National Taiwan University
Institute of Industrial Engineering
Jakey BLUE

Data Analytics

111-2 Homework #04

Due at 23h59, March 26, 2023; files uploaded to NTU-COOL

1. (10%) Given two sample covariance matrices:

$$\mathbf{S}_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \text{ and } \mathbf{S}_2 = \begin{bmatrix} 1 & -\frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & 1 & -\frac{1}{2} \\ -\frac{1}{2} & -\frac{1}{2} & 1 \end{bmatrix},$$

Calculate the "total sample variance" and "generalized sample variance" for $\mathbf{S}_1$ and $\mathbf{S}_2$, respectively.

2. (10%) Show that $|\mathbf{S}| = |\mathbf{R}| \prod_{i=1}^{p} s_{ii}$, where **S** and **R** are sample covariance and correlation matrices.

3. (15%) The energy consumption by state in 2001 is recorded in quadrillion of BTUs (10^{15})* in terms of the major sources: $[x_1 = \text{petroleum}, x_2 = \text{natural gas}, x_3 = \text{hydroelectric power}, x_4 = \text{nuclear power}]$. The resulting sample mean and covariance matrices are:

$$\bar{\mathbf{x}} = \begin{bmatrix} 0.766 \\ 0.508 \\ 0.438 \\ 0.161 \end{bmatrix} \text{ and } \mathbf{S} = \begin{bmatrix} 0.856 & 0.635 & 0.173 & 0.096 \\ 0.635 & 0.568 & 0.128 & 0.067 \\ 0.173 & 0.128 & 0.171 & 0.039 \\ 0.096 & 0.067 & 0.039 & 0.043 \end{bmatrix}.$$

*Source: Statistical Abstract of the United States 2006.

   a. Using the summary statistics above to calculate the sample mean and variance of the total energy consumption y_1 $(= x_1 + x_2 + x_3 + x_4)$.
   b. Determine the sample mean and variance of the excess of petroleum consumption over natural gas consumption y_2 $(= x_1 - x_2)$.
   c. What is the covariance between y_1 and y_2.

4. Given the data:

$$\mathbf{X} = \begin{bmatrix} 2 & 12 \\ 8 & 9 \\ 6 & 9 \\ 8 & 10 \end{bmatrix},$$

   a. (4%) Calculate T^2 for testing the hypothesis $H_0: \boldsymbol{\mu}^T = [7, 11]$.
   b. (4%) Specify the distribution of the T^2 in (a).
   c. (2%) From the results in (a) and (b), what is the conclusion you can make for the hypothesis test?

5. The relationship of size and shape for painted turtles are studied by Jolicoeur & Mosimann*. The measurements on the carapaces of 24 female and 24 male turtles can be seen in the following table.

Female			Male		
Length (x_1)	Width (x_2)	Height (x_3)	Length (x_1)	Width (x_2)	Height (x_3)
98	81	38	93	74	37
103	84	38	94	78	35
103	86	42	96	80	35
105	86	42	101	84	39
109	88	44	102	85	38
123	92	50	103	81	37
123	95	46	104	83	39
133	99	51	106	83	39
133	102	51	107	82	38
133	102	51	112	89	40
134	100	48	113	88	40
136	102	49	114	86	40
138	98	51	116	90	43
138	99	51	117	90	41
141	105	53	117	91	41
147	108	57	119	93	41
149	107	55	120	89	40
153	107	56	120	93	44
155	115	63	121	95	42
155	117	60	125	93	45
158	115	62	127	96	45
159	118	63	128	95	45
162	124	61	131	95	46
177	132	67	135	106	47

(10%) Test if the mean vectors of the two populations are equal, given $\alpha = 0.05$.
Hint: You may wish to consider log transformation on the observations.

*Jolicoeur, P., & Mosimann, J. E. (1960). Size and shape variation in the painted turtle. A principal component analysis. *Growth*, *24*(4), 339-354.

6. The spectral reflectance of three species of 1-year-old seedlings was measured at various wavelengths in one experiment involving remote sensing during the growing season. The seedlings were grown with two different levels of nutrients: the optimal level, coded +, and a suboptimal level, coded −. The species of seedlings used were Sitka Spruce (SS), Japanese Larch (JL), and Lodgepole Pine (LP). Two of the variables measured were:

$$x_1 = \text{percent spectral reflectance at wavelength 560 nm (green), and}$$
$$x_2 = \text{percent spectral reflectance at wavelength 720 nm (near} - \text{infrared).}$$

The Cell Means (CM) for each combination of species and nutrient level is as follows. These averages are based on four replications.

560CM	720CM	Species	Nutrient
10.35	25.93	SS	+
13.41	38.63	JL	+
7.78	25.15	LP	+
10.40	24.25	SS	−
17.78	41.45	JL	−
10.40	29.20	LP	−

a. (10%) Treating the cell means as individual observations, perform two MANOVAs to test for the species effect and the nutrient effect, respectively, with $\alpha = 0.05$.

b. (10%) Construct a two-way ANOVA for the 560CM observations and another two-way ANOVA for the 720CM observations. Are these results consistent with the MANOVA results in (a)? If not, can you explain any differences?