

STATISTICS 568 SAMPLE MIDTERM EXAM

Instructions: Answer all five equally weighted questions.

Time: 80 minutes.

1. Consider a randomized complete block design to compare five treatments using eight blocks. Let y_{ij} denote the response for the j^{th} treatment in the i^{th} block. Write down a formula describing how you would obtain simultaneous 95% confidence intervals for all pairwise comparisons of treatment effects. Explain how estimates, standard errors, and critical values would be calculated and define any notation that you use. Be as specific as possible.
2. For each of the following experiments, consider the usual test for the null hypothesis of “no treatment effect”. Assume a standard model with additive treatment, block and (where appropriate) covariate effects. State the degrees of freedom determining the null distribution; i.e. $df1$ for the numerator and $df2$ for the denominator of the F statistic.
 - (a) A completely randomized design with 2 treatments, 10 replicates for the first treatment and 20 for the second.
 - (b) A 6×6 Latin square design, where row and columns are blocks.
 - (c) A randomized complete block design with 5 blocks and 8 treatments.
 - (d) An experiment comparing 4 diets with 5 men and 5 women assigned to each diet. Gender is a blocking factor and initial weight is a covariate (assuming parallel lines).
3. In a one way analysis, suppose the data follow the model

$$y_{ij} = \eta + \tau_i + \varepsilon_{ij} \quad (i = 1, \dots, k, j = 1, \dots, n),$$

with random errors $\varepsilon_{ij} \stackrel{i.i.d.}{\sim} N(0, \sigma^2)$ and random effects $\tau_i \stackrel{i.i.d.}{\sim} N(0, \sigma_\tau^2)$. Note that all k of the group sizes are the same ($= n$).

- (a) Write down the F-statistic which is used to test the hypothesis that there are no treatment effects, i.e. that $\sigma_\tau^2 = 0$. State the null distribution.
- (b) Derive (without assuming the truth of the hypothesis) the expected value of the mean square in the numerator of this F-statistic. *...over*

4. Recall the statement of Cochran's Theorem: *Suppose that the squared norm of a random vector $\mathbf{x} = (X_1, \dots, X_n)'$, with i.i.d. $N(0, \sigma^2)$ elements, is represented as $\mathbf{x}'\mathbf{x} = \sum_{i=1}^k \mathbf{x}'\mathbf{Q}_i\mathbf{x}$ for symmetric matrices \mathbf{Q}_i with ranks r_i , $i = 1, \dots, k$. Then the following are equivalent: (i) the random variables (r.v.s) $\mathbf{x}'\mathbf{Q}_i\mathbf{x}$ are independently distributed as $\sigma^2\chi_{r_i}^2$, $i = 1, \dots, k$; (ii) $\sum_{i=1}^k r_i = n$.*

In proving this we first established three preliminary results:

Lemma 1 The r.v. $\mathbf{x}'\mathbf{Q}\mathbf{x}$ is $\sim \sigma^2\chi_r^2$ iff $\mathbf{Q}^2 = \mathbf{Q}$ and $rk(\mathbf{Q}) = r$.

Lemma 2 The r.v.s $\mathbf{x}'\mathbf{Q}_1\mathbf{x}$, $\mathbf{x}'\mathbf{Q}_2\mathbf{x}$ are independently distributed iff $\mathbf{Q}_1\mathbf{Q}_2 = \mathbf{0}$.

Lemma 3 Suppose that $\mathbf{x}'\mathbf{Q}\mathbf{x} = \sum_{i=1}^k \mathbf{x}'\mathbf{Q}_i\mathbf{x}$ for symmetric matrices $\mathbf{Q}, \mathbf{Q}_1, \dots, \mathbf{Q}_k$, that $\mathbf{x}'\mathbf{Q}\mathbf{x} \sim \sigma^2\chi_r^2$, that $\mathbf{x}'\mathbf{Q}_i\mathbf{x} \sim \sigma^2\chi_{r_i}^2$ for $i = 1, \dots, k-1$, and that $\mathbf{x}'\mathbf{Q}_k\mathbf{x} \geq 0$ for all \mathbf{x} . Then $\{\mathbf{x}'\mathbf{Q}_i\mathbf{x}\}_{i=1}^k$ are independently distributed, and hence $\mathbf{x}'\mathbf{Q}_k\mathbf{x} \sim \sigma^2\chi_{r_k}^2$ with $r_k = r - \sum_{i=1}^{k-1} r_i$.

Complete the proof of Cochran's Theorem.

5. Suppose you are planning to study five different hardwood concentrations to determine their effect on the strength of the paper produced at a pilot plant. The experiment is to be carried out over five days. Uncontrollable factors affecting paper quality can vary from day to day, and the plant is limited to four production runs on each day. What experimental design would you recommend in this situation? Be specific: show how concentrations might be assigned to the 20 possible production runs over the five-day period and describe how the assignments would be randomized. Describe a desirable feature of your design.