- Additional Sample Descriptive Measures

  ➢ Matrices of Errors of Approximations → approximation of cov. matrix.

  - Since $\hat{\mathbf{U}} = \hat{\mathbf{A}}\mathbf{x}^{(1)}$ and $\hat{\mathbf{V}} = \hat{\mathbf{B}}\mathbf{x}^{(2)}$ we can write

$$\underset{(p\times 1)}{\mathbf{x}^{(1)}} = \underset{(p\times p)}{\hat{\mathbf{A}}^{-1}}\ \underset{(p\times 1)}{\hat{\mathbf{U}}} \qquad \underset{(q\times 1)}{\mathbf{x}^{(2)}} = \underset{(q\times q)}{\hat{\mathbf{B}}^{-1}}\ \underset{(q\times 1)}{\hat{\mathbf{V}}}$$

Because sample $\boxed{\text{Cov}(\hat{\mathbf{U}}, \hat{\mathbf{V}})} = \hat{\mathbf{A}}\mathbf{S}_{12}\hat{\mathbf{B}}'$, sample $\text{Cov}(\hat{\mathbf{U}}) = \hat{\mathbf{A}}\mathbf{S}_{11}\hat{\mathbf{A}}' = \underset{(p\times p)}{\mathbf{I}}$, and sample $\text{Cov}(\hat{\mathbf{V}}) = \hat{\mathbf{B}}\mathbf{S}_{22}\hat{\mathbf{B}}' = \underset{(q\times q)}{\mathbf{I}}$, LNp.7-6

1st column of A^{-1} 1st column of B^{-1}

$$\mathbf{S}_{12} = \hat{\mathbf{A}}^{-1}\left[\begin{array}{cccc|c} \widehat{\rho_1^*} & 0 & \cdots & 0 & \\ 0 & \widehat{\rho_2^*} & \cdots & 0 & \\ \vdots & \vdots & \ddots & \vdots & \mathbf{0} \\ 0 & 0 & \cdots & \widehat{\rho_p^*} & \end{array}\right](\hat{\mathbf{B}}^{-1})' = \widehat{\rho_1^*}\hat{\mathbf{a}}^{(1)}\hat{\mathbf{b}}^{(1)\prime} + \widehat{\rho_2^*}\hat{\mathbf{a}}^{(2)}\hat{\mathbf{b}}^{(2)\prime} + \cdots + \widehat{\rho_p^*}\hat{\mathbf{a}}^{(p)}\hat{\mathbf{b}}^{(p)\prime}$$

c.f. Result 2A.14, LNp.2-12

c.f. PCA

$$\mathbf{S}_{11} = (\hat{\mathbf{A}}^{-1})(\hat{\mathbf{A}}^{-1})' = \hat{\mathbf{a}}^{(1)}\hat{\mathbf{a}}^{(1)\prime} + \hat{\mathbf{a}}^{(2)}\hat{\mathbf{a}}^{(2)\prime} + \cdots + \hat{\mathbf{a}}^{(p)}\hat{\mathbf{a}}^{(p)\prime}$$

$$\mathbf{S}_{22} = (\hat{\mathbf{B}}^{-1})(\hat{\mathbf{B}}^{-1})' = \hat{\mathbf{b}}^{(1)}\hat{\mathbf{b}}^{(1)\prime} + \hat{\mathbf{b}}^{(2)}\hat{\mathbf{b}}^{(2)\prime} + \cdots + \hat{\mathbf{b}}^{(q)}\hat{\mathbf{b}}^{(q)\prime}$$

$= \hat{U}_1\hat{a}^{(1)} + \hat{U}_2\hat{a}^{(2)} + \cdots + \hat{U}_r\hat{a}^{(r)} + \cdots + \hat{U}_p\hat{a}^{(p)}$

  - Since $\mathbf{x}^{(1)} = \hat{\mathbf{A}}^{-1}\hat{\mathbf{U}}$ and $\hat{\mathbf{U}}$ has sample covariance $\mathbf{I}$, the first r columns of $\hat{\mathbf{A}}^{-1}$ contain the sample covariances of the first r canonical variates $\hat{U}_1, \hat{U}_2, \ldots, \hat{U}_r$ with their component variables $X_1^{(1)}, X_2^{(1)}, \ldots, X_p^{(1)}$. Similarly, the first r columns of $\hat{\mathbf{B}}^{-1}$ contain the sample covariances of $\hat{V}_1, \hat{V}_2, \ldots, \hat{V}_r$ with their component variables.

