Amy Hoeksema

STATISTICS 512, QUIZ #2, 11/7/08

This is a closed book, closed notes test; you should have only a pencil (no calculator, cell phone, et cetera). 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 quiz, but be clear about which problem you are working. Please do not separate the stapled sheets of the quiz; this can easily lead to lost pages.

Unless otherwise specified, the notation used in this quiz is as defined in class, e.g. τ_i for the parameter associated with the ith experimental treatment in an "effects" model parameterization. If you have any doubt about an answer, put as much detail as possible into your solution so that I can consider giving you partial credit.

Give concise and numerically specific answers to all questions. For example, if I ask for degrees of freedom in a problem where you are given the specific characteristics of a design, "$N - b - t + 1$" and "$53 - 14 - 8...$" are not acceptable, but "9" may be. If a solution involves matrix short-hand notation, be sure to be complete in your answer, e.g. write "$\frac{1}{6}\mathbf{J}_{4\times 2}$" rather than "$c \times \mathbf{J}$". When distribution quantiles are needed, expressions like "$t(.99, 16)$" or "$t_{.99}(16)$" should be used.

$Q_1 = 39/70$

$Q_2 = 51/70$

$Q_3 = 63/70$

$\frac{51}{70}$

Look at #3

$\frac{3}{2} \cdot \frac{5}{2} = \frac{15}{4}$

-9

1. The following data were collected using a Balanced Incomplete Block Design with block size 2, in 3 treatments:

	trt 1	trt 2	trt 3
blk 1		25	15
blk 2	30		20
blk 3	28	32	
blk 4		27	13
blk 5	33		17
blk 6	30	30	
$\bar{\tau}_{i\cdot}$ =			

$k=2$ $t=3$

$\lambda = 2$
$r = 4$

Using a fixed-block model, MSE for these data is 5, but for the purposes of this problem, say it is 6.

(a) What is the standard error of $\widehat{\tau_1 - \tau_2}$?

$\text{Var}(\hat{c'\tau}) = \frac{k\sigma^2}{\lambda t} c'c = \frac{2(6)}{2(3)}(1^2 + (-1)^2) = 2(2) = 4$

$\Rightarrow$ Std error of $\tau_1 - \tau_2 = \sqrt{4} = 2$

(b) What would the standard error of $\widehat{\tau_1 - \tau_2}$ be for a Complete Block Design with the same number of units assigned per treatment, if MSE had the value of 8?

$\text{Var}(c'\hat{\tau}) = \frac{1}{b}\sigma^2 c'c = \frac{8}{4}(1^2 + (-1)^2) = 16$ $\quad b=4$

SE $= \sqrt{4} = 2$

Then Std error is $\sqrt{16} = 4$

(c) Give the data vector (y) and model matrix (X) for the *inter-block* analysis (i.e. with $N = 6$):

$\frac{k}{(r-\lambda)} = \frac{2}{4-2} = 1$

add them up → totals, not means

$\mathbf{y} = \begin{pmatrix} 40 \\ 50 \\ 60 \\ 40 \\ 50 \\ 60 \end{pmatrix}$ $\quad \bar{y} = 25$

$\mathbf{X} = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}$ $\quad N(i,j)$

intercepts (2)

what you get when you add all data values up in a block

$k=2$ $r=4$ $b=6$ $t=3$ $\lambda=2$

(d) What is the *inter-block* estimate of $\tau_1 - \tau_2$?

$\widehat{c'\tau} = \sum c_j Q_j = 1Q_1 - 1Q_2$

$= 36.25 - 25.5$

$= 10.75$

rows 1-3 are identical to rows 4-6

col rank of X is 3 $\Rightarrow \hat{y}$ must be $\bar{y}$

row 2 − row 1 $= \bar{\tau}_2 - \bar{\tau}_1 = 50 - 40 = 10$

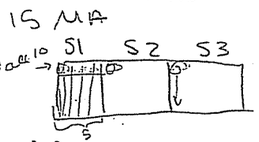
$w_1 = \frac{MSE_{inter}}{r-\lambda}$ $\quad w_2 = \frac{k \cdot MSE_{intra}}{\lambda t}$

*any time col rank is depleted, the fit will be "perfect" b/c data must agree w/ fitted values exactly

a=3, r=5, b=10 — 3 Sub — 10 Lt

2. A certain kind of microarray plate can be made using one of three different substrates (WP), and can be scanned using one of ten different intensities (SP) of light. In an array-development experiment, 15 microarrays were produced, 5 on each kind of substrate. All were treated with the same genetic material. Data were collected from each microarray using all 10 light intensities. Hence, data can be represented as $y_{i,m(i),j}$, where

- $i = 1, 2, 3$ for the substrate type
- $m = 1, 2, ..., 5$ for the individual microarray of a given substrate type
- $j = 1, 2, ..., 10$ for the light intensity levels used for each microarray



Suppose differences among arrays made with the same substrate can be regarded as random, with associated variance component $\sigma^2_{array} = 5$, and that differences among scans of the same array using the same light intensity level would be regarded as random with variance $\sigma^2_{scan} = 2$. Under the usual assumptions for this type of design:

