Homework # 8
Answer Key

1. Problem 6.8 in Johnson & Wichern, p 338. Do any computations using IML or your favorite matrix computer program.

(a) Break up the observations into mean, treatment, and residual components, as in (6-39). Construct corresponding arrays for each variables (See Example 6.9)

For the First variable:
\[
\begin{align*}
\text{Observation} &= \text{Mean} + \text{Treatment Effect} + \text{Residual} \\
X_{ij} &= \bar{X} + (\bar{X}_l - \bar{X}) + (X_{lj} - \bar{X}_l) \\
\begin{pmatrix}
6 & 5 & 8 & 4 & 7 \\
3 & 1 & 2 \\
2 & 5 & 3 & 2
\end{pmatrix}
&= \begin{pmatrix}
4 & 4 & 4 & 4 & 4 \\
4 & 4 & 4 \\
4 & 4 & 4 & 4
\end{pmatrix}
+ \begin{pmatrix}
2 & 2 & 2 & 2 & 2 \\
-2 & -2 & -2 \\
-1 & -1 & -1 & -1
\end{pmatrix}
+ \begin{pmatrix}
0 & -1 & 2 & -2 & 1 \\
1 & -1 & 0 \\
-1 & 2 & 0 & -1
\end{pmatrix}
\end{align*}
\]

For the Second variable:
\[
\begin{align*}
\text{Observation} &= \text{Mean} + \text{Treatment Effect} + \text{Residual} \\
X_{ij} &= \bar{X} + (\bar{X}_l - \bar{X}) + (X_{lj} - \bar{X}_l) \\
\begin{pmatrix}
7 & 9 & 6 & 9 & 9 \\
3 & 6 & 3 \\
3 & 1 & 1 & 3
\end{pmatrix}
&= \begin{pmatrix}
5 & 5 & 5 & 5 & 5 \\
5 & 5 & 5 \\
5 & 5 & 5 & 5
\end{pmatrix}
+ \begin{pmatrix}
3 & 3 & 3 & 3 & 3 \\
-1 & -1 & -1 \\
-3 & -3 & -3 & -3
\end{pmatrix}
+ \begin{pmatrix}
-1 & 1 & -2 & 1 & 1 \\
-1 & 2 & -1 \\
1 & -1 & -1 & -1
\end{pmatrix}
\end{align*}
\]

(b) Using the information in Part (a), construct a 1-way MANOVA table.

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>df</th>
<th>SSCP</th>
<th>(\Lambda^*)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>3 - 1 = 2</td>
<td>(H = \begin{pmatrix} 36 &amp; 48 \ 48 &amp; 84 \end{pmatrix})</td>
<td>.0362</td>
</tr>
<tr>
<td>Residual</td>
<td>5 + 3 + 4 - 3 = 9</td>
<td>(E = \begin{pmatrix} 18 &amp; -13 \ -13 &amp; 18 \end{pmatrix})</td>
<td></td>
</tr>
<tr>
<td>Total (corrected)</td>
<td>11</td>
<td>(T = \begin{pmatrix} 54 &amp; 35 \ 35 &amp; 102 \end{pmatrix})</td>
<td></td>
</tr>
</tbody>
</table>
(c) Evaluate Wilk’s lambda, $\Lambda^*$, and use Table 6.3 to test for treatment effects. Set $\alpha = .01$.

\[ n_+ = 12 \]
\[ p = 2 \]
\[ g = 3 \]
\[ \nu_e = 9 \]
\[ \nu_h = 2 \]
\[ F = \left( \frac{\nu_e-1}{\nu_h} \right) \left( \frac{1 - \sqrt[n]{\Lambda^*}}{\sqrt[n]{\Lambda^*}} \right) = 17.03 \]
\[ df_{num} = 4, \, df_{den} = 16, \, p < .01. \]
Reject $H_0$. The data support the conclusion that the mean vectors are not equal.

(d) The MANOVA model can be expressed in terms of the multivariate general linear model, $X = AB + e$. Provide two possible versions of the design matrix $A$.

