Maria L. Joseph

STATISTICS 512, QUIZ #1, 9/29/06

71

This is a closed book, closed notes, "closed calculator and cell phone" test; you should have only a pencil. Please do all work on this paper and hand it in at the end of the class period; you may use the back of these pages or the additional blank sheets at the back of the exam, but be clear about which problem you are working. Unless otherwise specified, the notation used on this quiz is as defined in class, e.g. τ_i for the parameter associated with the ith experimental treatment in the "effects" model parameterization. If you have any doubt about your answers, put as much detail as possible in your solutions so that I can consider giving you partial credit. (I can't give partial credit for any incorrect answer that includes no information about how you derived it.)

59
-2
-3
-6
-14(3)

A.) Consider a Randomized Complete Block Design as discussed in class, with $v = 4$ treatments and $b = 5$ blocks.

$r_1 = r_2 = r_3 = r_4 = b = 5$

-1 1. The ANOVA-based test for equality of treatment effects requires assumptions that the random variables denoted by ϵ in the model are independent, mean zero, of equal variance, and are normally distributed (although as we've discussed, the last of these assumptions isn't as important as the others). What additional modeling assumption is necessary for the validity of this test?

$\sum_i \hat{\tau}_i = 0$ or $\hat{\tau}_4 = 0$

(sum zero constraint) (last effect = 0)

This isn't an assumption, it can be used as an additional constraint but isn't necessary.

✓ 2. Suppose that in fact,

$$\tau_1 = \tau_2 = +1,\ \tau_3 = \tau_4 = -1, \text{ and } \sigma = 2$$

where each of these parameters has the meaning we've used in class. What is the value of the noncentrality parameter associated with the test mentioned in part 1?

$\bar{\tau} = \frac{1+1-1-1}{4} = 0$

$Q(\tau) = 5(1-0)^2 + 5(1-0)^2 + 5(-1-0)^2 + 5(-1-0)^2$
$= 5 + 5 + 5 + 5$
$= 20$

$$\boxed{NCP = \frac{Q(\tau)}{\sigma^2} = \frac{20}{4} = 5}$$

✓

✓ 3. What is the power of the test under the conditions stated in part 2, if the probability of a Type I error is set to be 0.10? The answer to this question is $Prob\{X > c\}$, for the appropriate random variable X and constant c. Fully characterize (i.e., express all constants as numerical quantities):

- the distribution of X, and

$X \sim F'(3, 12, ncp = 5)$ ✓

- the value of c.

c = the 90^{th} percentile of the central F dsn w/ (3,12) df ✓

source	df
blk	4
trt	3
error	12
C.tot	19

21/25

C.) Consider an application in which "professional tasters" are asked to taste and evaluate 5 different kinds of brownies. All 5 types of brownies are made with the same basic ingredients, but there are differences in "recipes," and the effects of these differences are of primary interest in this study.

1. Experiment 1 was designed as a single-replicate Latin Square, involving 5 tasters and 5 batches of material, yielding a total of 25 evaluations (one kind of brownie, made from one of the batches, evaluated by one of the tasters). Assuming each evaluation leads to one data value, the degrees of freedom for an analysis of variance following this experiment are (fill in the blanks):

taster

batch

source of variation	degrees of freedom
tasters	4
batches	4
recipes	4
residual	12
corrected total	24

24 − 12 = 12

2. Experiment 2 was carried out to confirm the results of Experiment 1, and was designed as a three-replicate Latin Square. The same tasters were used in each of the replicates. The degrees of freedom for an analysis of this experiment are (fill in the blanks):

source of variation	degrees of freedom
replicates	2
tasters (rep)	3(5−1) = 12 X
batches	12
recipes	4
residual	44 X
corrected total	74

74 − 30 = 44

Based on the information in this table, what can you tell me about the physical batches of material used in this experiment?

Since the batches and ~~the~~ tasters nested in reps have the same df, if we detect a difference in batches, we will also detect a difference in tasters within a replicate

3. Instead of designing Experiment 2 as a replicated Latin Square, a Randomized Complete Block design might have been used with $b = 15$ blocks, a different taster employed for each block, and all 5 brownies in each block (one from each recipe) made from a very large (and somewhat less uniform) common batch of material. For any estimable contrast of the treatment parameters, this Randomized Complete Block design would yield 95% two-sided confidence intervals of smaller expected squared length than the replicated Latin Square if (fill in the blank with the appropriate expression of fully defined quantities):

RCBD	
source	df
block	14
trt	4
error	56
total	74

$$\sigma_{RCBD}/\sigma_{LS} < \frac{t(.975, 44\,df)\ \cancel{20/15}}{t(.975, 56\,df)\ \cancel{20/15}}$$

ok

15 blocks
5 trt

20/25

B.) Consider a blocked design in $v = 4$ treatments, in $b = 3$ blocks of size 2 each, as depicted below:

$n=8$

h		
1	1	2
2	1	3
3	1	4

$y_{hij} = \theta_h + \tau_i + \varepsilon_{hij}$

where the numbers represent the treatments assigned to each of the $n = 6$ units.