  - If only the first r canonical pairs are used, so that for instance,

$$\tilde{\mathbf{x}}^{(1)} = [\hat{\mathbf{a}}^{(1)} \,\vdots\, \hat{\mathbf{a}}^{(2)} \,\vdots\, \cdots \,\vdots\, \hat{\mathbf{a}}^{(r)}]\begin{bmatrix}\hat{U}_1\\ \hat{U}_2\\ \vdots\\ \hat{U}_r\end{bmatrix} \in \mathbb{R}^r$$

in p-dim space

only occupy a r-dim subspace

and

$$\tilde{\mathbf{x}}^{(2)} = [\hat{\mathbf{b}}^{(1)} \,\vdots\, \hat{\mathbf{b}}^{(2)} \,\vdots\, \cdots \,\vdots\, \hat{\mathbf{b}}^{(r)}]\begin{bmatrix}\hat{V}_1\\ \hat{V}_2\\ \vdots\\ \hat{V}_r\end{bmatrix}$$

in q-dim space

but only occupy a r-dim space.

then $\mathbf{S}_{12}$ is approximated by sample $\text{Cov}(\tilde{\mathbf{x}}^{(1)}, \tilde{\mathbf{x}}^{(2)})$

$\rightarrow [\hat{a}^{(1)} | \cdots | \hat{a}^{(r)}] \begin{bmatrix} \hat{\rho}_1^* & & 0 \\ & \ddots & \\ 0 & & \hat{\rho}_r^* \end{bmatrix} \begin{bmatrix} \hat{b}^{(1)\prime} \\ \hat{b}^{(2)\prime} \\ \vdots \\ \hat{b}^{(r)\prime} \end{bmatrix}$

  - the matrices of error of approximation are

$$\mathbf{S}_{11} - (\hat{\mathbf{a}}^{(1)}\hat{\mathbf{a}}^{(1)\prime} + \hat{\mathbf{a}}^{(2)}\hat{\mathbf{a}}^{(2)\prime} + \cdots + \hat{\mathbf{a}}^{(r)}\hat{\mathbf{a}}^{(r)\prime}) = \hat{\mathbf{a}}^{(r+1)}\hat{\mathbf{a}}^{(r+1)\prime} + \cdots + \hat{\mathbf{a}}^{(p)}\hat{\mathbf{a}}^{(p)\prime}$$

$$\mathbf{S}_{22} - (\hat{\mathbf{b}}^{(1)}\hat{\mathbf{b}}^{(1)\prime} + \hat{\mathbf{b}}^{(2)}\hat{\mathbf{b}}^{(2)\prime} + \cdots + \hat{\mathbf{b}}^{(r)}\hat{\mathbf{b}}^{(r)\prime}) = \hat{\mathbf{b}}^{(r+1)}\hat{\mathbf{b}}^{(r+1)\prime} + \cdots + \hat{\mathbf{b}}^{(q)}\hat{\mathbf{b}}^{(q)\prime}$$

$$\mathbf{S}_{12} - (\widehat{\rho_1^*}\hat{\mathbf{a}}^{(1)}\hat{\mathbf{b}}^{(1)\prime} + \widehat{\rho_2^*}\hat{\mathbf{a}}^{(2)}\hat{\mathbf{b}}^{(2)\prime} + \cdots + \widehat{\rho_r^*}\hat{\mathbf{a}}^{(r)}\hat{\mathbf{b}}^{(r)\prime})$$

$$= \widehat{\rho_{r+1}^*}\hat{\mathbf{a}}^{(r+1)}\hat{\mathbf{b}}^{(r+1)\prime} + \cdots + \widehat{\rho_p^*}\hat{\mathbf{a}}^{(p)}\hat{\mathbf{b}}^{(p)\prime}$$

c.f. → residual matrix in factor analysis.

