Exam 2 Review

Monday March 27

Announcement: Office Hours this week

Monday	until 1:30pm	Dr. Burr
Tuesday	9am – 11am	Yanyi Liu (zoom)
Wednesday	NO OFFICE HOUR	

Note, correction to Review answers –
Q#5, the extra notes in class
had an error ~~yester~~ (a, b instead of r, ni),
which is now fixed.

STA 4211 Midterm 2 Review

Date: Exam 2 will be in class on Wednesday March 29. Bring your pencil, calculator (graphing is ok), crib sheet (one sheet, back and front, of your own creation), UF id.

General description: The test will be a reasonable length for a 50-minute class period. There will be some written questions, and some True/False and multiple-choice questions.

Coverage: Lectures 11 – 25, Homeworks 4 and 5, Quizzes 2 and 3, parts of Project 1 (no R coding will be asked for on the exam). Study everything that was gone over in class, and the solution sheets to the homeworks and Project 1.

Review Questions Answers will be given in class on Monday March 27. **Caution**: Not all possible topics are represented here, and not all topics covered by these questions will be on Exam 2.

1. Suppose George used a regression model for a pricing study involving three different price levels $X = \$50$, $\$60$, and $\$70$. The purpose was to study the effect of price on sales volume. George asks you: "Do you think it would be a good idea to use a one-way ANOVA model instead?" What would you say? [Note: Your answer should include some specifics. It is not sufficient to answer "It depends."]

Solution

I would say, "George, if the relationship between $X =$ Price and $Y =$ Sales Volume is not linear, it would be a good idea to use a one-way ANOVA model for your situation, in order to allow arbitrarily different mean sales volumes for the three prices."

2. True or False? To find the least-squares estimator in one-way ANOVA, you must know the value of the error variance.

Answer False

3. True or False? The method of least-squares in one-way ANOVA depends upon the assumption of normality.

Answer False

4. State what is the least-squares method for the one-way ANOVA setting.

Solution Define the criterion Q for the one-way ANOVA setting to be

$$Q = \sum_{i=1}^{r} \sum_{j=1}^{n_i} (Y_{ij} - \mu_i)^2$$

The least-squares method is to minimize the above criterion with respect to the μ_i, $i = 1, \ldots, r$.

Note You should know that least-squares estimates ~~st~~ satisfy an optimality criterion ("best linear unbiased estimate")
BLUE

5. Recall that in one-way ANOVA, we use the notation r = the number of groups, n_i the number of observations in Group i, and Y_{ij} the j^{th} observation in Group i. Which of the following formulas is correct for the grand mean? Select the single correct answer.

(A) $\bar{Y}_{..} = \sum_{i=1}^{r} \sum_{j=1}^{n_i} Y_{ij}$

(B) $\bar{Y}_{..} = \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij}$

(C) $\bar{Y}_{..} = \frac{1}{\sum_{i=1}^{r} n_i} \sum_{i=1}^{r} \sum_{j=1}^{n_i} Y_{ij}$

(D) None of the above formulas is correct.

Solution

Answer is C.

$$\bar{Y}_{..} = \frac{1}{n_T} \sum_{i=1}^{r} \sum_{j=1}^{n_i} Y_{ij} = \frac{1}{\sum_{i=1}^{r} n_i} \sum_{i=1}^{r} \sum_{j=1}^{n_i} Y_{ij}$$

Consider options A and B:

$$(A) \quad \sum_{i=1}^{r} \sum_{j=1}^{n_i} Y_{ij} = Y_{..}$$

$$(B) \quad \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij} = \bar{Y}_{i.}$$

6. In a one-way layout with three treatments, suppose the group sample sizes are $n_1 = 50$, $n_2 = 55$, and $n_3 = 60$, and that the value of $\hat{\sigma}^2$ from the ANOVA is MSE $= 2.0$. Consider the linear combination $L = (\mu_2 + \mu_3)/2$, for which the estimate is $\hat{L} = (\bar{Y}_{2.} + \bar{Y}_{3.})/2$. (A) Is L a contrast? (B) Find the estimate of the standard error of $\hat{L}$.

