STAT 3710 Homework 2

Chapter 3

1. (a) Evaluate T^2, for testing $H_0 : \boldsymbol{\mu}' = [7, 11]$ using data

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

   (b) Specifying the distribution T^2 for the situation in (a).

   (c) Using (a) and (b), test H_0 at the $\alpha = 0.05$ level . What conclusion do you reach ?

2. A physical anthropologist performed a mineral analysis of nine ancient Peruvian hairs. The results for the chroimum (x_1) and strontium (x_2) levels, in parts per million (ppm), were as follows

x_1(Cr)	.48	40.53	2.19	.55	.74	.66	.93	.37	.22
x_2(St)	12.57	73.68	11.13	20.03	20.29	.78	4.64	.43	1.08

It is known that low levels (less than or equal to 0.100ppm) of chromium suggest the presence of diabetes, while strontium is an indication of animal protein intake.

   (a) Construct and plot a 90% joint confidence ellipse for the population mean vector $\boldsymbol{\mu}' = [\mu_1, \mu_2]$, assuming that these nine Peruvian hairs represent a random sample from individual belongs to a particular ancient Peruvian culture.

(b) Obtain the individual simultaneous 90% confidence intervals for μ_1 and μ_2 by "projecting" the ellipse construct in Part a on each coordinate axis. Does it appear as if this Peruvian culture has a mean strontium level of 10 ? That is, any of the points (μ_1 arbitrary,10) in the confidence regions? Is $[.30, 10]'$ a plausible value for $\boldsymbol{\mu}$? Discuss.

(c) Do these data appear to be bivariate normal? Discuss their status with reference to Q-Q plots and a scatter diagram. If the data are not bivariate normal, what implications does this have for the results in Parts a and b?

(d) Repeat the analysis with the obvious "outlying" observation removed. Do the inference change? Comment.

3. Use the sweat data in Table 5.1. (See Example 4.2)

(a) Determine the axes of the 90% confidence ellipsoid for $\boldsymbol{\mu}$. Determine the lengths of these axes.

(b) Construct Q-Q plots for the observations on sweat rate, sodium content, and potassium content, respectively. Construct the three possible scatter plots for pairs of observations. Does the multivariate normal assumption seem justified in this case ? Comment.

4. Observation on two response are collected for two treatment. The observation vectors $[x_1, x_2]'$ are

$$\begin{array}{lcl} \text{Treatment 1} & : & [3,3]', [1,6]', [2,3]' \\ \text{Treatment 2} & : & [2,3]', [5,1]', [3,1]', [2,3]' \end{array}$$

(a) Calculate $\mathbf{S}_{\text{pooled}}$.

(b) Test $H_0 : \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = 0$ employing a two-sample approach with $\alpha = 0.01$.

(c) Construct 99% simultaneous confidence intervals for the differences $\mu_{1i} - \mu_{2i}, i = 1, 2$.

5. Samples of size $n_1 = 45$ and $n_2 = 55$ were taken of Wisconsin homeowners with and without air conditioning, respectively. (Data courtesy of Statistical Laboratory, University of Wisconsin.) Two measurements of electrical usage (in Kilowatt hours) were considered . The first is a measure of total on-peak consumption (X_1) during July, and the second is a measure of total off-peak consumption (X_2) during July. The resulting summary statistics are

$$\bar{\mathbf{x}}_1 = \begin{bmatrix} 204.4 \\ 556.6 \end{bmatrix}, \mathbf{S}_1 = \begin{bmatrix} 13825.3 & 23823.4 \\ 23823.4 & 73107.4 \end{bmatrix}, n_1 = 45$$

$$\bar{\mathbf{x}}_2 = \begin{bmatrix} 130.0 \\ 355.0 \end{bmatrix}, \mathbf{S}_2 = \begin{bmatrix} 8632.0 & 19616.7 \\ 19616.7 & 55964.5 \end{bmatrix}, n_1 = 55$$

Use Box' M-test to test the hypothesis $H_0 : \Sigma_1 = \Sigma_2 = \Sigma$. Here Σ_1 is the covariance matrix for the two measures of usage for the population of Wisconsin homeowners with air conditions, and Σ_2 is the electrical usage covariance matrix for the population of Wisconsin homeowners without air conditioning. Set $\alpha = 0.05$.

6. A likelihood argument provides additional support for pooling the two independent sample covariance matrices to estimate a common covariance matrix in the case of two normal populations. Give the likelihood function $L(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \Sigma)$, for two independent sample sizes n_1 and n_2 from $N_p(\boldsymbol{\mu}_1, \Sigma)$ and $N_p(\boldsymbol{\mu}_2, \Sigma)$ populations, respectively. Show that this likelihood is maximized by the choices $\hat{\boldsymbol{\mu}}_1 = \bar{x}_1, \boldsymbol{\mu}_2 = \bar{x}_2$ and

$$\hat{\Sigma} = \frac{1}{n_1 + n_2}[(n_1 - 1)\mathbf{S}_1 + (n_2 - 1)\mathbf{S}_2] = \left(\frac{n_1 + n_2 - 2}{n_1 + n_2}\right)\mathbf{S}_{\text{pooled}}$$

7. Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_{20}$ be independent with sample size $n = 20$ from an $N_6(\boldsymbol{\mu}, \Sigma)$ population. Specify each of the following completely.

   (a) The distribution of $(\mathbf{X}_1 - \boldsymbol{\mu})'\Sigma(\boldsymbol{\mu}_1 - \boldsymbol{\mu})$.

   (b) The distribution of $\bar{\mathbf{X}}$ and $\sqrt{n}(\bar{\mathbf{X}} - \boldsymbol{\mu})$

   (c) The distribution of $(n - 1)\mathbf{S}$.

   (d) Find the marginal distributions for each of the random vectors

   $$\mathbf{V}_1 = \frac{1}{4}\mathbf{X}_1 - \frac{1}{4}\mathbf{X}_2 + \frac{1}{4}\mathbf{X}_3 - \frac{1}{4}\mathbf{X}_4$$

   and

   $$\mathbf{V}_2 = \frac{1}{4}\mathbf{X}_1 + \frac{1}{4}\mathbf{X}_2 - \frac{1}{4}\mathbf{X}_3 - \frac{1}{4}\mathbf{X}_4$$

   (e) Find the joint density of the random vectors $\mathbf{V}_1$ and $\mathbf{V}_2$ defined in (d).