  - 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 $\mathbf{S}_{12}$ than the elements of $\mathbf{S}_{11}$ or $\mathbf{S}_{22}$ (**Q**: Why?) $(\hat{a}_1, \hat{b}_1), (\hat{a}_2, \hat{b}_2), \cdots$ are chosen to maximize cov/cor (i.e. S_{12}). PCs would better reproduce them.

➢ Proportions of Explained Sample Variance

- When the observations are standardized, the sample covariance matrices $\mathbf{S}_{kl}$ are correlation matrices $\mathbf{R}_{kl}$. The canonical coefficient vectors are the *rows* of the matrices $\hat{\mathbf{A}}_\mathbf{z}$ and $\hat{\mathbf{B}}_\mathbf{z}$ and the *columns* of $\hat{\mathbf{A}}_\mathbf{z}^{-1}$ and $\hat{\mathbf{B}}_\mathbf{z}^{-1}$ are the sample correlations between the canonical variates and their component variables.

- $$\text{sample Cov}(\mathbf{z}^{(1)}, \hat{\mathbf{U}}) = \text{sample Cov}(\hat{\mathbf{A}}_\mathbf{z}^{-1}\hat{\mathbf{U}}, \hat{\mathbf{U}}) = \hat{\mathbf{A}}_\mathbf{z}^{-1}$$
  $$\text{sample Cov}(\mathbf{z}^{(2)}, \hat{\mathbf{V}}) = \text{sample Cov}(\hat{\mathbf{B}}_\mathbf{z}^{-1}\hat{\mathbf{V}}, \hat{\mathbf{V}}) = \hat{\mathbf{B}}_\mathbf{z}^{-1}$$

- so,

$$\hat{\mathbf{A}}_\mathbf{z}^{-1} = [\hat{\mathbf{a}}_\mathbf{z}^{(1)}, \hat{\mathbf{a}}_\mathbf{z}^{(2)}, \ldots, \hat{\mathbf{a}}_\mathbf{z}^{(p)}] = \begin{bmatrix} r_{\hat{U}_1, z_1^{(1)}} & r_{\hat{U}_2, z_1^{(1)}} & \cdots & r_{\hat{U}_p, z_1^{(1)}} \\ r_{\hat{U}_1, z_2^{(1)}} & r_{\hat{U}_2, z_2^{(1)}} & \cdots & r_{\hat{U}_p, z_2^{(1)}} \\ \vdots & \vdots & \ddots & \vdots \\ r_{\hat{U}_1, z_p^{(1)}} & r_{\hat{U}_2, z_p^{(1)}} & \cdots & r_{\hat{U}_p, z_p^{(1)}} \end{bmatrix}$$

$$\hat{\mathbf{B}}_\mathbf{z}^{-1} = [\hat{\mathbf{b}}_\mathbf{z}^{(1)}, \hat{\mathbf{b}}_\mathbf{z}^{(2)}, \ldots, \hat{\mathbf{b}}_\mathbf{z}^{(q)}] = \begin{bmatrix} r_{\hat{V}_1, z_1^{(2)}} & r_{\hat{V}_2, z_1^{(2)}} & \cdots & r_{\hat{V}_q, z_1^{(2)}} \\ r_{\hat{V}_1, z_2^{(2)}} & r_{\hat{V}_2, z_2^{(2)}} & \cdots & r_{\hat{V}_q, z_2^{(2)}} \\ \vdots & \vdots & \ddots & \vdots \\ r_{\hat{V}_1, z_q^{(2)}} & r_{\hat{V}_2, z_q^{(2)}} & \cdots & r_{\hat{V}_q, z_q^{(2)}} \end{bmatrix}$$

where $r_{\hat{U}_i, z_k^{(1)}}$ and $r_{\hat{V}_i, z_k^{(2)}}$ are the sample correlation coefficients between the quantities with subscripts.

