1. Problem 4.26 p 206, Johnson & Wichern (2007). Only do parts (a) and (b).

(a) Method involving the least work when using SAS/IML or other linear algebra program. The square statistical distances equals the diagonal elements of the following matrix

\[
\begin{pmatrix}
(x_{11} - \bar{x}_1) & (x_{12} - \bar{x}_2) \\
(x_{21} - \bar{x}_1) & (x_{22} - \bar{x}_2) \\
\vdots & \vdots \\
(x_{n1} - \bar{x}_1) & (x_{n2} - \bar{x}_2)
\end{pmatrix}
\begin{pmatrix}
S^{-1}
\end{pmatrix}
\begin{pmatrix}
(x_{11} - \bar{x}_1) & \cdots & (x_{n1} - \bar{x}_1) \\
(x_{12} - \bar{x}_2) & \cdots & (x_{n2} - \bar{x}_2)
\end{pmatrix}
\]

The \((j, j^*)^{th}\) element of the above matrix equals \((x_j - \bar{x})' S^{-1} (x_{j^*} - \bar{x})\) and the \((j, j)^{th}\) elements of the above matrix equal \((x_j - \bar{x})' S^{-1} (x_j - \bar{x})\), which are the squared statistical distance. Computing the matrix and pulling out the diagonals gives

<table>
<thead>
<tr>
<th>Squared Statistical Distance</th>
<th>Within 50% prob contour?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.8753045</td>
<td>no</td>
</tr>
<tr>
<td>2.0203262</td>
<td>no</td>
</tr>
<tr>
<td>2.9009088</td>
<td>no</td>
</tr>
<tr>
<td>0.7352659</td>
<td>yes</td>
</tr>
<tr>
<td>0.3105192</td>
<td>yes</td>
</tr>
<tr>
<td>0.0176162</td>
<td>yes</td>
</tr>
<tr>
<td>3.7329012</td>
<td>no</td>
</tr>
<tr>
<td>0.8165401</td>
<td>yes</td>
</tr>
<tr>
<td>1.3753379</td>
<td>yes</td>
</tr>
<tr>
<td>4.2152799</td>
<td>no</td>
</tr>
</tbody>
</table>

(b) Comparing the squared statistical distances to \(\chi^2_{(0.5)} = 1.3862944\), we find that 5/10 = .5 (half of them) are within the 50% probability contour (which is what we would expect if the observations are from a bivariate normal distribution).


(a) The \( T^2 \) for testing \( H_o: \mu' = (7, 11) \) versus \( H_a: \mu \neq (7, 11) \).

\[
T^2 = 4 \left( \begin{array}{c} 6 - 7 \\ 10 - 11 \end{array} \right) \left( \begin{array}{cc} 8 & -3.33 \\ -3.33 & 2 \end{array} \right)^{-1} \left( \begin{array}{c} 6 - 7 \\ 10 - 11 \end{array} \right) = 13.636
\]

(b) If assumptions and \( H_o \) are true then

\[
T^2 \sim \frac{(n - 1)p}{n - p} F_{p,n-p} = \frac{(4 - 1)^2}{4 - 2} F_{2,2} = 3 F_{2,2}
\]

(c) Since \( F_{2,2}(0.95) = 19 \) and \( (3)(19) = 57 \), retain \( H_o \) because \( T^2 = 13.636 < 57.00 \) (or \( p - value = 0.18 \)). The data are more consistent with the null hypothesis.


Using the \( X \) data, we have

<table>
<thead>
<tr>
<th>( X )-data</th>
<th>( Y )-data</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \begin{array}{cc} 6 &amp; 9 \ 10 &amp; 6 \ 8 &amp; 3 \end{array} )</td>
<td>( \begin{array}{cc} -3 &amp; 15 \ 4 &amp; 16 \ 5 &amp; 11 \end{array} )</td>
</tr>
</tbody>
</table>

Mean vectors \( \left( \begin{array}{c} 8 \\ 6 \end{array} \right) \) \( \left( \begin{array}{c} 2 \\ 14 \end{array} \right) \)

