ASSIGNMENT 3

NAME

| Reading Assignment: | Chapters 5 and 6 and section 7.7 in Johnson and Wichern. |
|---------------------|---|
| Written Assignment: | Due on Monday, February 28 |
| EXAM: | Thursday, March 3, at 7 pm in a location to be announced. |

1. Disturbing the equilibrium of an organism "in vivo" by administering a treatment, usually a drug, and monitoring the reaction is a commonly used procedure in biochemistry. In one such study, 7 human subjects were given an alcoholic drink, and blood samples were taken at 1, 2, 3, and 4 hours after the alcoholic drink was consumed. The blood glucose concentration (mg/10 liters of blood) was determined for each blood sample. The sample mean vector and sample covariance matrix, for these data are

| $\overline{\mathbf{X}}_{\sim} = \begin{bmatrix} \\ \\ \\ \\ \\ \end{bmatrix}$ | 6.84 | S = | [10.1 | 2.0 | 1.1 | 0.2 | | |
|---|------|-----|-------|-----|-----|-----|-----|--|
| | 5.06 | | S _ | 2.0 | 6.5 | 2.1 | 0.5 | |
| | 4.18 | | 5 = | 1.1 | 2.1 | 4.9 | 1.1 | |
| | 3.41 | | | 0.2 | 0.5 | 1.1 | 1.4 | |

- A. Plot the sample means against the four time points and connect them with straight line segments. To help visualize whether differences in means are large relative to variation in responses for different subjects, insert vertical bars representing one-at-a-time 95% confidence intervals for the individual means. Note how variation decreases as the mean glucose concentration decreases.
- B. Let μ_1 , μ_2 , μ_3 , μ_4 denote the population means for blood glucose concentrations at 0, 1,2, 3, and 4 hours, respectively. Test the null hypothesis that the mean blood glucose concentrations are the same for the four time points, i.e., $\mu_1 = \mu_2 = \mu_3 = \mu_4$. First write the null hypothesis in the form

$$H_{0}:\begin{bmatrix} & & \\ &$$

State your conclusion.

C. Use the Bonferroni inequality to compute simultaneous 95% confidence intervals for $\mu_1 - \mu_2$, $\mu_1 - \mu_3$, $\mu_2 - \mu_3$, $\mu_1 - \mu_4$, $\mu_2 - \mu_4$, $\mu_3 - \mu_4$.

D. Compute the T² and F values for testing H₀: $C \mu = 0$ where $C = \begin{bmatrix} -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \end{bmatrix}$.

What does this null hypothesis imply about how mean blood glucose concentrations change in the first two hours? Report

 $T^2 =$ ____ F =____ d.f. =____ p-value =____

State your conclusion,

E. Test the null hypothesis that all 6 correlations are zero. Report values for

 $X^2 =$ _____ d.f. = _____ p-value = _____.

State your conclusion.

F. Test the null hypothesis that all 6 correlations are equal and all four variances are equal. Report the value of the large sample chi-squared test using the Bartlett correction.

 $X^2 =$ _____ d.f. = _____ p-value = _____.

State your conclusions.

G. Consider the linear mixed model $Y_{ij} = \mu + S_j + \tau_i + \varepsilon_{ij}$, where Y_{ij} corresponds to the measurement of blood glucose concentration taken on the *j*-th subject at the *i*-th time point. Here there are random effects for subjects and fixed effects for time points. Assume the $S_j \sim \text{NID}(0, \sigma_S^2)$ and $\varepsilon_{ij} \sim \text{NID}(0, \sigma_e^2)$ and any S_j is independent of any ε_{ij} . Complete the following ANOVA table.

| Source of Variation | d.f. | Sums of Squares | Mean Square | F | p-value | Conserv- ative df | p-value for conserv. df |
|---------------------|------|--------------------|----------------|---|---------|----------------------|-------------------------|
| Subjects | | - | - | | - | | |
| Time Points | | | | | | | |
| Error | | 82.05 | | | | | |
| Corrected total | l | 183.07 | | | | | |

Should conservative degrees of freedom be used in this case? Explain.

H. For the "mixed" model in Part G, the correlations between responses on a single subject are the same for all pairs of time points. Estimate the correlation between blood glucose concentrations at 1 and 2 hours.

