

**Multivariate Statistical Analysis**  
**Statistics 4223/5223 — Spring 2018**

**Assignment 5**

*Reading:*

By Thursday, March 29, read Chapters 11–12 and 16 of *Applied Multivariate Statistical Analysis*, fourth edition; by Wolfgang Härdle and Léopold Simar.

By Tuesday, April 3, read Chapter 14 of Härdle and Simar.

*Homework 5:*

The following problems are nominally due on Thursday, April 5, but can also be submitted in class on Tue or Thu, April 10 or 12, or to the course mailbox in Room 904 SSW, any time before 5:00pm on Friday, April 13.

1. A sample of  $n = 140$  seventh-grade children received four tests on  $x_1 =$  reading speed,  $x_2 =$  reading power,  $y_1 =$  arithmetic speed, and  $y_2 =$  arithmetic power. The correlations for performance are

$$\mathbf{R} = \begin{pmatrix} \mathbf{R}_{XX} & \mathbf{R}_{XY} \\ \mathbf{R}_{YX} & \mathbf{R}_{YY} \end{pmatrix} = \begin{pmatrix} 1.0000 & 0.6328 & 0.2412 & 0.0586 \\ 0.6328 & 1.0000 & -0.0553 & 0.0655 \\ 0.2412 & -0.0553 & 1.0000 & 0.4248 \\ 0.0586 & 0.0655 & 0.4248 & 1.0000 \end{pmatrix}$$

- (a) Find the sample canonical correlation coefficients and sample canonical variables.
- (b) Find the  $p$ -value for a test of  $H_0 : \Sigma_{XY} = \mathbf{0}$ .
- (c) If you reject  $H_0 : \rho_1 = \rho_2 = 0$  at the  $\alpha = .05$  level of significance, find the  $p$ -value for a test of  $H_0 : \rho_2 = 0$ .
- (d) Does reading ability (as measured by the two tests) correlate with arithmetic ability (as measured by the two tests)? Discuss.

2. A random sample of  $n = 70$  families will be surveyed to determine the association between certain “demographic” variables and certain “consumption” variables.

Define the *Criterion* set of variables by

$y_1$  = annual frequency of dining at a restaurant

$y_2$  = annual frequency of attending movies

and the *Predictor* set by

$x_1$  = age of head of household

$x_2$  = annual family income

$x_3$  = educational level of head of household

Suppose 70 observations on the preceding variables give the sample correlation matrix

$$\mathbf{R} = \begin{pmatrix} \mathbf{R}_{XX} & \mathbf{R}_{XY} \\ \mathbf{R}_{YX} & \mathbf{R}_{YY} \end{pmatrix} = \begin{pmatrix} 1.00 & 0.37 & 0.21 & 0.26 & 0.33 \\ 0.37 & 1.00 & 0.35 & 0.67 & 0.59 \\ 0.21 & 0.35 & 1.00 & 0.34 & 0.34 \\ 0.26 & 0.67 & 0.34 & 1.00 & 0.80 \\ 0.33 & 0.59 & 0.34 & 0.80 & 1.00 \end{pmatrix}$$

- Determine the sample canonical correlation coefficients, and find the  $p$ -value for a test of the null hypothesis  $H_0 : \Sigma_{XY} = \mathbf{0}$ .
- If  $H_0 : \Sigma_{XY} = \mathbf{0}$  is rejected at the  $\alpha = .05$  level, test the null hypothesis that the second and higher canonical correlations are all zero.
- Using standardized variables, construct the canonical variables corresponding to the significant ( $\alpha = .05$ ) canonical correlation(s).
- Interpret the canonical variates corresponding to significant ( $\alpha = .05$ ) canonical correlation(s).
- Do the demographic variables have something to say about the consumption variables? Do the consumption variables provide information about the demographic variables?