Name: ________________________

ID: __________________________

Stat 421, Test 2, Fall, Nov. 16, 2015; Marzban

9+25

ONLY a half-size "cheat sheet" is allowed

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

hw-T, X

1 **1.** Which of the following statements is/are true regarding contrasts?

a) **All** comparisons of μ_i can be written in terms of zero-sum contrasts constructed from the μ_i.

b) A specific comparison between several μ_i, constructed from a zero-sum contrast, can be written in terms of a zero-sum contrast constructed from the corresponding effects $\alpha_i = \mu - \mu_i$.

c) For a treatment levels, there exists a **unique** set of $(a-1)$ orthogonal contrasts.

d) If an anova F-test has led to the rejection of the null hypothesis of equal means, then testing a set of orthogonal contrasts can help in identifying a specific combination of means that is "responsible" for the rejection.

0.5 each

hw-Z

~ 1 **2.** In using the maximum likelihood criterion for estimating model parameters, if we change the constraints, then ______ of the parameter estimates will change.

a) none b) some c) all

hw-Z

~ 1 **3.** In using the maximum likelihood criterion for estimating model parameters, if we change the constraints, then ______ functions of the parameters will change.

a) no b) some c) all

Lect 11

1 **4.** Suppose we have developed the model $y_{ij} = \mu + \alpha_i + \epsilon_{ij}\ i = 1 \cdots a,\ j = 1 \cdots n$ based on a CRD. Circle all of the following quantities that will be different in an RCBD model $y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}$ of the same data.

~~0.5 each~~

a) SST b) SSA c) SSE

$SSA = \sum_i^a (\bar{Y}_{i.} - \bar{Y}_{..})^2$ in both models.

Lect 11

~ 1 **5.** In the model $y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}\ i = 1 \cdots a,\ j = 1 \cdots b$ based on an RCBD design, how does F_A generally change with increasing b?

a) Generally decreases b) Generally remains constant c) Generally increases d) None of the above.

(*)

6. In which of the following models should we be concerned if the **"grand mean" of the residuals** turns out to be nonzero?

all models have $e_{....} = 0$.

2 a) $y_{ij} = \mu + \alpha_i + \epsilon_{ij}$ b) $y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}$ c) $y_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ijk}$

d) $y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \epsilon_{ijk}$ e) $y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \epsilon_{ijk}$

0.5 each, max = 2

Lect 13, 14

1 **7.** In which of the following models should we not be surprised if some "conditional mean" of the residuals turns out to be nonzero?

0.5 each

a) $y_{ij} = \mu + \alpha_i + \epsilon_{ij}$ $e_{i.} = 0$ b) $y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}$ $e_{i.} = 0$ c) $y_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ijk}$ $e_{ij.} \neq 0$ Lect 14

d) $y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \epsilon_{ijk}$ $e_{ij.} = 0$ e) $y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \epsilon_{ijk}$ $e_{ij.} \neq 0$ (lect 13)

hw-AO

~ 1 **8.** In a problem with quantitative factors, one can develop a regression model that has a comparable number of parameters as the anova model. This statement is

a) always true, period.

or b) true, but the highest possible order depends on the number of factors in the problem.

c) true, but the highest possible order depends on the number of levels in the factors.

d) never true, period.

5) (*)

$$F = \frac{MSA}{MS_E} = \frac{b \sum_i (\bar{Y}_{i.} - \bar{Y}_{..})^2 / (\cancel{a-1})}{\sum_{ij}^{ab} (Y_{ij} - \cdots)^2 / ab(n-1)} \sim b(n-1)$$

The more levels in B, the easier to find A effect!

Yadav example

9. We are preparing to conduct a study to find if IQ depends on gender and on whether or not one wears glasses. Due to practical limitations, the data must be collected on 4 different days, and in 4 different cities, and it will be impossible to randomize across the days and the cities.

~ 2 **a)** In words, what kind of design is most appropriate if we have sufficient resources to conduct 4^3 runs? Make sure you also define the treatment factors, and the block factors, if any.

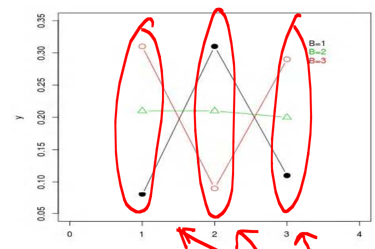
RCBD factorial, with Gender (B/G) and Glasses (Yes/No) as 2 treatment factors, and Day and City as 2 block factors.

~ 2 **b)** What if we have sufficient resources to conduct only 16 runs?

LSD with Day and City as 4-level row and col. factors, and combination of Gender and Glasses (B,Y), (B,N), (G,Y), (G,N) as the 4-level treatment factor.

Lect 13,15

10. Consider the data (shown in the adjacent figure) on the response y and two 3-level factors A and B.



~ 2 **a)** Suppose, initially we are not given the information on B at all, i.e., we have no idea that it even exists. Do we expect to find that A has an effect on the response? Yes/No, explain in words.