3. Observations on the lengths of tails and wings were collected for a number of males and females of a particular bird species. The data are:

| females: | $\begin{bmatrix} 4 \\ 7 \end{bmatrix} \begin{bmatrix} 5 \\ 8 \end{bmatrix} \begin{bmatrix} 4 \\ 6 \end{bmatrix} \begin{bmatrix} 6 \\ 8 \end{bmatrix} \begin{bmatrix} 5 \\ 7 \end{bmatrix} \begin{bmatrix} 5 \\ 10 \end{bmatrix} \begin{bmatrix} 6 \\ 10 \end{bmatrix}$ |
|----------|--|
| males: | $\begin{bmatrix} 6\\9 \end{bmatrix} \begin{bmatrix} 5\\7 \end{bmatrix} \begin{bmatrix} 7\\11 \end{bmatrix} \begin{bmatrix} 7\\10 \end{bmatrix} \begin{bmatrix} 5\\8 \end{bmatrix}$ |

- A. Compute the pooled estimate of the covariance matrix, call it S, and report its degrees of freedom.
- B. Test the null hypothesis that the covariance matrices are the same for the male and female birds. Report values for

 $M = _ ___ C^{-1} = _ ___ d.f. = _ ___ p-value=____$

State your conclusion:

D. Compute the value of the two-sample T²-statistic for testing the null hypothesis of equal mean vectors for male and females. Report

 $T^2 =$ _____ F =_____ d.f. =_____ p-value=_____

State your conclusion:

E. Use the Bonferonni method to construct simultaneous 95% confidence intervals for the differences between females and males in mean tail lengths and mean wing lengths:

 $\underline{\qquad } \leq \mu_{1F} - \mu_{1M} \leq \underline{\qquad } \leq \mu_{2F} - \mu_{2M} \leq \underline{\qquad }$

4. The data for this problem come from a comparative bioassay of potency of insulin made by two different manufacturers, coded 1 and 2. The two insulins were supposed to have the same potency. The same set of dosage levels (0.5 units, 1.0 units, and 1.5 units) were used for each insulin. A total of 54 rabbits were used in this experiment with 9 rabbits randomly assigned to each combination of dosage levels and manufacturer. Blood sugar concentrations, Y₁, Y₂, Y₃, Y₄, Y₅, were measured for each rabbit at 1, 2, 3, 4, and 5 hours after the appropriate insulin dose was administered. An initial measurement, Y₀, of blood sugar concentration was taken just before the insulin was administered. Percent deviations from the initial concentration for each rabbit were computed as

$$X_i = 100(Y_i - Y_0)/Y_0$$
, $i = 1, 2, 3, 4, 5$.

The data are posted on the course web page as rabbits.dat. There is one line of data for each rabbit and eight numbers on each line in the following order:

rabbit identification number, manufacturer, dosage level, X1, X2, X3, X4, X5.

- A. Construct a profile plot of the mean percent deviations versus time with one profile for each of the six combinations of dosage level and manufacturer. Submit the plot with this assignment. Which profiles appear to be the same? Which appear to be parallel?
- B. Test the null hypothesis that all 6 population profiles are the same. Report

 $T^2 = _$ $F = _$ $d.f. = _$ $p-value = _$

State your conclusion:

C. Test the null hypothesis that all 6 population profiles are parallel. Report

 $T^2 = _$ $F = _$ $d.f. = _$ $p-value = _$

State your conclusion:

D. Let $\mu_{ij} = (\mu_{1ij}, \mu_{2ij}, \mu_{3ij}, \mu_{4ij}, \mu_{5ij})'$ denote the vector of mean percent

deviations of blood sugar concentrations at 1, 2, 3, 4, and 5 hours when the j -th dosage of the insulin made by the i-th manufacturer is used. Define

$$\beta = \begin{bmatrix} \mu_{111}, \mu_{211}, \mu_{311}, \mu_{411}, \mu_{511} \\ \mu_{112}, \mu_{212}, \mu_{312}, \mu_{412}, \mu_{512} \\ \mu_{113}, \mu_{213}, \mu_{313}, \mu_{413}, \mu_{513} \\ \mu_{121}, \mu_{221}, \mu_{321}, \mu_{421}, \mu_{521} \\ \mu_{122}, \mu_{222}, \mu_{322}, \mu_{422}, \mu_{522} \\ \mu_{123}, \mu_{223}, \mu_{323}, \mu_{423}, \mu_{523} \end{bmatrix}$$

If possible, give the M and C matrices needed to write each of the following null hypotheses in the form H_0 : $C\beta M = 0$ and report values for approximate F tests (derived from the Wilks Criterion), corresponding degrees of freedom, and p-values. If this cannot be done, say so.