Solution

(A) No, L is not a contrast, because $\sum_{i=1}^{r} c_i = 0 + \frac{1}{2} + \frac{1}{2} = 1$ (so the sum of the coefficients of the linear combination is not zero).

(B)

$$\text{SE}\left(\frac{1}{2}\bar{Y}_{2.} + \frac{1}{2}\bar{Y}_{3.}\right) = \sqrt{\text{MSE}}\sqrt{(1/2)^2/55 + (1/2)^2/60} = \sqrt{2.0}\sqrt{.008712} = 0.132$$

$$\text{Var}\left(\sum_{i=1}^{r} c_i \bar{Y}_{i.}\right) = \sigma^2 \sum_{i=1}^{r} \frac{c_i^2}{n_i}$$

7. For a three-group ANOVA with $n_1 = 35, n_2 = 35, n_3 = 34$, suppose we know that the true means are $\mu_1 = 0, \mu_2 = -6, \mu_3 = -12$, and that the error variance is $\sigma^2 = 120$. (A) What is the expected value of the mean square for error? (B) What is the expected value of the mean square for treatment?

Solution

(A) $E(MSE) = \sigma^2 = 120$

(B) Find the weighted mean of the cell means,

$$\mu_{.} = \frac{\sum_{i=1}^{r} n_i \mu_i}{n_T} = \frac{35(0) + 35(-6) + 34(-12)}{104} = -5.94$$

Then,

$$\begin{aligned} E(MSTR) = \sigma^2 + \frac{\sum n_i(\mu_i - \mu_{.})^2}{r-1} &= 120 + \frac{35(0-(-5.94))^2 + 35(-6+5.94)^2 + 34(-12+5.94)^2}{2} \\ &= 120 + 1242 \\ &= 1362 \end{aligned}$$

Recall also: We used these to explain behavior of the F statistic under <u>Ho</u>, and under <u>Ha</u>.

8. For a four-group ANOVA with $n_1 = 5, n_2 = n_3 = n_4 = 6$, consider the observation denoted Y_{41}.

(A) What position (index) does Y_{41} have in the vector of responses, with the observations listed in the usual order (Group 1 followed by Group 2, etc.)? ____________________

(B) Write out the model equation for Y_{41} using the factor-effects formulation (use either the non-full-rank or the full-rank version).

Solution

(A) Since $n_1 + n_2 + n_3 = 17$, Y_{41} is the 18^{th} obs.

(B) $Y_{41} = \mu_{\cdot} + \tau_4 + \varepsilon_{41}$

or $Y_{41} = \mu_{\cdot} - \tau_1 - \tau_2 - \tau_3 + \varepsilon_{41}$

9. Consider a four-group ANOVA situation with $n_1 = 5, n_2 = n_3 = n_4 = 6$. Recall the matrix model equation for any general linear model is given by:

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

For the factor-effects formulation of this model, write out the parameter $\boldsymbol{\beta}$. Describe the design matrix $\mathbf{X}$: State its row and column dimensions, and give one row of the matrix for each of the four groups. Note: Use the full-rank formulation of the factor-effects model.

Solution

$$\beta_{FR} = \begin{pmatrix} \mu_{\cdot} \\ \tau_1 \\ \tau_2 \\ \tau_3 \end{pmatrix}$$

$\underset{\sim}{X}$: n_T rows, $n_T = 23$

$r = 4$ columns

Group	Row
1	1 1 0 0
2	1 0 1 0
3	1 0 0 1
4	1 −1 −1 −1

10. In a scientific article, a one-way analysis of variance is reported, for which all the group sizes are equal. The following table is reported:

Analysis of Variance

Source	SS	df	MS	F
Treatment	722.7	4	180.68	15.3
Error	473.3	40	11.83	
Total				

(A) In the Total row, fill in the blanks for SS and df.

(B) How many groups were there, and what was the sample size in each group?

Solution

(A) 1195.0, 44 (by the decomposition of sums of squares)

(B) df(Treatment) $= r - 1 = 4$, so $r = 5$ is the number of groups. df(Total) $= n_T - 1 = 44$, so $n_T = 45$, and so there were 9 observations in each of the five groups.