Here are three possibilities

A rank 3 design matrix:

\[
A_0 = \begin{pmatrix}
1 & 1 & 0 & 0 \\
1 & 1 & 0 & 0 \\
1 & 1 & 0 & 0 \\
1 & 1 & 0 & 0 \\
1 & 1 & 0 & 0 \\
1 & 1 & 0 & 0 \\
1 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 \\
1 & 0 & 0 & 1 \\
1 & 0 & 0 & 1 \\
1 & 0 & 0 & 1
\end{pmatrix}
\]
A full rank design matrix One:

\[
A_1 = \begin{pmatrix}
1 & 1 & 0 \\
1 & 1 & 0 \\
1 & 1 & 0 \\
1 & 1 & 0 \\
1 & 1 & 0 \\
1 & 0 & 1 \\
1 & 0 & 1 \\
1 & 0 & 1 \\
1 & 0 & 0 \\
1 & 0 & 0 \\
1 & 0 & 0 \\
1 & 0 & 0 \\
\end{pmatrix}
\]

A full rank design matrix Two:

\[
A_2 = \begin{pmatrix}
1 & 1 & 0 \\
1 & 1 & 0 \\
1 & 1 & 0 \\
1 & 1 & 0 \\
1 & 1 & 0 \\
1 & 0 & 1 \\
1 & 0 & 1 \\
1 & 0 & 1 \\
1 & -1 & -1 \\
1 & -1 & -1 \\
1 & -1 & -1 \\
\end{pmatrix}
\]

(e) For each design matrix, given in part (d)

i. Estimate \( \mathbf{B} \)

<table>
<thead>
<tr>
<th>Effect</th>
<th>( \hat{\mathbf{B}}_0 )</th>
<th>( \hat{\mathbf{B}}_1 )</th>
<th>( \hat{\mathbf{B}}_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>intercept</td>
<td>2.75</td>
<td>3.5</td>
<td>3.66</td>
</tr>
<tr>
<td>Rx 1</td>
<td>3.25</td>
<td>4.5</td>
<td>3.66</td>
</tr>
<tr>
<td>Rx 2</td>
<td>-0.75</td>
<td>0.5</td>
<td>-1</td>
</tr>
<tr>
<td>Rx 3</td>
<td>0.25</td>
<td>-1.5</td>
<td>2</td>
</tr>
</tbody>
</table>

We get different values for \( \hat{\mathbf{B}} \); however, the differences between elements for the groups are the same. For example, taking the difference between treatments 1 and 2 on the first variable we get

\[
\hat{\mathbf{B}}_0 \rightarrow 3.25 + 0.75 = 4 \\
\hat{\mathbf{B}}_1 \rightarrow 3.00 + 1.00 = 4 \\
\hat{\mathbf{B}}_2 \rightarrow 2.33 + 1.67 = 4 
\]
These are the difference between means (treatment effects) for groups 1 and 2 on the first variable. We could do this for the other groups and other variables, but the same basic result would be found:

Let $C_{r \times g}$ be a contrast matrix and $C_0 = (0|C)$, then

$$C_0 \hat{B}_0 = C \hat{B}_1 = C \hat{B}_2$$

**ii. Estimate the residuals.**

Regardless of design matrix they are all the same; namely,

$$\hat{\epsilon} = \begin{pmatrix}
0 & -1 \\
-1 & 1 \\
2 & -2 \\
-2 & 1 \\
1 & 1 \\
1 & -1 \\
-1 & 2 \\
0 & -1 \\
-1 & 1 \\
2 & -1 \\
0 & -1 \\
-1 & 1
\end{pmatrix}$$

**iii. Compute the cell means (i.e., $\hat{X} = A \hat{B}$).**

Regardless of which design matrix we use, we get the same result; namely,

$$\bar{x}_1 = \begin{pmatrix} 6 \\ 8 \end{pmatrix} \quad \bar{x}_2 = \begin{pmatrix} 2 \\ 4 \end{pmatrix} \quad \bar{x}_3 = \begin{pmatrix} 3 \\ 2 \end{pmatrix}$$

(f) *Compare the estimates obtained in all parts of this problem.*

I’ve done this above.

2. Baten, Tack, and Baeder (1958) compared judge’s scores on fish prepared by three methods. Twelve fish were cooked by each method, and several judges tasted the fish samples and rated each on four variables: $y_1 =$ aroma, $y_2 =$ flavor, $y_3 =$ texture, and $y_4 =$ moisture. The raw data are given below and will be mailed to you. Each entry is an average score for the judges on that fish. You can use PROC GLM and/or IML for this problem.

(a) *Compare the three methods using MANOVA.*
Reject $H_0 : \mu_1 = \mu_2 = \mu_3$. The data support the conclusion that the methods don’t have the same mean vectors.

(b) Using multivariate contrasts, test the following two comparisons:

- **1 and 2 versus 3**

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Value</th>
<th>F Value</th>
<th>Num DF</th>
<th>Den DF</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wilks’ Lambda</td>
<td>0.2248</td>
<td>8.33</td>
<td>8</td>
<td>60</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

The data support the conclusion that the means of methods 1 and 2 differ from those of method 3.