Null \( H_o: \mu = \left( \begin{array}{c} 9 \\ 5 \end{array} \right) \) \( \left( \begin{array}{c} 4 \\ 14 \end{array} \right) \)

\[
S = \left( \begin{array}{cc} 4 & -3 \\ -3 & 9 \end{array} \right) \quad \left( \begin{array}{cc} 19 & -5 \\ -5 & 7 \end{array} \right) \]

\[
T^2 = 0.777 \quad 0.777
\]

4. A number of patients with bronchus cancer were treated with ascorbate and compared with matched control patients who received no ascorbate (Cameron & Pauling, 1978). The data are given below and are on the course web site. The variables measured were

\( y_1 = \) patient: survival time (days) from date of first hospital admission.
\( x_1 = \) matched control: survival time (days) from date of first hospital admission.
\( y_2 = \) patient: survival time from date of untreatability.
\( x_2 = \) match control: survival time from date of untreatability.
(a) Compare \( y_1 \) and \( y_2 \) with \( x_1 \) and \( x_2 \) using a paired comparison \( T^2 \)-test with \( \alpha = .05 \). (Be sure to state your null hypothesis).

I will take differences of patients minus controls in doing the paired \( T^2 \).

\[
H_o : \begin{pmatrix} \mu_{y_1} - \mu_{x_1} \\ \mu_{y_2} - \mu_{x_2} \end{pmatrix} = \delta = 0 \quad \text{versus} \quad H_A : \delta \neq 0
\]

This will be like a test on a single mean vector.

Descriptive statistics: \( n = 16, p = 2 \)

\[
\bar{d} = \begin{pmatrix} 49.5 \\ 106.875 \end{pmatrix} \quad S_d = \begin{pmatrix} 23915.06717461.467 \\ 17461.46716619.45 \end{pmatrix} \quad S_d^{-1} = \begin{pmatrix} 0.0001796 & -0.000189 \\ -0.000189 & 0.0002584 \end{pmatrix}
\]

Using these we find

\[
T^2 = n\bar{d}'S_d^{-1}\bar{d} = 16(46.5, 106.875) \begin{pmatrix} 0.0001796 & -0.000189 \\ -0.000189 & 0.0002584 \end{pmatrix} \begin{pmatrix} 46.5 \\ 106.875 \end{pmatrix} = 22.324
\]

Note that \( F_{2,14}(.05) = 3.739 \). Since our test statistic is \( T^2 = 22.324 \) is greater than \((16 - 1)2/(16 - 2)(3.739) = (2.143)(3.739) = 8.012\), reject \( H_o \) and conclude that the data do not support the null.

(b) If your reject \( H_o \) in part (a),

i. Compute and sketch 95% confidence region:

major \( \rightarrow \) \((-57.700, 156.700) \) to \((19.757, 193.993)\)

minor \( \rightarrow \) \((71.495, 27.505) \) to \((79.811, 133.939)\)

ii. Compute 95% univariate intervals for each variable.

For \( \mu_{y_1-x_1} \) \( \rightarrow \) \((-32.904, 131.904)\)

For \( \mu_{y_2-x_2} \) \( \rightarrow \) \((38.180, 175.570)\)

iii. Compute 95% simultaneous \( T^2 \) intervals for each variable.

For \( \mu_{y_1-x_1} \) \( \rightarrow \) \((-59.932, 158.932)\)

For \( \mu_{y_2-x_2} \) \( \rightarrow \) \((15.649, 198.101)\)

iv. Compute 95% Bonferroni component intervals.

For \( \mu_{y_1-x_1} \) \( \rightarrow \) \((-46.762, 145.762)\)

For \( \mu_{y_2-x_2} \) \( \rightarrow \) \((26.628, 187.122)\)
v. Compute the linear combination that gives largest value of $T^2$ (i.e., a discriminant function).

$$a' = (-0.0113, 0.0183)$$

Both variables appear to contribute equally to rejection (weights are of similar value).

vi. Using what you learned in parts (i)–(iv), interpret. Would you want the treatment if you had this type of cancer?

Although the value of 0 is within the confidence intervals for the difference between treatment and control groups in terms of survival time from date of first hospitalization (i.e., $y_1 - x_2$), this value is unlikely given the covariability between survival time from date of first hospital admission and survival time from date of untreatability.

On average treatment increases the survival time from date of first hospital admission and from date of untreatability. Also survival beyond date of
untreatability is greater for the treatment group... If one desires to increase time to live, then one would probably want the treatment.