- Total (standardized) sample variance in first set

$$= \text{tr}(\mathbf{R}_{11}) = \text{tr}(\hat{\mathbf{a}}_\mathbf{z}^{(1)}\hat{\mathbf{a}}_\mathbf{z}^{(1)\prime} + \hat{\mathbf{a}}_\mathbf{z}^{(2)}\hat{\mathbf{a}}_\mathbf{z}^{(2)\prime} + \cdots + \hat{\mathbf{a}}_\mathbf{z}^{(p)}\hat{\mathbf{a}}_\mathbf{z}^{(p)\prime}) = p$$

Total (standardized) sample variance in second set

$$= \text{tr}(\mathbf{R}_{22}) = \text{tr}(\hat{\mathbf{b}}_\mathbf{z}^{(1)}\hat{\mathbf{b}}_\mathbf{z}^{(1)\prime} + \hat{\mathbf{b}}_\mathbf{z}^{(2)}\hat{\mathbf{b}}_\mathbf{z}^{(2)\prime} + \cdots + \hat{\mathbf{b}}_\mathbf{z}^{(q)}\hat{\mathbf{b}}_\mathbf{z}^{(q)\prime}) = q$$

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

$$\text{tr}(\hat{\mathbf{a}}_\mathbf{z}^{(1)}\hat{\mathbf{a}}_\mathbf{z}^{(1)\prime} + \hat{\mathbf{a}}_\mathbf{z}^{(2)}\hat{\mathbf{a}}_\mathbf{z}^{(2)\prime} + \cdots + \hat{\mathbf{a}}_\mathbf{z}^{(r)}\hat{\mathbf{a}}_\mathbf{z}^{(r)\prime}) = \sum_{i=1}^{r}\sum_{k=1}^{p} r^2_{\hat{U}_i, z_k^{(1)}}$$

$$\text{tr}(\hat{\mathbf{b}}_\mathbf{z}^{(1)}\hat{\mathbf{b}}_\mathbf{z}^{(1)\prime} + \hat{\mathbf{b}}_\mathbf{z}^{(2)}\hat{\mathbf{b}}_\mathbf{z}^{(2)\prime} + \cdots + \hat{\mathbf{b}}_\mathbf{z}^{(r)}\hat{\mathbf{b}}_\mathbf{z}^{(r)\prime}) = \sum_{i=1}^{r}\sum_{k=1}^{q} r^2_{\hat{V}_i, z_k^{(2)}}$$

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

$$R^2_{\mathbf{z}^{(1)}|\hat{U}_1, \hat{U}_2, \ldots, \hat{U}_r} = \frac{\text{tr}(\hat{\mathbf{a}}_\mathbf{z}^{(1)}\hat{\mathbf{a}}_\mathbf{z}^{(1)\prime} + \cdots + \hat{\mathbf{a}}_\mathbf{z}^{(r)}\hat{\mathbf{a}}_\mathbf{z}^{(r)\prime})}{\text{tr}(\mathbf{R}_{11})} = \frac{\sum_{i=1}^{r}\sum_{k=1}^{p} r^2_{\hat{U}_i, z_k^{(1)}}}{p}$$

S_{11}

similar to PCA

$$R^2_{\mathbf{z}^{(2)}|\hat{V}_1, \hat{V}_2, \ldots, \hat{V}_r} = \frac{\text{tr}(\hat{\mathbf{b}}_\mathbf{z}^{(1)}\hat{\mathbf{b}}_\mathbf{z}^{(1)\prime} + \cdots + \hat{\mathbf{b}}_\mathbf{z}^{(r)}\hat{\mathbf{b}}_\mathbf{z}^{(r)\prime})}{\text{tr}(\mathbf{R}_{22})} = \frac{\sum_{i=1}^{r}\sum_{k=1}^{q} r^2_{\hat{V}_i, z_k^{(2)}}}{q}$$

