Results of survey on whether students want to see all the stat. I show on tests/grades
Out of 24 responses,
Yes : 19
No : 3
No opinion : 2

Name: ___________________________

ID: _____________________________

8 + 17.5

Stat 421, Test 2, Fall, Nov. 9, 2018; Marzban

ONLY a half-size "cheat sheet" is allowed $\boxed{\text{80 minutes}}$

Multiple choice: Circle all the correct answers; there is wrong-answer penalty

For rest, SHOW answer & work; NO CREDIT for correct answer without explanation

Points

radar example

1 **1.** We have come across a data set regarding the amount of vibration inside buses. The company that collected the data didn't have any experts in experimental design, and so this is what they did: They asked a driver to drive one bus while they randomly changed two factors: one 2-level factor (A) and one 3-level factor (B). For each of these 6 combinations, they recorded the vibration. Then, they asked the same operator to drive another bus, and again took 6 measurements in random order. Finally, someone commented that the driver may have an effect on the vibrations, and so the company repeated everything using the same procedure but now with a different driver. Based on all information available to you, it's not even clear why the company did this experiment. What is the design of this experiment?

a) CRD b) RCBD with 1 block factor
c) RCBD with 2 block factors d) RCBD with 3 block factors

lect 14, p.4

1 **2.** In the previous question, given the design, circle the factor(s) that can be considered treatment factor(s), and hence tested.

a) A b) B c) Driver d) Bus e) None of the above

lect 9, p.6

1 **3.** In an RCBD, the sample variance of the block-conditional-means (i.e.,means within blocks)
a) is zero b) is equal to that in a CRD c) measures the effect of blocking d) none of the above.

$\begin{pmatrix} A & B & C & \cdots \\ B \\ C \end{pmatrix} \begin{pmatrix} \alpha & \beta & \gamma & \cdots \\ \beta \\ \gamma \end{pmatrix} = \begin{pmatrix} A\alpha & B\beta & \cdots \\ B\beta \end{pmatrix}$

1 **4.** Which of the following statememts is/are FALSE?

a) There is only 1 standard LS of order 3. b) Two Standard LS cannot be orthogonal.
lect. and book → c) There are no orthogonal LSs of order 6. d) One can have a LS with no orthogonal counterpart.
e) None of the above.

$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \\ 3 & 1 & 2 \end{pmatrix}$ ↖ lab 6 and hw

spread out

1 **5.** For an additive model $y_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ijk}$, which of the following is FALSE?

a) $E[\hat{y}_{ijk}] = \mu + \alpha_i + \beta_j$ b) $E[y_{ijk}] = \mu + \alpha_i + \beta_j$ c) $E[e_{ij.}] = 0$ d) non of the above.

"$E[\hat{\mu} + \hat{\alpha} + \hat{\beta}]$" unbiased

follows from a, b.

hw-lect 13-3

1 **6.** Which of the following is FALSE? The residual plot (residuals vs. predictions) should
a) display no correlation for a full model.
b) be centered around the horizontal line for a full model.
c) display no correlation for an additive model.
d) be centered around the horizontal line for an additive model.

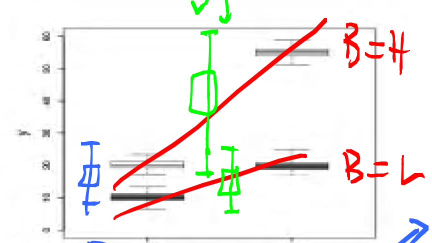
See hw-lect 13-3 and the bottom of next page, below.

hw-lect 15-4, lect 13

1 **7.** The adjacent figure shows the boxplots of data y involving two binary factors A and B. The clear (filled) boxplots correspond to the high (low) level of the B factor. Based on this data, circle all the effects that appear to exist.

