• Additional Sample Descriptive Measures

Matrices of Errors of Approximations

- Since \( \hat{U} = \hat{A}(1) \) and \( \hat{V} = \hat{B}(2) \) we can write
  \[
  \begin{align*}
  \mathbf{x}^{(1)} &= \hat{A}^{-1} \hat{U} \\
  \mathbf{x}^{(2)} &= \hat{B}^{-1} \hat{V}
  \end{align*}
  \]

Because sample \( \text{Cov}(\hat{U}, \hat{V}) = \hat{A} \hat{S}_{12} \hat{B}^\prime \), sample \( \text{Cov}(\hat{U}) = \hat{A} \hat{S}_{11} \hat{A}^\prime = I \), and sample \( \text{Cov}(\hat{V}) = \hat{B} \hat{S}_{22} \hat{B}^\prime = I \), the matrices of error of approximation may be interpreted as descriptive summaries of how well the first \( r \) sample canonical variates reproduce the sample covariance matrices. Patterns of large entries in the rows and/or columns of the error matrices indicate a poor “fit” to the corresponding variables.

If only the first \( r \) canonical pairs are used, so that for instance,

\[
\mathbf{x}^{(1)} = [\hat{a}^{(1)} | \hat{a}^{(2)} | \ldots | \hat{a}^{(r)}]
\]

and

\[
\mathbf{x}^{(2)} = [\hat{b}^{(1)} | \hat{b}^{(2)} | \ldots | \hat{b}^{(r)}]
\]

then \( S_{12} \) is approximated by sample \( \text{Cov}(\hat{x}^{(1)}, \hat{x}^{(2)}) \).

The matrices of error of approximation are

\[
\begin{align*}
S_{11} &= (\hat{a}^{(1)} \hat{a}^{(1)\prime} + \hat{a}^{(2)} \hat{a}^{(2)\prime} + \ldots + \hat{a}^{(r)} \hat{a}^{(r)\prime}) = \hat{a}^{(r+1)} \hat{a}^{(r+1)\prime} + \ldots + \hat{a}^{(p)} \hat{a}^{(p)\prime}, \\
S_{22} &= (\hat{b}^{(1)} \hat{b}^{(1)\prime} + \hat{b}^{(2)} \hat{b}^{(2)\prime} + \ldots + \hat{b}^{(r)} \hat{b}^{(r)\prime}) = \hat{b}^{(r+1)} \hat{b}^{(r+1)\prime} + \ldots + \hat{b}^{(q)} \hat{b}^{(q)\prime}, \\
S_{12} &= (\rho_1 \hat{a}^{(1)} \hat{b}^{(1)\prime} + \rho_2 \hat{a}^{(2)} \hat{b}^{(2)\prime} + \ldots + \rho_r \hat{a}^{(r)} \hat{b}^{(r)\prime})
\end{align*}
\]

The approximation error matrices may be interpreted as descriptive summaries of how well the first \( r \) sample canonical variates reproduce the sample covariance matrices. Patterns of large entries in the rows and/or columns of the error matrices indicate a poor “fit” to the corresponding variables.
ordinarily, the first \( r \) variates do a better job of reproducing the elements of \( S_{12} \) than the elements of \( S_{11} \) or \( S_{22} \) (Q: Why?)

Proportions of Explained Sample Variance

- When the observations are standardized, the sample covariance matrices \( S_{kl} \) are correlation matrices \( R_{kl} \). The canonical coefficient vectors are the rows of the matrices \( \hat{A}_z \) and \( \hat{B}_z \) and the columns of \( \hat{A}_z^{-1} \) and \( \hat{B}_z^{-1} \) are the sample correlations between the canonical variates and their component variables.