No.
without any info on B, there is too much overlap over the data in the 3 categories/populations.

$\sim$ ~~2~~ 1 **b)** Now, suppose it is revealed to us that B affects the data as shown in the figure. If we fit a model that continues to ignore B, do we expect to find that A has an effect? Yes/No, explain in words.

No.
A model without the B term is the same as the model above.

~ 2 **c)** Write down the appropriate model if all the information in the figure is provided to you, but our main focus is the effect of A on the response. Explain your reasoning for the proposed model.

$y_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ij}$ more indices if replication

more indices if repl.

There is clearly an interaction between A,B because the effect of A depends on the level of B.

~ 2 **11.** Consider a CRD and the model $y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \epsilon_{ijk}$, where $i = 1 \cdots a,\ j = 1 \cdots b, k = 1 \cdots c$. Show that the sum of the predictions is equal to the sum of the observed values of y_{ijk}.

$$\hat{y}_{ijk} = \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}_k = \bar{y}_{...} + (\bar{y}_{i..} - \bar{y}_{...}) + (\bar{y}_{.j.} - \bar{y}_{...}) + (\bar{y}_{..k} - \bar{y}_{...})$$

$$\frac{1}{bc} y_{i..} \qquad = \bar{y}_{i..} + \bar{y}_{.j.} + \bar{y}_{..k} - 2\bar{y}_{...}$$

$$\sum_{ijk}^{abc} \hat{y}_{ijk} = \sum_{ijk} \bar{y}_{i..} + \sum_{ijk} \bar{y}_{.j.} + \sum_{ijk} \bar{y}_{..k} - 2\bar{y}_{...} \sum_{ijk} 1$$

$$\to \frac{1}{\cancel{ab}} \left(\cancel{\sum_{ij} 1}\right) \sum_k y_{..k}$$

$$= \frac{1}{bc} \sum_{ijk} y_{i..} + \frac{1}{ac} \sum_{ijk} y_{.j.} + \frac{1}{ab} \sum_{ijk} y_{..k} - 2 \frac{1}{abc} y_{...} \sum_{ijk}$$

$$= y_{...} + y_{...} + y_{...} - 2y_{...} = \boxed{y_{...}}.$$

~ 2 hw and Sample test **12.** Show that $y_{1..} + y_{2..} + y_{3..} = y_{...}$ where all "dots" refer to restricted sums over the LSD given by

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

$$\begin{pmatrix} 111 & 122 & 133 \\ 221 & 232 & 213 \\ 331 & 312 & 323 \end{pmatrix}$$

$$Y_{1..} = Y_{111} + Y_{122} + Y_{133}$$

$$Y_{2..} = Y_{221} + Y_{232} + Y_{213}$$

$$Y_{3..} = Y_{331} + Y_{312} + Y_{323}$$

$\therefore Y_{1..} + Y_{2..} + Y_{3..}$ = sum of all 9 entries

$= Y_{...}$

Lect 14

$\sim \cancel{3}$ 2.5 **13.** Consider the model $y_{ij} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \epsilon_{ij}$, where $i = 1 \cdots a$, $j = 1 \cdots b$, and $\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$.
a) Starting from the expression for the Likelihood of data, compute the maximum-likelihood (ML) estimate of the interaction term. You may use the fact that the ML estimates of μ, α_i, β_j are, respectively, $\overline{y_{..}}$, $(\overline{y_{i.}} - \overline{y_{..}})$ and $(\overline{y_{.j}} - \overline{y_{..}})$.

$$L = \prod_i \prod_j \frac{1}{\sqrt{2\pi\sigma_\epsilon^2}} e^{-\frac{1}{2}\left(\frac{Y_{ij} - \mu - \alpha_i - \beta_j - (\alpha\beta)_{ij}}{\sigma_\epsilon}\right)^2} = e^{-\frac{1}{2}(}$$

$$= e^{-\frac{1}{2\sigma_\epsilon^2}\sum_{ij}(Y_{ij} - \mu - \alpha_i - \beta_j - (\alpha\beta)_{ij})^2 - \overbrace{\sum_{ij}}^{ab} \frac{1}{2}\ln(2\pi\sigma_\epsilon^2)}$$

$$\frac{\partial}{\partial(\alpha\beta)_{kl}} : \frac{\partial}{\partial(\alpha\beta)_{kl}} \sum_{ij} (Y_{ij} - \mu - \alpha_i - \beta_j - (\alpha\beta)_{ij})^2 = -2(Y_{ij} - \mu - \alpha_i - \beta_j - (\alpha\beta)_{ij})$$

$$\therefore \widehat{(\alpha\beta)}_{ij} = Y_{ij} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_j = Y_{ij} - \overline{Y}_{..} - (\overline{Y}_{i.} - \overline{Y}_{..}) - (\overline{Y}_{.j} - \overline{Y}_{..})$$

$$= Y_{ij} - \overline{Y}_{i.} - \overline{Y}_{.j} + \overline{Y}_{..}$$

~ 2 **b)** Find the expression for the predictions $\hat{y_{ij}}$.