a) A effect b) B effect c) AB effect
d) BA effect e) none of the above

B effect



A effect

hw-lect 15-1

1 **8.** Fill in the blanks: In 2^2, given the Contrast vector for A, and the Contrast vector for B, one can can compute the Contrast vector for AB.

hw-lect9-3, part c

~~2~~ 1.5

9. For data from a replicated CRD with two factors A and B, we have the following two models:
Model 1: $y_{ijk} = \mu + \alpha_i + (\alpha\beta)_{ij} + \epsilon_{ijk}$; $\hat{(\alpha\beta)}_{ij} = \overline{y}_{ij.} - \overline{y}_{i..}$; $\hat{y}_{ijk} = \overline{y}_{ij.}$
Model 2: $y_{ijk} = \mu + \alpha_i + \epsilon_{ijk}$; $\hat{y}_{ijk} = \overline{y}_{i..}$

Show that at the least-squares estimates, $SSE2 - SSE1 = \sum_{ijk} \hat{(\alpha\beta)}_{ij}^2$.

Note that I have not given $\hat{\alpha}_i$. So, $\cancel{SSA_1} + SSE_1 = SSA_2 + SSAB + SSE_2$ doesn't work, because we don't know $SSA_1 = SSA_2$.

$$SSE_2 - SSE_1 = \sum_{ijk} (y_{ijk} - \overline{y}_{i..})^2 - \sum_{ijk} (y_{ijk} - \overline{y}_{ij.})^2$$

$$= \sum_{ijk} \left[(\cancel{y_{ijk}} - \overline{y}_{i..}) - (\cancel{y_{ijk}} - \overline{y}_{ij.})\right]\left[(y_{ijk} - \overline{y}_{i..}) + (y_{ijk} - \overline{y}_{ij.})\right]$$

$$= \sum_{ij} (\overline{y}_{ij.} - \overline{y}_{i..}) \underbrace{\sum_k (2y_{ijk} - \overline{y}_{i..} - \overline{y}_{ij.})}_{2y_{ij.} - n\overline{y}_{i..} - n\overline{y}_{ij.} = 2n\overline{y}_{ij.} - n\overline{y}_{i..} - n\overline{y}_{ij.} = n(\overline{y}_{ij.} - \overline{y}_{i..})} = \sum_{ij} n(\overline{y}_{ij.} - \overline{y}_{i..})^2 = \sum_{ijk} \hat{(\alpha\beta)}_{ij}^2$$

	C_1	C_2	C_3
A_1	B_1	B_2	B_3
A_2	B_2	B_3	B_1
A_3	B_3	B_1	B_2

hw-lect11-2

~2

10. Consider the model $y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \epsilon_{ijk}$ for the LSD shown. **Using** $\sum'_{ijk} y_{.j.} = 3y_{...}$ (and its analog for $\sum'_{ijk} y_{i..}$ and $\sum'_{ijk} y_{..k}$) show that $\sum'_{ijk}$ of the predictions $\hat{y}_{ijk} = \overline{y}_{i..} + \overline{y}_{.j.} + \overline{y}_{..k} - 2\overline{y}_{...}$ is equal to $\sum'_{ijk}$ of the observations.

$$\sum{}' \hat{y}_{ijk} = \sum{}' (\overline{y}_{i..} + \overline{y}_{.j.} + \overline{y}_{..k} - 2\overline{y}_{...}) = \frac{1}{3}\underbrace{\sum{}' y_{i..}}_{3y_{...}} + \frac{1}{3}\underbrace{\sum{}' y_{.j.}}_{3y_{...}} + \frac{1}{3}\underbrace{\sum{}' y_{..k}}_{3y_{...}} - \frac{2}{\cancel{9}} y_{...} \cancel{\sum{}'}$$

$$= 3y_{...} - 2y_{...} = y_{...} = \sum{}' y_{ijk}$$

without orthogonality, the cross-terms prevent us from testing a given effect.

