STATISTICS 368
SAMPLE MIDTERM EXAM

Note: You should attempt these questions only after reviewing the relevant material. If you have difficulty on any of them, then you should take this as an indication of an area in which more review is required. Solutions will not be posted. You are most welcome to see me to check your solutions, etc.

Time: 50 minutes.

1. The following questions all require only very short answers.

   (a) Suppose that independent observations $Y_1, ..., Y_n$ are taken from a population with <u>a mean of zero</u> and a variance of σ^2. How can these be combined to form a random variable X which has a χ^2_n distribution? What then is the expected value of X?

   (b) An F-test is carried out, to test for the equality of the treatment effects following a completely randomized designed experiment. There were 4 treatments, the sum of squares SS_E was on 20 degrees of freedom, and the computed F-ratio was $F_0 = 6.1$. What is the p-value? Give your answer in terms of the probability that one thing is larger than, or smaller than, another. Illustrate it with a rough diagram of the relevant probability density, on which you mark the location of F_0 and indicate the desired probability.

   (c) We commonly use one of two procedures for making multiple comparisons among mean effects. Which of the two - Fisher's or Tukey's - maintains the <u>experimentwise</u> error rate at a fixed value, regardless of how many comparisons are made?

2. In order to compare automobile repair costs at the 2 repair shops in a certain locale, investigators collect a sample of 12 cars which have been in accidents and need repairs. Each car is taken to both repair shops, and estimates of the repair costs are obtained. In terms of

$$y_{ij} = \text{ estimated cost from shop } i \text{ to repair car } j \ (i = 1, 2, \ j = 1, ..., 12)$$

   what is the statistic that should be calculated in order to test the hypothesis that the mean repair costs are the same at the two shops?

3. Complete the following ANOVA table for a RCBD, in which each treatment appeared once in each block.

Source	SS	df	MS	F_0
Treatments	12	_____	_____	_____
Blocks	_____	4	6	
Error	36	_____	_____	
Total	_____	19		

4. (a) Suppose that you wish to investigate 3 fertilizer types ("A", "B", "C"), by applying them to crops in 3 regions of land. There are thought to be two nuisance factors at play - the regions themselves ("R1", "R2", "R3") and the farmers (Brown, Smith, and Doe). Assuming that these factors do not interact, write down a design which will allow for the estimation of all of their main effects, with only 9 crops being planted.

   (b) In a notation such as is used in class, write down the <u>effects model</u> for the data.

   (c) How many degrees of freedom will be associated with SS_E?

ADDITIONAL QUESTIONS FOR REVIEW

1. In order to test the null hypothesis that a randomly chosen coin from my pocket is properly balanced (so that it comes up heads or tails with equal probability), I toss the coin 5 times and obtain 5 heads. What is the p-value associated with my null hypothesis?

2. Consider the simple "location" model

$$Y_i = \mu + \varepsilon_i, \ i = 1, ..., n,$$

with independent random errors ε_i with means $= 0$ and variances $= \sigma^2$.

   (a) The sum of squares function for this model is

   $$S(\mu) = \sum (Y_i - E[Y_i])^2 = \sum (Y_i - \mu)^2 .$$

   What is the least squares estimate $\hat{\mu}$?

   (b) Prove your answer in a) by carrying out the decomposition of $S(\mu)$ into a sum of two non-negative terms, only one of which involves μ and which equals 0 when evaluated at $\hat{\mu}$.

3. For the analysis of a RCB designed experiment, we used the effects model

$$y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij} \qquad (*)$$

in which the τ_i are the treatment effects. Suppose that there are $a = 5$ treatments and $b = 6$ blocks. Our decomposition of the sum of squares $S(\mu, \boldsymbol{\tau}, \boldsymbol{\beta}) = \sum_{i=1}^{a} \sum_{j=1}^{b} \{y_{ij} - E[y_{ij}]\}^2$ was

$$S(\mu, \boldsymbol{\tau}, \boldsymbol{\beta}) = SS_E + ab(\mu - \hat{\mu})^2 + b\sum_{i=1}^{a} (\tau_i - \hat{\tau}_i)^2 + a\sum_{j=1}^{b} \left(\beta_j - \hat{\beta}_j\right)^2,$$

from which we obtained the LSEs and the minimum value of S under the 'Full' model (*). Suppose now that want to test the hypothesis that treatments 1 and 2 have no effect, while saying nothing about the other 3 treatments.

   (a) Write down the null hypothesis in terms of the τ_i's.

   (b) What is the $SS_{Reduced}$, the minimum value of S in the 'Reduced' model that results if the null hypothesis is assumed to be true?

   (c) What is F_0 to test this hypothesis?