Name:________________________________________________________

Midterm Exam
Due Thursday March 16, 2017
At the Beginning of Lecture

There are 5 questions, most of which have multiple parts. The total number of points is 100. The point values for each question is give next to the problem, as well as points for each part.

Data sets for the problems 2, 3, and 4 are available from the course web-site.

All work should be your own! You may use notes, books, internet, or whatever you desire EXCEPT consulting with another person. You may not talk to or correspond with your classmates or others about this exam until after the due date. You may not get help from anyone on any of these problems. Any and all questions or problems with this exam should be directed to me (the instructor). The only exceptions are problems with SAS, which may be directed to either me or Wes Crues.

In addition to answering the questions, include an appendix with the code that you wrote to obtain answers to all problems. This should be labeled by problem and part of problem.
1. (30 points) You have observations on a random sample of 25 individuals on two attitude measures, $X_1$ and $X_2$. Assume that $X' = (X_1, X_2)$ comes from a population that is $N_2(\mu, \Sigma)$. Based on your $n = 25$ sample, you compute $\bar{x}$ and $S$ in the usual way. Using the sample data, you obtain a 95% confidence region for $\mu$ and one-at-a-time confidence intervals for $\mu_1$ and $\mu_2$. These are plotted below and drawn to scale.

![Diagram of confidence region and confidence intervals for $\mu_1$ and $\mu_2$.]

(a) (2 points) What does $\bar{x}$ equal? (Give a numerical value for your answer).

(b) (2 points) Is $s_{11}$ or $s_{22}$ larger? Explain.

(c) (3 points) Is the GSV based on $S$ greater than 0? Explain.

(d) (3 points) What does the GSV based on $R$ (standardized variables) equal? Give the numerical value and explain how you got it.

(e) (3 points) What do the eigenvectors of $S$ equal (i.e., $e_1$ and $e_2$)? You should give numerical values (note: there are multiple correct answers for this problem. You just need to give one set of correct eigenvectors).

(f) (3 points) Would the hypothesis $H_0 : \mu' = (3.9, 2.0)$ be rejected if you used Hotellings $T^2$ with $\alpha = .05$? Why or why not? (note: no numerical computations are required to answer this question).
(g) (3 points) Would the hypothesis \( H_0 : \mu' = (3.9, 2.0) \) be rejected if you used a likelihood ratio test with \( \alpha = .05 \)? Why or why not? (note: no numerical computations are required to answer this question).

(h) (3 points) Would either of the hypotheses \( H_{o1} : \mu_1 = 3.9 \) or \( H_{o2} = \mu_2 = 2.0 \) (2-tailed tests) be rejected if you used univariate t tests to test each of them? Why or why not?

(i) (3 points) Would 3.50 fall within the simultaneous \( T^2 \) interval for \( \mu_1 \)? Would 3.25 fall within the simultaneous \( T^2 \) interval for \( \mu_2 \)? Explain how you know?

(j) (5 points) In this case, what (if any) are the advantages of using a multivariate approach versus a univariate approach to test \( H_0 : \mu_1 = \mu_{1o} \) and \( \mu_2 = \mu_{2o} \) where \( \mu_{1o} \) and \( \mu_{2o} \) are specified values of means for variables \( X_1 \) and \( X_2 \), respectively? Are there any advantages to using a univariate approach rather than a multivariate one? Is so, what are they?
2. (20 points) The data given below are also available from the course web-site. The data are results from an experiment where 11 subjects responded to probe words at 5 points in a sentence. The variables are response times for the $i^{th}$ probe word.

(a) (6 points) Does the position of the probe word lead to different mean reaction times? (state hypotheses, assumptions, what test you are doing, show results, interpretation/conclusion).

(b) (6 points) If the test done in part (a) was significant, determine which positions have different mean reaction times? Which is on average the fastest? The slowest? (Hint: consider linear confidence intervals of linear combinations of means).

(c) (2 points) Summarize your results in a few sentences (something you might include in a paper describing the study).

<table>
<thead>
<tr>
<th>Position of probe word</th>
<th>Subject 1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>51</td>
<td>36</td>
<td>50</td>
<td>35</td>
<td>42</td>
</tr>
<tr>
<td>2</td>
<td>27</td>
<td>20</td>
<td>26</td>
<td>17</td>
<td>27</td>
</tr>
<tr>
<td>3</td>
<td>37</td>
<td>22</td>
<td>41</td>
<td>37</td>
<td>30</td>
</tr>
<tr>
<td>4</td>
<td>42</td>
<td>36</td>
<td>32</td>
<td>34</td>
<td>27</td>
</tr>
<tr>
<td>5</td>
<td>27</td>
<td>18</td>
<td>33</td>
<td>14</td>
<td>27</td>
</tr>
<tr>
<td>6</td>
<td>43</td>
<td>32</td>
<td>43</td>
<td>35</td>
<td>40</td>
</tr>
<tr>
<td>7</td>
<td>41</td>
<td>22</td>
<td>36</td>
<td>25</td>
<td>38</td>
</tr>
<tr>
<td>8</td>
<td>38</td>
<td>21</td>
<td>31</td>
<td>20</td>
<td>16</td>
</tr>
<tr>
<td>9</td>
<td>36</td>
<td>23</td>
<td>27</td>
<td>25</td>
<td>28</td>
</tr>
<tr>
<td>10</td>
<td>26</td>
<td>31</td>
<td>31</td>
<td>32</td>
<td>36</td>
</tr>
<tr>
<td>11</td>
<td>29</td>
<td>20</td>
<td>25</td>
<td>26</td>
<td>25</td>
</tr>
</tbody>
</table>
3. (30 points) The data for this problem along with SAS code to create a SAS data set are on the course web-site. The data consist of $n = 88$ students on five 100 point exams: mechanics, vectors, algebra, analysis, and statistics. The first two exams were closed book (i.e., mechanics and vectors) and the last two exams were open book (i.e., algebra, analysis, and statistics). The mean of the two open book exams and the mean of the three close book exams were computed.