~~2~~ 1

11. In the previous problem, suppose we are now supposed to design an experiment that includes another 3-level nuisance factor. Specify the necessary runs if we want to be able to perform tests.

$$B \quad \begin{array}{c|ccc} & C_1 & C_2 & C_3 \\ A_1 & 1 & 2 & 3 \\ A_2 & 2 & 3 & 1 \\ A_3 & 3 & 1 & 2 \end{array}$$

constructing orthogonal matrix for GLSD $\rightarrow$

$$\begin{array}{c|ccc} & C_1 & C_2 & C_3 \\ A_1 & 1 & 2 & 3 \\ A_2 & 3 & & \\ A_3 & & & \end{array} \rightarrow \begin{array}{c|ccc} & C_1 & C_2 & C_3 \\ A_1 & 1 & 2 & 3 \\ A_2 & 3 & & \\ A_3 & 2 & & \end{array} \rightarrow \begin{array}{c|ccc} & C_1 & C_2 & C_3 \\ A_1 & 1 & 2 & 3 \\ A_2 & 3 & 1 & 2 \\ A_3 & 2 & 3 & 1 \end{array} \quad D$$

~~2~~ 1.5

12. Here is one more interpretion of the interaction term: Consider the model $y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}$. To simplify the language, let's say that $\hat{\alpha}_i$ is the effect of A; etc.. Show that $\hat{(\alpha\beta)}_{ij}$ can be written in terms of the conditional effect of A (given B=j), and the effect of A. This takes 1 line of algebra!

The cond'l part comes from Ch.6.

$$\hat{(\alpha\beta)}_{ij} = \overline{y}_{ij.} - \overline{y}_{i..} - \overline{y}_{.j.} + \overline{y}_{...} = \underbrace{(\overline{y}_{ij.} - \overline{y}_{.j.})}_{\text{Cond'l effect of A given B=j}} - \underbrace{(\overline{y}_{i..} - \overline{y}_{...})}_{\text{A effect}}$$

#6 More detail:

a) Cov[residual, prediction] = 0 for full model $\Rightarrow$ no corr. $\Rightarrow$ True

b) $e_{...} = 0$ for full model $\Rightarrow$ centered about horiz. line. $\Rightarrow$ True

c) Cov[,] $\neq 0$ for add. model $\Rightarrow$ There will be a corr. $\Rightarrow$ False

d) $e_{ijk} = y_{ijk} - \overline{y}_{i..} - \overline{y}_{.j.} + \overline{y}_{...} \Rightarrow e_{...} = 0 \Rightarrow$ centered about horiz. line. $\Rightarrow$ True

The fact that $e_{ij.} \neq 0$ for add. model means that at each $\hat{y}_{ij}$ on the x-axis, the residuals will be generally above, or below, the horiz. line. But the residual plot is still centered no hw-lect14-1 (w/o μ)

$\sim \cancel{2}$ 1.5

13. Consider the model $y_{ijk} = \alpha_i + (\alpha\beta)_{ij} + \epsilon_{ijk}$, with $i, j, k = 1 \cdots a, b, n$. Note that there is no μ.
a) Starting from the expression for SSE, derive the least-squares equations, and write them in dot-bar notation (i.e., conditional means). DO NOT impose/introduce any constraints.

Models w/o intercept are important in problems where Theory/fact/physics/etc requires the intercept to be zero.

$$SSE = \sum_{ijk} (y_{ijk} - \alpha_i - (\alpha\beta)_{ij})^2$$

$$\frac{\partial}{\partial \alpha_i}\Big|_{\hat\alpha, \hat{\alpha\beta}} \sim \sum_{jk} (y_{ijk} - \hat\alpha_i - (\hat{\alpha\beta})_{ij}) \sim \bar{y}_{i..} - \hat\alpha_i - (\hat{\alpha\beta})_{i.} = 0 \quad \mathrm{I}$$

$$\frac{\partial}{\partial (\alpha\beta)_{ij}}\Big|_{\hat\alpha, \hat{\alpha\beta}} \sim \sum_{k} (y_{ijk} - \hat\alpha_i - (\hat{\alpha\beta})_{ij}) \sim \bar{y}_{ij.} - \hat\alpha_i - (\hat{\alpha\beta})_{ij} = 0 \quad \mathrm{II}$$