0k 1. Using "effects" notation (with θ's for blocks and τ's for treatments, but omitting μ since it is linearly redundant) write:

- X_1
- H_1
- X_2
- $H_1 X_2$

$H_1 = X_1(X_1'X_1)^{-}X_1'$

$$X_1 = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}_{6,3} \qquad X_2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}_{6,4}$$

$$H_1 = \frac{1}{2}\begin{bmatrix} J_{22} & & \\ & J_{22} & \\ & & J_{22} \end{bmatrix}_{66} \qquad H_1 X_2 = \frac{1}{2}\begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix}$$

-5 2. Is the last of these matrices the same as you would have seen with a Completely Randomized Design with $r_1 = 3$, $r_2 = 1$, $r_3 = 1$, and $r_4 = 1$? Regardless of whether your answer is "yes" or "no", state (very precisely) what this would mean about inferences based on the two designs.

$\boxed{\cancel{\text{Yes}}}$ no; more precise variances in the CRD design since more error df are available

(but not because of H_1X_2)

-4 3. Is $\tau_2 - \tau_3$ estimable under this design? Use your results from part (1) to prove your answer.

$\tau_2 - \tau_3 = (\underbrace{0\ 0\ 0}\ \underbrace{0\ 1\ -1\ 0}) \cdot \beta$

Since $0+0+0 = 0 = 0+1-1+0$, $(0\ 0\ 0\ 0\ 1\ -1\ 0)' \in \mathcal{R}(X)$

show this

$\Rightarrow \tau_2 - \tau_3$ is estimable

$\boxed{\text{yes}}$

16/25

D.) Consider a Completely Randomized Design for investigating $v = 3$ treatments, and for which the total number of units to be used is $n = 20$. The primary functions of interest in this experiment are:

$$\psi_1 = \tau_1 + \tau_2 + 3\tau_3,\ \psi_2 = \sqrt{5}\tau_1 - \sqrt{5}\tau_3, \text{ and } \psi_3 = -2\tau_1 + 4\tau_2 - 2\tau_3$$

-10

1. What must be true of the CRD so that each of the following is true? (If a statement can't be true regardless of the details of the CRD, indicate that this is the case.)
   - ψ_1 is estimable.
   - ψ_2 is estimable.
   - ψ_3 is estimable.

✓

2. If the units are appropriately randomized in this design, what is the number of equally probable assignments of units to treatment groups? Give your answer as a combinatorial expression in r_1, r_2, r_3, and any other defined quantity you need.

$$\frac{n!}{r_1!\, r_2!\, r_3!} = \boxed{\frac{20!}{r_1!\, r_2!\, r_3!}}$$

3. Derive the design (i.e. values of r_1, r_2, and r_3 that satisfy the total size requirement) that minimizes average (or sum) of variances for ψ_2 and ψ_3 only, using the method of Lagrangian Multipliers as follows:

   -1

   - Give the total of the variances of these two functions (this is easier to work with than the average):

$$\mathrm{Var}\ \psi_2 + \mathrm{Var}\ \psi_3 = \mathrm{Var}(\sqrt{5}\tau_1 - \sqrt{5}\tau_3) + \mathrm{Var}(-2\tau_1 + 4\tau_2 - 2\tau_3)$$
$$= \sigma^2\left(\frac{5}{r_1} + \frac{5}{r_3}\right) + \sigma^2\left(\frac{4}{r_1} + \frac{16}{r_2} - \frac{4}{r_3}\right)$$
$$= \boxed{\sigma^2\left(\frac{9}{r_1} + \frac{16}{r_2} + \frac{1}{r_3}\right)}\quad 9$$

   ✓

   - Remove the common factor of σ^2, and add something to the total variance to get the "augmented" objective function (i.e. the full expression which must be differentiated):

$$\phi = \frac{9}{r_1} + \frac{16}{r_2} + \frac{1}{r_3} + \lambda(r_1 + r_2 + r_3 - 20)\ ✓$$

   ok

   - Give the three equations in three unknowns (sample sizes) that follow from the augmented objective function:

$$\frac{\partial\phi}{\partial r_1} = -\frac{9}{r_1^2} + \lambda = 0 \qquad \frac{\partial\phi}{\partial r_3} = -\frac{1}{r_3^2} + \lambda$$

ok $$\frac{\partial\phi}{\partial r_2} = -\frac{16}{r_2^2} + \lambda = 0 \qquad \frac{\partial\phi}{\partial \lambda} = r_1 + r_2 + r_3 - 20 = 0 \Leftarrow \text{constraint}$$

   - Give the solution to the problem, i.e. the sample sizes the satisfy the constraint:

$$\boxed{r_1 = 3 \quad r_2 = 10 \quad r_3 = 7} \quad \text{or} \quad \boxed{r_1 = 2 \quad r_2 = 10 \quad r_3 = 8}$$

work on back of previous page

14/25