The 7 variables in the data set are

- **mech** = mechanics
- **vectors** = vectors
- **algebra** = algebra
- **analysis** = analysis
- **stats** = statistics
- **closed** = mean of mechanics and vectors
- **open** = mean of algebra, analysis and statistics

(a) (4 points) Compute the covariance matrix for the 7 variables and find the eigenvalues and eigenvectors of the matrix. What is the rank of the matrix? Given the rank of the matrix, what implications does this have regarding the data and various descriptive statistics that can be computed.

(b) Compute the covariance matrix of the 5 test scores (i.e., exclude closed and open), which for and clarity we’ll refer to as $S_x$.

i. (2 points) Find (& report) the eigenvalues and eigenvectors of the covariance matrix $S_x$.

ii. Suppose that we define two “new” variables $Y_1$ and $Y_2$ as

$$Y_{1j} = e_1' X_j$$
$$Y_{2j} = e_2' X_j$$

where $X_j$ is a $(5 \times 1)$ vector of scores on the 5 tests for the $j^{th}$ individual and $e_1$ and $e_2$ are the eigenvectors of the covariance matrix for the 5 test scores (i.e., $S_x$). We’ll call this the “$Y$-data”.

A. (2 points) What is the covariance matrix for $Y' = (Y_1, Y_2)$? For clarity, we’ll refer to this covariance matrix as $S_y$.

B. (2 points) What is the total sampling variance of the $Y$-data?
C. (2 points) What is the generalized sample variance for the Y-data?

D. (3 points) How are the results from the previous three parts, (b)iiA, (b)iiB and (b)iiC, related to the eigenvalues of $S_x$?

(c) Repeat part (b) and all of it’s sub-parts except standardize the X-data, which we’ll refer to as Z-data, so that the means for each of the 5 variables are all 0 and the variances are all 1... try the following to standardize the variables

$$Z_{n \times p} = (X_{n \times p} - 1_n \bar{x}')(\text{diag}(s_{x,ii}))^{-1/2}$$

Compute the covariance matrix of the 5 standardized test scores (i.e., exclude closed and open), which for and clarity we’ll refer to as $S_z$.

i. (2 points) Find ( & report) the eigenvalues and eigenvectors of the covariance matrix $S_z$.

ii. Suppose that we define two “new” variables $Y^*_1$ and $Y^*_2$ as

$$Y^*_{1j} = e_1' Z_j$$
$$Y^*_{2j} = e_2' Z_j$$

where $Z_j$ is a $(5 \times 1)$ vector of scores on the 5 standardized test scores for the $j^{th}$ individual and $e_1$ and $e_2$ are the eigenvectors of the covariance matrix for the 5 test scores (i.e., $S_z$). We’ll call this the “$Y^*$-data”.

A. (2 points) What is the covariance matrix for $Y^* = (Y^*_1, Y^*_2)$? For clarity, we’ll refer to this covariance matrix as $S_{y^*}$.

B. (2 points) What is the total sampling variance of the $Y^*$-data?

C. (2 points) What is the generalized sample variance for the $Y^*$-data?

D. (3 points) How are the results from the three previous parts, (c)iiA, (c)iiB and (c)iiC, related to the eigenvalues of $S_z$?

(d) (4 points) Compare your results from parts (b) (unstandardized variables) and (c) (standardized variables). Are they the same or different? Why or why not? (Your explanation for “Why or why not” may be intuitive and/or mathematical).
4. (20 points) Ten pairs of college freshman with matching IQ scores were used in a problem–solving experiment, members of each pair being assigned to the experimental and control groups. The task was to solve two sets of riddle-like problems, one set being entirely verbal and the other set involving some numerical reasoning. The problems in the two sets had similar logical structures, so that some practice effect could be anticipated even in the control group. The experimental group was given instructions that should facilitate performance, especially on the second set of problems.

The scores earned by the 10 experimental subjects and their matched controls on the two sets of problems were as follows:

<table>
<thead>
<tr>
<th></th>
<th>Experimental</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_1$</td>
<td>19 28 30 31 33 34 41 35 45 53</td>
<td>19 27 31 30 34 34 39 36 42 50</td>
</tr>
<tr>
<td>$X_2$</td>
<td>26 33 37 34 41 36 45 40 42 56</td>
<td>18 31 36 31 36 37 39 41 43 52</td>
</tr>
</tbody>
</table>

(a) Test the significance of the difference between the experimental and control group means at the $\alpha = .05$ level? Be sure to state the statistical hypotheses and assumptions, and give the formula for the test statistic, its sampling distribution, and a decision rule for rejecting or retaining the null hypotheses.

(b) If you reject the null, do follow up analyses to investigate why the null was rejected.