Alana L. Joseph

STATISTICS 512, QUIZ #2, 11/3/06

72

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.)

All of Part D?

A. As part of an industrial safety program, an experiment was carried out to investigate factors that effect the frequency of workplace accidents. The two factors of particular interest in this study were the training program used to educate workers about workplace safety, and the fatigue level of the worker. The experiment was designed to evaluate the effects of three different training programs, and five different levels of fatigue defined as 2, 4, 6, 8, and 10 continuous hours of work.

Thirty randomly chosen new employees were recruited into the study, and each employee was randomly assigned to one of the training programs. This randomization was restricted so that the number of employees enrolled in each training program was ten. After training was completed, each employee was given a series of five simulated job tasks to complete over the course of a 10-hour workday. Tasks took only a few minutes to complete, and were carried out after 2, 4, 6, 8, and 10 hours of work (that is, corresponding to the five levels of the fatigue factor). Each employee's performance was scored by a trained observer for the likelihood that an accident might occur. (No real accidents could result due to the way in which the tasks were simulated.) So, at the end of the experiment, five data values corresponding to the five levels of fatigue were associated with each employee.

✓ 1. Write a statistical model for the data generated in this experiment, including effects for safety program and fatigue and their interaction, and for employees and measurement error. Carefully index all terms and explain the indexing system including the range of each index. Indicate which model terms must be considered as random effects if the primary point of the experiment is to understand the marginal and joint effects of training program and fatigue.

α_i = safety program i; $i = 1,2,3$

$\beta_{j(i)}$ = j^{th} employee in the i^{th} program; $j = 1,\ldots,11$

γ_k = k^{th} level of fatigue; $k = 1,2,\ldots 5$

$$y_{ij(i)k} = \mu + \alpha_i + \beta_{j(i)} + \gamma_k + (\alpha\gamma)_{ik} + \varepsilon_{ij(i)k}$$

where $\beta_{j(i)}$ and $\varepsilon_{ij(i)k}$ are random effects ✓

-3 2. Based on your model, complete the following analysis of variance table. Fill in the blanks with appropriate number (not formulae) in the degrees-of-freedom column, and with appropriate formulae (not numbers) in the sums-of-squares column. You may write "difference" for one line in the sums-of-squares column if that entry can be indirectly calculate after the others are known.

source of variation	degrees of freedom	sum of squares
Safety Program (SF)	2	$\sum_i 50(y_{i\cdot\cdot} - y_{\cdot\cdot\cdot})^2$
Fatigue (F)	4	$\sum_k 30(y_{\cdot\cdot k} - \bar{y}_{\cdot\cdot\cdot})^2$
SF-by-F Interaction	8	$\sum_{i,k} X(\bar{y}_{i\cdot k} - \bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot k} - \bar{y}_{\cdot\cdot\cdot})^2$
W.P. error → Employees	27	
S.P. error → Residual	108	difference
Corrected Total	149	$\sum (y_{ij(i)k} - \bar{y}_{\cdot\cdot\cdot})^2$

-8

✓ 3. What is the reference distribution (i.e. the distribution from which the critical value would be calculated) for testing for the presence of:

- the fatigue main effect $F(4, 108)$
- the training program main effect $F(2, 27)$
- the fatigue-by-training program interaction $F(8, 108)$

5 4. Suppose instead of recruiting thirty random employees for this study, the investigator had recruited three random employees from each of 10 non-overlapping age groups. There is some belief that age might affect performance in the simulated tasks, but that this effect would be additive (i.e. would not interact with the treatments of interest). Briefly, describe how (if at all) you would modify the assignment of treatments to units, and the analysis of these data, as a result of this change. 7

I would add an effect for blocks (i.e θ_h, $h=1,\ldots,10$), and the rest of the model would remain the same since blocks (i.e age groups are additive).

$$\Rightarrow \quad y_{hij(i)k} = \theta_h + \alpha_i + \beta_{j(i)} + \gamma_k + (\alpha\gamma)_{ik} + \varepsilon_{hij}$$

This would change the ANOVA table df to:

Source	df
age group	9
SF	2
Employee	27
F	4
SF	8
Error	99
C. total	149

50

OK

$v=5$

B. A researcher wants to conduct a blocked experiment to compare five treatments, but operational constraints require that the blocks be of size three. $k=3$

$b=5$

✓ 1. She likes the idea of performing the experiment in five blocks since that would allow her to assign each treatment to three different units in the experiment. Can a BIBD be for this experiment be constructed under these conditions? (Prove your answer.)

not necessary but sufficient

$r=\frac{bk}{v}=\frac{3b}{5}=\frac{15}{5}=5\equiv$ Integer ✓

$\lambda=\frac{r(k-1)}{v-1}=\frac{3(2)}{4}=\frac{10}{4}\not\equiv$ Integer X

OK

$\binom{v}{k}=\frac{5!}{3!2!}=\frac{5\cdot4}{2\cdot1}=10$

(b)	1	2	3	4	5
1	x	x	x		
2					
3					
4					
5					

✓ 2. Write a BIBD design for this experiment using ten blocks (not five). Write the full design by filling in treatment numbers 1-5 in the blanks below.