$\sim \cancel{2}$ 1 — recorded jointly

hw-lect14-1

b) Recall that "uniquely estimable effects" are those that can be estimated from data without/before imposing constraints. Are there any such effects involving both $\hat\alpha_i$ and $(\hat{\alpha\beta})_{ij}$ (and/or their sums)? If so, write at least one. If not, explain why not. (If you don't know, go to part d.)

$$\hat\alpha_i + (\hat{\alpha\beta})_{ij} = \bar{y}_{ij.}$$

$\sim \cancel{2}$ 1

" c) Are there any such effects involving **only** $(\alpha\beta)_{ij}$ (and/or its sums)? If so, write at least one. If not, explain why not. (If you don't know, go to part d.)

$$\mathrm{II} - \mathrm{I} \quad (\hat{\alpha\beta})_{ij} - (\hat{\alpha\beta})_{i.} = \bar{y}_{ij.} - \bar{y}_{i..}$$

$\sim \cancel{1}$ 2 d) The number of parameters in the model (excluding σ_ϵ) is $\underbrace{a}_{\alpha_i} + \underbrace{ab}_{(\alpha\beta)_{ij}}$.

~ 1 e) The number of independent least-squares equations in part a is $\underbrace{a}_{\mathrm{I}} + \underbrace{ab}_{\mathrm{II}} - a \leftarrow \sum_j \mathrm{II} = \mathrm{I}$

$\sim \cancel{1}$ 2 f) Specify a natural set of constraints, and estimate all the effects.

standard

We need $\underline{a}$ constraints, e.g. $(\hat{\alpha\beta})_{i.} = 0 \xRightarrow{\mathrm{I},\mathrm{II}} \hat\alpha_i = \bar{y}_{i..}, \; (\hat{\alpha\beta})_{ij} = \bar{y}_{ij.} - \bar{y}_{i..}$

~ 2 g) Perform the ANOVA decomposition of $SST = \sum_{ijk} y_{ijk}^2$ into SSA, SSAB, and SSE. You may assume the cross-terms are zero.

$$y_{ijk} = \underbrace{\bar{y}_{i..}}_{\hat\alpha} + \underbrace{(\bar{y}_{ij.} - \bar{y}_{i..})}_{\hat{\alpha\beta}} + \underbrace{(y_{ijk} - \bar{y}_{ij.})}_{\text{left-over}}$$

$$SST = \sum_{ijk} y_{ijk}^2 = \sum_{ijk} \bar{y}_{i..}^2 + \sum_{ijk} (\bar{y}_{ij.} - \bar{y}_{i..})^2 + \sum_{ijk} (y_{ijk} - \bar{y}_{ij..})^2 + \underbrace{0}_{\text{cross-terms.}}$$

~ 2 h) Write the corresponding *df* for each term (SST, SSA, SSAB, SSE). Don't explain.

$$\sum_{ijk}^{abn} y_{ijk}^2 = bn \sum_{i}^{a} \bar{y}_{i..}^2 + n \sum_{ij}^{ab} (\bar{y}_{ij.} - \bar{y}_{i..})^2 + \sum_{ijk}^{abn} (y_{ijk} - \bar{y}_{ij.})^2$$

$$df: \quad abn = a + (ab - a) + (abn - ab)$$

$\sum_j^b (\bar{y}_{ij.} - \bar{y}_{i..}) = 0$ — $\underline{a}$ constraints

$\sum_k^n (y_{ijk} - \bar{y}_{ij.}) = 0$ — $\underline{ab}$ constraints