- (i) For each dosage level, the profiles for the two manufacturers are parallel, but profiles for different dosage levels need not be parallel.
- (ii) For each dosage level, the profiles for the two manufactures are the same, but profiles for two different dosage levels need not be the same.

- (iii) At each time point there are no interactions between manufacturers and dosage levels.
- (iv) For each manufacturer there are no interactions between dosage levels and Time.
- (v) All six profiles are straight lines (there is no curvature across time), but the lines are <u>not</u> necessarily parallel or equal.
- (vi) The six profiles are <u>parallel</u> and <u>straight lines</u> but they are not necessarily equal.
- E. Consider a "mixed" model ANOVA for these data with fixed effects for dosage levels, manufacturers and time points and random effects for rabbits. Assume a compound symmetry covariance structure for repeated measurements taken on the same rabbit. Complete the following ANOVA table.

| Source of Variation | d.f. | Sum of d.f. Squares F | | p-values for the compound | Conserv- ative | Geisser- Greenhouse | p-values for G-G |
|---------------------|------|--------------------------|--|---------------------------|-------------------|------------------------|---------------------|
| | | | | symmetry model | df | df | df |
| | | | | · · | | | |

Corrected total

- F. Are conservative degrees of freedom needed in Part E? Explain.
- G. Run the REPEATED statement in PROC GLM with the POLYNOMIAL OPTION. What model is suggested by the output from this analysis?
- H. Repeat the analysis in part E (without conservative or Geisser-Greenhouse degreees of freedom) using an arbitrary (unstructured) covariance matrix for repeated measures on the same rabbit. Did your conclusions change? Which is the better analysis to use in this case? Give some justification for your answer.
- 5. Consider the data shown in problem 6.27 on page 346 in Johnson and Wichern. Each person gave responses on a five point scale (1=none, 2=very little, 3=some, 4=a great deal, 5=tremendous amount) to the following four questions:
 - 1. What is the level of passionate love you feel for your partner?
 - 2. What is the level of passionate love your partner feels for you?
 - 3. What is the level of companionate love you feel for your partner?
 - 4. What is the level of companionate love your partner feels for you?

The responses to these questions are denoted by X_1, X_2, X_3, X_4 , respectively. The data

are posted on the course web page as rating.dat.

(a) Suppose the data on page 346 came from a sample of n=30 married couples. Test the null hypothesis that the husband rating of his wife profile is parallel to the wife rating of her husband profile. Report

F=_____ df=_____ p-value=_____.

(b) Suppose the data on page 346 came from a sample of n=30 married couples. Test the null hypothesis that the husband rating wife profile is coincident with the wife rating husband profile. Report

F=_____ df=_____ p-value=_____.

(c) Suppose the data on husbands came from a sample of 30 married men and the data on wives came from an independent sample of n=30 married women. Test the null hypothesis that the husband rating wife profile is parallel to the wife rating husband profile. Report

F=_____ df=_____ p-value=_____.

(d) Suppose the data on husbands came from a sample of 30 married men and the data on wives came from an independent sample of n=30 married women. Test the null hypothesis that the husband rating wife profile is coincident with the wife rating husband profile. Report

F=_____ df=_____ p-value=_____.

(e) When married couples are sampled, often the husband's responses are positively correlated with the wife's responses. What are the consequences of ignoring this and incorrectly assuming that responses given by husbands and responses given by wives correspond to two independent random samples?

Additional problems from Chapters 5 and 6 in the text that you may find informative are problems 5.8, 5.9, 5.11, 5.13, 5.14, 5.15, 6.5, 6.6, 6.7, 6.8, 6.12, 6.13, 6.14. Most of these problems do not require extensive use of a computer.