$$\hat{Y}_{ij} = \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \widehat{(\alpha\beta)}_{ij}$$

$$= \cancel{\hat{\mu}} + \cancel{\hat{\alpha}_i} + \cancel{\hat{\beta}_j} + Y_{ij} - \cancel{\hat{\mu}} - \cancel{\hat{\alpha}_i} - \cancel{\hat{\beta}_j}$$

$$= Y_{ij}$$

~ ~~3~~ 3.5 Lect 12 hw AG

14. Consider a CRD with a single qualitative factor A, and the model $y_{ij} = \mu + \alpha_i + \epsilon_{ij}$, $i = 1 \cdots a$, $j = 1 \cdots n$. We want to test whether A has an effect on the response. We know that $MS_{Tr} = \sum_{i,j}(\overline{y_{i.}} - \overline{y_{..}})^2$. Compute $SSE_{reduced}$ and SSE_{full}, and then show that $\cancel{M}S_{Tr} = SSE_{reduced} - SSE_{full}$. You may use the anova decomposition without deriving/proving it.

Reduced: $y_{ij} = \mu + \epsilon_{ij} \Rightarrow \hat{\mu} = \bar{y}_{..}$

$\therefore \hat{y}_{ij} = \hat{\mu} = \bar{y}_{..} \Rightarrow SSE_{reduced} = \sum_{ij} (y_{ij} - \bar{y}_{..})^2$

Full: $y_{ij} = \mu + \alpha_i + \epsilon_{ij} \Rightarrow \hat{\mu} = \bar{y}_{..}, \ \hat{\alpha}_i = \bar{y}_{i.} - \bar{y}_{..}$

$\therefore \hat{y}_{ij} = \hat{\mu} + \hat{\alpha}_i = \bar{y}_{i.} \Rightarrow SSE_{full} = \sum_{ij} (y_{ij} - \bar{y}_{i.})^2$

$$\underline{SSE_{red} - SSE_{full}} = \underbrace{\sum_{ij} (y_{ij} - \bar{y}_{..})^2}_{SST} - \underbrace{\sum_{ij} (y_{ij} - \bar{y}_{i.})^2}_{SSE} \underset{\text{ANOVA decomp.}}{=} \sum_{ij} (\bar{y}_{i.} - \bar{y}_{..})^2 = \underline{SS_{Tr}}$$

15. Consider the following data on a response y (numbers in the table) from two treatement levels of a factor A.

A	Runs		
1	1	2	3
2	2	3	4

~ 2 a) Suppose the data have been collected in a CRD. Use an appropriate model and report the numerical value of SSE. Hint: find $\overline{y_{i.}}$ first.

$y_{ij} = \mu + \alpha_i + \epsilon_{ij}$ $\quad a=2, n=3 \Rightarrow \hat{y}_{ij} = \bar{y}_{i.} = \begin{cases} \frac{1}{3}(1+2+3) = 2 & i=1 \\ \frac{1}{3}(2+3+4) = 3 & i=2 \end{cases}$

$$SSE = \sum_{ij} (y_{ij} - \bar{y}_{i.})^2 = (1-2)^2 + (2-2)^2 + (3-2)^2 \leftarrow i=1$$
$$+ (2-3)^2 + (3-3) + (4-3)^2 \leftarrow i=2$$
$$= 1+0+1 + 1+0+1 = \boxed{4}$$

~ 2 b) Now, suppose the 3 Runs occurred on 3 different days, and there was no randomization between days. Use an appropriate RCBD model and report the numerical value of SSE. Hint: find $\overline{y_{.j}}, \overline{y_{..}}$.

$y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij} \Rightarrow \hat{y}_{ij} = \bar{y}_{i.} + \bar{y}_{.j} - \bar{y}_{..}$, $\quad \bar{y}_{.j} = \begin{cases} \frac{1}{2}(1+2) = 3/2 \\ \frac{1}{2}(2+3) = 5/2 \\ \frac{1}{2}(3+4) = 7/2 \end{cases}$

$SSE = \sum_{ij} (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2$ $\quad \bar{y}_{..} = \frac{15}{6} = 5/2$

$$= \left(1 - 2 - \frac{3}{2} + \frac{5}{2}\right)^2 + \left(2 - 2 - \frac{5}{2} + \frac{5}{2}\right)^2 + \left(3 - 2 - \frac{7}{2} + \frac{5}{2}\right)^2 \leftarrow i=1$$
$$\left(2 - 3 - \frac{3}{2} + \frac{5}{2}\right)^2 + \left(3 - 3 - \frac{5}{2} + \frac{5}{2}\right)^2 + \left(4 - 3 - \frac{7}{2} + \frac{5}{2}\right)^2 \leftarrow i=2$$
$$= \boxed{0}$$