$b=10$

$\Rightarrow r=\frac{10\cdot3}{5}=6$

$\lambda=\frac{6(2)}{4}=3$

block

1	2	3	4	5	6	7	8	9	10
1	1	1	1	1	1	2	2	2	3
2	2	2	3	3	4	3	3	4	4
3	4	5	4	5	5	4	5	5	5

? -2

?

3. Suppose we think of this experiment with reference to the model:

$$y_{ij}=\theta_i+\tau_j+\epsilon_{ij}.$$

where y is the response, θ's are block effects, and τ's are treatment effects. Suppose that, in fact (although we don't know it as experimenters),

$$\tau_1=\tau_2=-1,\quad \tau_3=0,\quad \tau_4=\tau_5=+1,\quad \sigma=2 = \boxed{\sigma_{CRD} \text{ or } \sigma_{BIBD}}$$

Doesn't matter, just divide once, not twice.

$$\frac{v/(v-1)}{k/(k-1)}$$

Given this information, completely characterize the distribution of the F statistic that would be used to test for equality of treatments. (That is, give the distribution along with the numerical values of all parameters.)

$Q(T)=r\sum_j(\tau_j-\bar{\tau}_.)^2=6\sum_j(\tau_j)^2=6(1+1+0+1+1)=24 \;/\sigma^2$

$\bar{\tau}_.=\frac{0}{5}=0$

$\frac{v(v-1)}{k(k-1)}=\frac{5(4)}{3(2)}=\frac{20}{6}$

$\frac{20}{6}\cdot24\cdot\frac{1}{\sigma^2}=80/4=20$

$F'(4,16,20)$ ← NCP

why is this right?

-4

4. Continuing with the part three of this problem, suppose the researcher had been able to execute a randomized complete block design in ten blocks, rather than the BIBD we've been discussing. Note that this would have been a larger experiment, since each block would contain five units. Using the same model information provided in the last part of this problem, give a complete characterization of the distribution of the F statistic that would be used to test for equality of treatments in this case.

so don't use N=30, use N=50?

$\frac{Q(T)}{\sigma^2}=\frac{24}{4}=6$

$F'(4,16,6)$ ← NCP

denom. df = 50 - 9 - 4 - 4

-6

why not $F(4,36,10)$?

C. A chemical engineer would like to investigate the effects of four factors – acid strength, reaction time, amount of acid, and reaction temperature – on the yield of a process. He is primarily interested in knowing the overall importance of each of the factors and their interactions, so he decides that an experiment in which each factor is represented at two levels is adequate for his purposes. Suppose he decides to replicate each of the 2^4 treatments $r = 3$ times (that is, each "cell" in his 5-dimensional data table will contain 3 numbers), and that he completely randomizes treatments to units (e.g. no blocks or split plots). Suppose also that he is willing to bet, based on his knowledge of the chemical kinetics involved in his process, that there are *no* interactions involving the first factor, acid strength, but that all other factorial effects should be considered.

α_h β_j γ_k δ_l

Answer each of the following questions assuming the usual parameterization we've used in class (i.e. model matrix of ± 1 elements, et cetera).

-1

1. What is the variance of the least-squares estimate of the main effect for reaction time? (Write this as $\sigma^2\times$ "a number", where you give me the value of "a number".)

$$Var(\hat{\beta}) = \frac{\sigma^2}{N} = \sigma^2 * \frac{1}{r2^4} = (?)\ \sigma^2\frac{1}{3\cdot 2^4} = \frac{1}{48}\sigma^2$$

-5

2. What is the variance of the least-squares estimate of the contrast in which:

need to write out M?

- all cell means at high acid strength and high reaction time are weighted +1 $\frac{\sigma^2}{r}$
- all cell means at low acid strength and high reaction time are weighted -1 $\rightarrow \frac{\sigma^2}{r}$
- all cell means at high acid strength and low reaction time are weighted -1 $\rightarrow \frac{\sigma^2}{r}$
- all cell means at low acid strength and low raction time are weighted +1 $\rightarrow \frac{\sigma^2}{r}$

2x this many...?

$$Var(\widehat{contrast}) = 0$$

contrast proportional to $(\alpha\beta)$, assumed known

-8

3. Suppose the fitted model is:

$$y_{ijkl} = 3 + 2x_1 - 2x_2 + 1x_3 - 1x_4 + 1x_2x_3 + 0x_2x_4 + 1x_3x_4 - 0x_2x_3x_4$$

fitted param = 9

where x's are ± 1 as discussed in class, and suppose that the sum of squared residuals for this model is 60. Compute values of the indicated components of the F statistic for testing:

$$Hyp_0 : \beta\gamma = \beta\delta = \gamma\delta = 0$$

- numerator sum-of-squares: $48(1^2 + 0^2 + 1^2) = 96$
- numerator degrees-of-freedom: 3 ✓
- denominator sum-of-squares: 60
- denominator degrees-of-freedom: $17 - 9 = 8$ ✗

$N\hat{\theta}_2'\hat{\theta}_2 = 48(\beta\gamma^2 + \beta\delta^2 + \gamma\delta^2)$
$= 48(x_2x_3^2 + x_2x_4^2 + x_3x_4^2)$
$= 48(1^2 + 0^2 + 1^2) = 96$

or $2^4(r-1) = 32$ for full model

N − # fitted param

$48 - 9 = 39$ $48 - 9 = 39$ for this model

-14