S_{22}

◆ 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

$$\frac{1}{p}\,\mathrm{tr}\big[\underbrace{\mathbf{R}_{11} - \hat{\mathbf{a}}_{\mathbf{z}}^{(1)}\hat{\mathbf{a}}_{\mathbf{z}}^{(1)\prime} - \hat{\mathbf{a}}_{\mathbf{z}}^{(2)}\hat{\mathbf{a}}_{\mathbf{z}}^{(2)\prime} - \cdots - \hat{\mathbf{a}}_{\mathbf{z}}^{(r)}\hat{\mathbf{a}}_{\mathbf{z}}^{(r)\prime}}_{\parallel\ \hat{a}_z^{(r+1)}\hat{a}_z^{(r+1)\prime} + \cdots + \hat{a}_z^{(p)}\hat{a}_z^{(p)\prime}}\big] = 1 - R^2_{\mathbf{z}^{(1)}|\hat{U}_1,\hat{U}_2,\ldots,\hat{U}_r}$$

$$\frac{1}{q}\,\mathrm{tr}\big[\mathbf{R}_{22} - \hat{\mathbf{b}}_{\mathbf{z}}^{(1)}\hat{\mathbf{b}}_{\mathbf{z}}^{(1)\prime} - \hat{\mathbf{b}}_{\mathbf{z}}^{(2)}\hat{\mathbf{b}}_{\mathbf{z}}^{(2)\prime} - \cdots - \hat{\mathbf{b}}_{\mathbf{z}}^{(r)}\hat{\mathbf{b}}_{\mathbf{z}}^{(r)\prime}\big] = 1 - R^2_{\mathbf{z}^{(2)}|\hat{V}_1,\hat{V}_2,\ldots,\hat{V}_r}$$

- Large Sample Inferences
  - Note: $\Sigma_{12}=\mathbf{0} \Rightarrow$ no point in pursuing a CCA. $\Rightarrow$ **Q**: how to test H_0: $\Sigma_{12}=\mathbf{0}$?
  - **Result 10.3.** Let

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

assume normality.

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

Recall: for normal density, maximum likelihood $\propto |\hat{\Sigma}|^{-n/2}$ $\quad |\Sigma| = |\Sigma_{11}|\cdot|\Sigma_{22}|$

Then the likelihood ratio test of H_0: $\boldsymbol{\Sigma}_{12} = \underset{(p\times q)}{\mathbf{0}}$ (all $\rho_i^* = 0$) versus H_1: $\boldsymbol{\Sigma}_{12} \neq \underset{(p\times q)}{\mathbf{0}}$ (at least one $\rho_i^* \neq 0$) rejects H_0 for large values of

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

$$\Lambda = \frac{\text{maximum likelihood under } H_0}{\text{maximum likelihood under } H_0 \cup H_1} = \frac{(|S_{11}||S_{22}|)^{-n/2}}{|S|^{-n/2}}$$

- 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. $= \#$ of parameters in Σ_{12}.
- Bartlett (1939) suggests

Reject H_0: $\boldsymbol{\Sigma}_{12} = \mathbf{0}$ $(\rho_1^* = \rho_2^* = \cdots = \rho_p^* = 0)$ at significance level α if

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

- **Q**: What if H_0: $\Sigma_{12}=\mathbf{0}$ is rejected? Next step?
  - H_0^k: $\rho_1^* \neq 0, \rho_2^* \neq 0, \ldots, \rho_k^* \neq 0, \rho_{k+1}^* = \cdots = \rho_p^* = 0$

    H_1^k: $\rho_i^* \neq 0$, for some $i \geq k + 1$
  - Reject $H_0^{(k)}$ at significance level α if

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

  - the issue of multiple testing should be taken into consideration

❖ **Reading**: Reference, 10.4, 10.5, 10.6