(a) What are the numerator and denominator degrees of freedom for testing the hypothesis of no main effect associated with substrate type?

num = 3 − 1 = 2

denom = a(r − 1) = 3(5 − 1) = 3(4) = 12

(b) What are the numerator and denominator degrees of freedom for testing the hypothesis of no main effect associated with light intensity?

num = b − 1 = 10 − 1 = 9

denom = 149 − [18 + 9 + 12 + 2] (AB, B, WP res A)

= 149 − 41

= 108

(c) What is the noncentrality parameter governing the power of the light intensity main effect test if:

$E(\bar{y}_{..1}) = E(\bar{y}_{..2}) = E(\bar{y}_{..3}) = E(\bar{y}_{..4}) = E(\bar{y}_{..5})$, ← same 5

$E(\bar{y}_{..6}) = E(\bar{y}_{..7}) = E(\bar{y}_{..8}) = E(\bar{y}_{..9}) = E(\bar{y}_{..10})$,

$E(\bar{y}_{..1}) = E(\bar{y}_{..6}) + 0.4$ ⇒ $E(\bar{y}) = .2$

$Q = \sum n_i (\bar{T}_i - \bar{\bar{T}})$

$= 5 \times 3 [5(0 - .2)^2 + 5(.4 - .2)^2]$

$= 15 [5(.2)^2 + 5(.2)^2]$

$= 15 [5(.04) + 5(.04)]$

$= 15 (.2 + .2) = 15(.4) = 6$

noncentrality: $\frac{Q}{\sigma^2_{scan}} = \frac{6}{2} = 3$

-5

$2^4 = 16 \quad N = r2^f = 2(16) = 32$

3. Consider a 2^4 factorial experiment with $r = 2$ replicates, where we "know" $\sigma^2 = 1$.

-2 (a) Under the full model, what is $Var(\widehat{\mu_{2111} - \mu_{1111}})$?

what is c?

$c'M$ has elements of ± 2 in each of M for which the two tmts have opposite signs (any one w/ α)

$\alpha, (\alpha\beta), (\alpha\delta), (\alpha\beta\gamma), (\alpha\beta\delta), (\alpha\gamma\delta), (\alpha\beta\gamma\delta)$

$c'MM'c = 8 \cdot 2^2 = 32$

$Var = \frac{1}{32} \cdot 32 = 1$

Why N? $\frac{\sigma^2}{N} c'c = \frac{1}{32}(1^2 + (-1)^2) = \frac{2}{32} = \boxed{\frac{1}{16}}$ (MM')

full factorial... → independently estimating every cell means... add variances of cell means

I think this is why

-3 (b) Under a model containing only μ, main effects, and two-factor interactions, what is $Var(\widehat{\mu_{2111} - \mu_{1111}})$?

$p = 1 + 4 + 6 = 11$

$\left(\frac{\sigma^2}{N} c'c\right) p = \frac{1}{16}(11) = \boxed{\frac{11}{16}}$ (MM')

Do above but now only $\alpha, (\alpha\beta), (\alpha\gamma), (\alpha\delta)$

$c'M_1M_1'c = 4 \cdot 2^2$

$Var = \frac{1}{32} \cdot 16 = 1/2$

✓ (c) Under a model containing only μ and main effects, what is $Var(\hat{\alpha})$?

2.

$Var(\hat{\alpha}) = \frac{\sigma^2}{N} = \frac{\sigma^2}{r2^4} = \frac{1}{2(16)} = \boxed{\frac{1}{32}}$

✓ Nope (d) If $\hat{\alpha} = 2$, $\hat{\beta} = 2$, $\hat{\gamma} = 2$, $\widehat{(\alpha\beta\gamma)} = 1$, and all other effects estimates are exactly zero, what is the sum of squares for treatments? $N = r2^f$

$SST = N(\hat{\alpha}^2 + \hat{\beta}^2 + \hat{\gamma}^2 + \widehat{(\alpha\beta\gamma)}^2)$

$= 32[4 + 4 + 4 + 1]$

$= 32(13) = \boxed{416}$

32
13
96
320
416

✓ Got (e) Given the effect estimates in part (d), what additional effects, if any, would need to be added to the model to satisfy the Heredity Principle? $\alpha\beta\gamma$

Heredity: At least one of the two way interactions, $\alpha\beta$, $\alpha\gamma$ or $\beta\gamma$; picking anyone would work

(-5)

4. Consider an experiment designed as a (fairly odd) BIBD, with $k = 1$ units per block, $t = 3$ treatments, and $b = 9$ blocks.

-5 (a) Under the usual fixed-block model, no linear combination of the treatment parameters is estimable. Using only words (no math), clearly explain in one or two sentences why this is so.

All variation among observations can be attributable to blocks; after "correction" for blocks, there is no variation left to model

Because there is confounding of blocks + tmts; we can't make the assumption of no block x tmt interaction because we don't know where the influence is coming from

✓ (b) If the problem is changed by adding one very powerful modeling assumption, all contrasts in the treatment parameters become estimable. What assumption is that?

✓

Assume the blocks are random - types, then we can assume no block x tmt interaction + get estimable functions for all contrasts.