- sample \( \text{Cov}(z^{(1)}, \hat{U}) = \text{Cov}(\hat{A}_z^{-1}\hat{U}, \hat{U}) = \hat{A}_z^{-1} \)

- sample \( \text{Cov}(z^{(2)}, \hat{V}) = \text{Cov}(\hat{B}_z^{-1}\hat{V}, \hat{V}) = \hat{B}_z^{-1} \)

so,

\[
\hat{A}_z^{-1} = [\hat{a}_z^{(1)}, \hat{a}_z^{(2)}, \ldots, \hat{a}_z^{(p)}] = \begin{bmatrix}
    r_{\hat{U}_1, \hat{z}_1^{(1)}} & r_{\hat{U}_2, \hat{z}_1^{(1)}} & \cdots & r_{\hat{U}_p, \hat{z}_1^{(1)}} \\
    r_{\hat{U}_1, \hat{z}_1^{(2)}} & r_{\hat{U}_2, \hat{z}_1^{(2)}} & \cdots & r_{\hat{U}_p, \hat{z}_1^{(2)}} \\
    \vdots & \vdots & \ddots & \vdots \\
    r_{\hat{U}_1, \hat{z}_1^{(p)}} & r_{\hat{U}_2, \hat{z}_1^{(p)}} & \cdots & r_{\hat{U}_p, \hat{z}_1^{(p)}} \\
\end{bmatrix}
\]

\[
\hat{B}_z^{-1} = [\hat{b}_z^{(1)}, \hat{b}_z^{(2)}, \ldots, \hat{b}_z^{(q)}] = \begin{bmatrix}
    r_{\hat{V}_1, \hat{z}_1^{(2)}} & r_{\hat{V}_2, \hat{z}_1^{(2)}} & \cdots & r_{\hat{V}_q, \hat{z}_1^{(2)}} \\
    r_{\hat{V}_1, \hat{z}_1^{(3)}} & r_{\hat{V}_2, \hat{z}_1^{(3)}} & \cdots & r_{\hat{V}_q, \hat{z}_1^{(3)}} \\
    \vdots & \vdots & \ddots & \vdots \\
    r_{\hat{V}_1, \hat{z}_1^{(q)}} & r_{\hat{V}_2, \hat{z}_1^{(q)}} & \cdots & r_{\hat{V}_q, \hat{z}_1^{(q)}} \\
\end{bmatrix}
\]

where \( r_{\hat{U}_i, \hat{z}_i^{(j)}} \) and \( r_{\hat{V}_i, \hat{z}_i^{(j)}} \) are the sample correlation coefficients between the quantities with subscripts.

- Total (standardized) sample variance in first set

\[
= \text{tr}(R_{11}) = \text{tr}(\hat{a}_z^{(1)}\hat{a}_z^{(1)\prime} + \hat{a}_z^{(2)}\hat{a}_z^{(2)\prime} + \cdots + \hat{a}_z^{(p)}\hat{a}_z^{(p)\prime}) = p
\]

- Total (standardized) sample variance in second set

\[
= \text{tr}(R_{22}) = \text{tr}(\hat{b}_z^{(1)}\hat{b}_z^{(1)\prime} + \hat{b}_z^{(2)}\hat{b}_z^{(2)\prime} + \cdots + \hat{b}_z^{(q)}\hat{b}_z^{(q)\prime}) = q
\]

- the contribution of the first \( r \) canonical variates to the total sample variance:

\[
\text{tr}(\hat{a}_z^{(1)}\hat{a}_z^{(1)\prime} + \hat{a}_z^{(2)}\hat{a}_z^{(2)\prime} + \cdots + \hat{a}_z^{(r)}\hat{a}_z^{(r)\prime}) = \sum_{i=1}^{r} \sum_{k=1}^{p} r_{\hat{U}_i, \hat{z}_k^{(i)}}
\]

\[
\text{tr}(\hat{b}_z^{(1)}\hat{b}_z^{(1)\prime} + \hat{b}_z^{(2)}\hat{b}_z^{(2)\prime} + \cdots + \hat{b}_z^{(r)}\hat{b}_z^{(r)\prime}) = \sum_{i=1}^{r} \sum_{k=1}^{q} r_{\hat{V}_i, \hat{z}_k^{(i)}}
\]

- proportions of total sample variances explained by 1st \( r \) canonical variates:

\[
R^2_{U} = \frac{\text{tr}(\hat{a}_z^{(1)}\hat{a}_z^{(1)\prime} + \cdots + \hat{a}_z^{(r)}\hat{a}_z^{(r)\prime})}{\text{tr}(R_{11})} = \frac{\sum_{i=1}^{r} \sum_{k=1}^{p} r_{\hat{U}_i, \hat{z}_k^{(i)}}}{p}
\]

\[
R^2_{V} = \frac{\text{tr}(\hat{b}_z^{(1)}\hat{b}_z^{(1)\prime} + \cdots + \hat{b}_z^{(r)}\hat{b}_z^{(r)\prime})}{\text{tr}(R_{22})} = \frac{\sum_{i=1}^{r} \sum_{k=1}^{q} r_{\hat{V}_i, \hat{z}_k^{(i)}}}{q}
\]