- **1 versus 2**

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Value</th>
<th>F Value</th>
<th>Num DF</th>
<th>Den DF</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wilks’ Lambda</td>
<td>0.2701</td>
<td>20.27</td>
<td>4</td>
<td>30</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

It looks like methods 1 and 2 don’t have the same mean vectors on the four variables.

(c) Compute simultaneous confidence intervals for each of the variables comparing methods

- **1 and 2 versus 3**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>std error</th>
<th>95% Confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aroma</td>
<td>0.6917</td>
<td>0.4507</td>
<td>(−0.225, 1.610)</td>
</tr>
<tr>
<td>Flavor</td>
<td>1.3000</td>
<td>0.3584</td>
<td>(0.571, 2.029) **</td>
</tr>
<tr>
<td>Texture</td>
<td>−1.0667</td>
<td>0.4194</td>
<td>(−1.920, −0.207) **</td>
</tr>
<tr>
<td>Moisture</td>
<td>−0.6083</td>
<td>0.4001</td>
<td>(−1.422, 0.206)</td>
</tr>
</tbody>
</table>

- **1 versus 2**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>std error</th>
<th>95% Confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aroma</td>
<td>0.1250</td>
<td>0.2602</td>
<td>(−0.404, 0.654)</td>
</tr>
<tr>
<td>Flavor</td>
<td>0.5000</td>
<td>0.2069</td>
<td>(0.079, 0.921) **</td>
</tr>
<tr>
<td>Texture</td>
<td>0.1333</td>
<td>0.2421</td>
<td>(−0.359, 0.626)</td>
</tr>
<tr>
<td>Moisture</td>
<td>0.1083</td>
<td>0.2310</td>
<td>(−0.362, 0.578)</td>
</tr>
</tbody>
</table>

(d) If the test in part (a) is significant, run an ANOVA F-test on each $y_i$. 

<table>
<thead>
<tr>
<th>Variable</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aroma</td>
<td>1.29</td>
<td>.29</td>
</tr>
<tr>
<td>Flavor</td>
<td>9.50</td>
<td>&lt; .01</td>
</tr>
<tr>
<td>Texture</td>
<td>3.39</td>
<td>.05</td>
</tr>
<tr>
<td>Moisture</td>
<td>1.27</td>
<td>.30</td>
</tr>
</tbody>
</table>

Since both Flavor and Texture are significant, I did (univariate) multiple comparisons using Tukey’s method for these two variables (if you do the other two (moisture and aroma, there are no differences). I found

Pr > |t| for H0: LSMean(i)=LSMean(j)

**Dependent Variable: y2 (Flavor)**

<table>
<thead>
<tr>
<th>i/j</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>0.0543</td>
<td>0.0004</td>
</tr>
<tr>
<td>2</td>
<td>0.0543</td>
<td></td>
<td>0.1455</td>
</tr>
<tr>
<td>3</td>
<td>0.0004</td>
<td></td>
<td>0.1455</td>
</tr>
</tbody>
</table>

Methods 1 and 3 are significantly different (p < .001) with respect to Flavor; however, methods 1 and 2 (p = .05) and methods 2 and 3 (p = .15) are not.

Pr > |t| for H0: LSMean(i)=LSMean(j)

**Dependent Variable: y3 (Texture)**

<table>
<thead>
<tr>
<th>i/j</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>0.8468</td>
<td>0.1469</td>
</tr>
<tr>
<td>2</td>
<td>0.8468</td>
<td></td>
<td>0.0474</td>
</tr>
<tr>
<td>3</td>
<td>0.1469</td>
<td></td>
<td>0.0474</td>
</tr>
</tbody>
</table>

Only methods 2 and 3 are significantly different (p = .05) with respect to Texture.

(e) Using the results from parts (a), (b), (c) and (d), write a short summary of what you found out about the methods of cooking fish.

Summary: The methods overall seem fairly similar with respect to aroma and seem to differ with respect to flavor. In terms of moisture and texture, it
appears that texture is a point where there may be differences. However, when I look at the discriminant functions, which you didn’t have to do, I do notice that the variables flavor and moisture are the important ones.

In sum: Methods 1, 2, and 3 are similar on aroma and moisture (but it looks like texture).

The methods on the first two discriminant functions (flavor and moisture dominate — have the largest weight in the linear combinations)
Methods of Cooking Fish

Method: One, Two, Three