This is a 2-day take-home test, with 5 questions, each carrying equal weight. The test will be available after the end of class on Monday, March 10 and you must return it in class on Wednesday, March 12. Please sign this page and attach it to your answers. Your test paper will not be accepted without the signed cover page.

Note: You will submit your answers to this test for grading; by signing below, you acknowledge that you are familiar with the course policy on graded material and the university policy on academic integrity, cited in the course syllabus. In particular, your signature means that this is your own individual work, and that you neither gave nor received unauthorized aid.

All questions refer to the data in Table 1.6 of Johnson and Wichern:

http://www.stat.ncsu.edu/people/bloomfield/courses/st731/wrap/JandW/T01-06.dat

The table contains data comparing the responses of two groups of subjects to visual stimuli (the last column contains an indicator variable: 0 for a non-multiple-sclerosis group, 1 for a multiple-sclerosis group).
1. Consider variables $x_2$ and $x_4$.

   (a) Plot the scatter diagram for the multiple-sclerosis group, and comment on its appearance.

   (b) Make the same diagram for the non-multiple-sclerosis group, and comment on the comparison of the two diagrams.

   (c) Compute the $\bar{x}$, $S_n$, and $R$ arrays for both groups. Summarize the major differences between the statistics for the two groups.

2. Consider again variables $x_2$ and $x_4$. For the multiple-sclerosis group, find the eigenvalues and eigenvectors of:

   (a) the variance-covariance matrix $S_n$;

   (b) the correlation matrix $R$.

3. Investigate the marginal normality of variable $x_4$ for the non-multiple-sclerosis group.

   (a) Construct a $Q$-$Q$ plot for this variable. Is it consistent with a marginal normal distribution?

   (b) Construct $Q$-$Q$ plots for various power transformations of the variable. What power gives transformed data that seem most consistent with a marginal normal distribution?

4. Consider again variables $x_2$ and $x_4$, for the multiple-sclerosis group.

   (a) Find the maximum likelihood estimators of the $2 \times 1$ mean vector $\mu$ and the $2 \times 2$ covariance matrix $\Sigma$.

   (b) If $X_2$ and $X_4$ have the bivariate normal distribution with parameters as estimated in part 4a, what is the conditional distribution of $X_4$, given that $X_2 = 200$?

5. For the non-multiple-sclerosis group data, consider again variables $x_2$ and $x_4$.

   (a) Calculate the statistical distances

   \[
   (x_j - \bar{x})' S^{-1} (x_j - \bar{x})
   \]

   and determine the proportion falling within the 50% probability contour of a bivariate normal distribution.

   (b) Construct the corresponding chi-square plot. Does the plot suggest the presence of any outliers, or any other departure from bivariate normality?