$ or $\{x_{j1}^{(2)}, x_{j2}^{(2)}, ..., x_{jn}^{(2)}\}$ or both is a random sample
2. Either $x_i^{(1)}$ or $x_i^{(2)}$ (or both) is univariate normal
3. $x_i^{(1)}$ and $x_i^{(2)}$ are independent,

Under these conditions,

$$t_{ij} = t_{f_e-1} = t_{n-2}, \text{ Student's t}$$

In particular, *Bi*variate normality is *not* required to test *independence*.

When $(x_i^{(1)}, x_j^{(2)})$ is not bivariate normal, $\rho_{ij}^{12} = 0$ is not enough to ensure that t_{ij} is t_{f_e-1}. You need actual independence.

Since there are K = pq t-statistics t_{ij}, one for each $r_{ij}^{(12)}$ in $\mathbf{R}_{12}$ you should Bonferronize them using $K = p \times q$ to test H_0: $\boldsymbol{\rho}_{12} = \mathbf{0}$.:

Reject H_0 when

$$\max_{i,j}|t_{ij}| > t_{f_e-1}((\alpha/(pq))/2)$$

or when $pq \times \min_{i,j} P_{ij} < \alpha$, P_{ij} = two-tail P-value based on t_{ij}

And, for all i and j such that

$$|t_{ij}| > t_{f_e-1}((\alpha/(pq))/2) \text{ or } pq \times P_{ij} < \alpha$$

you can reject $H_0^{(ij)}: \rho_{ij}^{12} = 0$ and declare that $x_i^{(1)}$ and $x_j^{(2)}$ are apparently correlated.