MIDTERM

STA 207

Winter Quarter, 2015.

Please show all work

1. It is desired to estimate the mean magnesium concentration (μ) in the leaves of a species of plants. A random sample of r plants of this species is chosen at random and k leaves are randomly selected at random from each plant. Let $Y_{ij}, j = 1, \ldots, r, i = 1, \ldots k$, be the magnesium concentration for the j^{th} leaf of the i^{th} plant. A random effects model is considered appropriate for this study

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij},$$

where $\{\varepsilon_{ij}\}$ are iid $N(0, \sigma^2)$, $\{\alpha_i\}$ are iid $N(0, \sigma_1^2)$ and $\{\varepsilon_{ij}\}$ and $\{\alpha_i\}$ are independent. [Here α_i is the leaf effect.]

(a) Find the mean and variance of Y_{ij}. Express them explicitly in terms of the parameters of the model.

(b) Show that the correlation between Y_{ij} and $Y_{ij'}, j \neq j'$, is equal to $\sigma_1^2/(\sigma_1^2 + \sigma^2)$.

(c) Show that $\bar{Y}_{\cdot\cdot}$ is an unbiased estimate of μ. Find the variance of $\bar{Y}_{\cdot\cdot}$ and express it in terms of the parameters of the model.

(d) Find an unbiased estimate of $Var(\bar{Y}_{\cdot\cdot})$ and use this to construct a $(1-\alpha)100\%$ confidence interval for μ.

2. A large retail chain has three regional accounting centers for auditing. Each regional center employs three auditors who **serve only that center and no other**. One function of each center is to review whether a certain internal control operates properly in the processing of payroll. Data on the proportion of transactions (p) where the internal control was found to be operating properly were obtained from each auditor in each region for the previous two months. The response is $Y = (2 \arcsin \sqrt{p})100$ (this is done to stabilize the variance). The company is interested in finding out if there are differences in the centers and also if there are differences among the auditors in these centers. Let Y_{ijk} be the response at region i for j^{th} auditor in the k^{th} month. Here region is factor A and auditor is factor B. A summary of the data is give below.

Sample means for the three regions are: 141.300 (region 1), 150.283 (region 2), 155.783 (region 3).

$\sum_i \sum_j \sum_k (Y_{ijk} - \bar{Y}_{ij\cdot})^2 = 93.390$, $\sum\sum\sum(\bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot\cdot\cdot})^2 = 641.4344$, $SSTO = 1431.631$,

$\sum_j \sum_k (Y_{1j\cdot} - \bar{Y}_{1\cdot\cdot})^2 = 231.04$, $\sum_j \sum_k (Y_{2j\cdot} - \bar{Y}_{2\cdot\cdot})^2 = 424.643$, $\sum_j \sum_k (Y_{3j\cdot} - \bar{Y}_{3\cdot\cdot})^2 = 41.123$.

(a) Write down an appropriate model for this data. Explain the terms in the model along with the conditions on the parameters of this model.

(b) Set up the ANOVA table including the F-statistics and the approximate p-values.

(c) Are the regions different? Are there differences among the auditors? For each test, write down the null and the alternative hypotheses and carry out the test at $\alpha = 0.05$ level of significance.

(d) If you find the auditors are different, then carry out for each region a test to decide if there are differences among the auditors in that region. Thus there will be three tests. For each test, carefully write down the null and the alternative hypotheses and state your conclusions.

3. In a large university a random sample of 10 students were examined about their stress levels on a scale of 0 to 50 at three points of time: one week before, during and one week after the final exams. A summary is given below.

Source	df	SS	MS	F	p-val
Subject		654.3			
Week		204.8			
Error		273.2			
Total					

The mean stress levels for the three weeks are 26.3(before), 29.5(during), 23.1(after).

(a) Write down a model that is appropriate for the analysis of this data. Explain the terms in the model along with the conditions on the parameters of this model..

(b) Complete the ANOVA table. What are your conclusions?

(c) Obtain an estimate of the proportion of variability in stress that is due to subjects.

(d) Use Tukey's method to make pairwise comparisons in the means stress levels in the three weeks uing simultaneous 95% confidence intervals. What do you conclude?

4. A soft drink bottling company wishes to know effects of the following three factors on the response Y, the deviation from target fill:

factor A: percent carbonation (three levels; 10%, 12% and 14%),

factor B: operating pressure (two levels; 25psi and 30psi),

factor C: line speed (two levels; 200 and 250).

Two observations were obtained for every set of conditions, thus a total of 24 observations were recorded.

ANOVA Table

Source	df	SS	MS	F	p-val
A					
B		43.375			
C		22.042			
AB		5.250			
AC		0.583			
BC		1.042			
ABC		1.083			
Error		8.500			
Total		336.625			

(a) Complete the ANOVA table.

(b) Which terms in the three-factor model can be dropped? Assume a level of significance $\alpha = 0.05$.

(c) From part (b), you will find that some of the terms in a three factor model can be dropped. Write a model that excludes the terms that can be dropped.

(d) Set up the ANOVA table for the model you have written down in part (c).