- they provide some indication of how well the canonical variates represent their respective sets
• they also provide single-number descriptions of the matrices of errors, because
\[
\frac{1}{p} \text{tr} \left[ \mathbf{R}_{11} - \mathbf{a}_1^{(1)} \mathbf{a}_1^{(1)\top} - \mathbf{a}_2^{(2)} \mathbf{a}_2^{(2)\top} - \cdots - \mathbf{a}_r^{(r)} \mathbf{a}_r^{(r)\top} \right] = 1 - R^2_{0} \hat{v}_1, \hat{v}_2, \ldots, \hat{v}_r,
\]
\[
\frac{1}{q} \text{tr} \left[ \mathbf{R}_{22} - \mathbf{b}_1^{(1)} \mathbf{b}_1^{(1)\top} - \mathbf{b}_2^{(2)} \mathbf{b}_2^{(2)\top} - \cdots - \mathbf{b}_r^{(r)} \mathbf{b}_r^{(r)\top} \right] = 1 - R^2_{0} \hat{v}_1, \hat{v}_2, \ldots, \hat{v}_r,
\]

• Large Sample Inferences

➤ Note: $$\Sigma_{12} = 0 \Rightarrow$$ no point in pursuing a CCA. ➤ Q: how to test $$H_0: \Sigma_{12} = 0$$?

➤ Result 10.3. Let

$$\mathbf{X}_j = \begin{bmatrix} \mathbf{X}_j^{(1)} \\ \mathbf{X}_j^{(2)} \end{bmatrix}, \quad j = 1, 2, \ldots, n$$

be a random sample from an $$N_{p+q}(\mu, \Sigma)$$ population with $$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

Recall: for normal density, maximum likelihood $$\sigma_{\mathbf{X}_j} = \frac{1}{\sqrt{|\Sigma|}}$$.

Then the likelihood ratio test of $$H_0: \Sigma_{12} = 0$$ versus $$H_1: \Sigma_{12} \neq 0$$ rejects $$H_0$$ for large values of

$$-2 \ln \Lambda = n \ln \left( \frac{|\mathbf{S}_{11}| |\mathbf{S}_{22}|}{|\mathbf{S}|} \right) = -n \ln \left( \prod_{i=1}^{pq} (1 - \hat{\rho}_i^2) \right) \sim \chi^2_{pq}$$

maximum likelihood under $$H_0$$

➤ when $$n$$ is large, under $$H_0$$, the likelihood ratio test statistic is approximately distributed as a chi-square random variable with $$pq$$ d.f.

➤ Bartlett (1939) suggests

Reject $$H_0: \Sigma_{12} = 0$$ ($$\rho_1 = \rho_2 = \cdots = \rho_p = 0$$) at significance level $$\alpha$$ if

$$-n (n - 1 - \frac{1}{2} (p + q + 1)) \ln \prod_{i=1}^{p} (1 - \hat{\rho}_i^2) > \chi^2_{pq}(\alpha)$$

➤ Q: What if $$H_0: \Sigma_{12} = 0$$ is rejected? Next step?

• $$H_0^{(k)}: \rho_1 \neq 0, \rho_2 \neq 0, \cdots, \rho_k \neq 0, \rho_{k+1} = \cdots = \rho_p = 0$$

• $$H_1^{(k)}: \rho_i \neq 0$$, for some $$i \geq k + 1$$

• Reject $$H_0^{(k)}$$ at significance level $$\alpha$$ if

$$-n (n - 1 - \frac{1}{2} (p + q + 1)) \ln \prod_{i=1}^{p} (1 - \hat{\rho}_i^2) > \chi^2_{(p-k)(q-k)}(\alpha)$$

• the issue of multiple testing should be taken into consideration

➤ Reading: Reference, 10.4, 10.5, 10.6