11. In a balanced completely randomized design there were $r = 3$ groups with means $\overline{Y}_{1.} = 8.67$, $\overline{Y}_{2.} = 8$, $\overline{Y}_{3.} = 3.33$. The value of Tukey's HSD ("honest significant difference") at confidence level .95, $\frac{1}{\sqrt{2}}q(.95, r, n_T - r)\sqrt{\text{MSE}}\sqrt{2/n}$, is 3.44.

(A) Use Tukey's studentized range method to form confidence intervals for all pairwise comparisons, with familywise coverage level .95.

(B) Interpret this family of confidence intervals with a statement beginning: "We are 95% confident that ..."

Solution

In the table below, the Tukey intervals are given by: Estimate ± 3.44.

Comparison	Estimate	Interval
D1$= \mu_1 - \mu_2$	$8.67 - 8.0 = .67$	$(-2.77, 4.11)$
D2$= \mu_1 - \mu_3$	$8.67 - 3.33 = 5.34$	$(1.90, 8.78)$
D3$= \mu_2 - \mu_3$	$8.0 - 3.33 = 4.67$	$(1.23, 8.11)$

Interpretation: We are 95% confident that all three confidence intervals are correct, that is, contain the parameter being estimated.

Also (not required for this particular question); This corresponds to three simultaneous hypothesis tests at level $\alpha = 0.05$: By Tukey's method we fail to reject $H_0 : \mu_1 = \mu_2$ since that confidence interval contains the value 0, whereas we conclude that μ_1 and μ_3 are different and μ_2 and μ_3 are different.

12. Consider a four-group one-way ANOVA setup. Suppose a researcher plans to estimate all pairwise comparisons, plus $\mu_. = (\mu_1 + \mu_2 + \mu_3 + \mu_4)/4$.

- For the Bonferroni method, what would be the coverage level of each of the individual confidence intervals?
- Would the Bonferroni method be appropriate here? What do you have to know to determine whether the Bonferroni method is appropriate?
- Would the Scheffé method be appropriate here? Why or why not? Explain briefly.
- Would Tukey's studentized range be appropriate here?

Solution (A) The number of comparisons is $k = \binom{4}{2} + 1 = 6 + 1 = 7$, so the coverage level for each individual CI would be $100(1 - .05/7)\%$ for simultaneous coverage probability 95%.

Extra note: A different question that could be asked is what is the quantile of the t distribution for the individual CI, and this would be $1 - \frac{.05}{2(7)}$.

(B) The Bonferroni method would be appropriate if these seven comparisons were decided upon in advance of collecting the data, but would not be appropriate if the investigator decided on these parameters after seeing the results.

(C) The Scheffé method would not be appropriate because the overall mean is not a contrast.

(D) Tukey's studentized range can't be used with all seven parameters since the overall mean is not a pairwise difference.

13. In a two-factor study, the treatment means μ_{ij} are as follows:

	Factor B		
Factor A	B_1	B_2	B_3
A_1	6	12	9
A_2	8	18	13

$\mu_{1\cdot} = 9$

(Know notation as well)

(A) Obtain the factor B level means.

(B) Obtain the main effects of factor B.

(C) What is the value of the interaction effect $(\alpha\beta)_{12}$ for the above table of treatment means?

(D) Do Factors A and B interact? Explain your answer.

Solution

$\mu_{\cdot j}$: 7, 15, 11 $\qquad \mu_{\cdot\cdot} = 11$

β_j : -4 $\quad 4$ $\quad 0$

$$(\alpha\beta)_{12} = \mu_{12} \cancel{\mu_{ij}} - (\mu_{1\cdot} + \mu_{\cdot 2} - \mu_{\cdot\cdot})$$

$$= 12 - (9 + 15 - 11)$$

$$= \cancel{d}\ 12 - (13)$$

$$= -1$$

Yes, A & B interact since $(\alpha\beta)_{12} \neq 0$

14. For the two-way ANOVA model, with cell means μ_{ij}, $i = 1, \ldots, a;\ j = 1, \ldots b$, show that the overall mean is the average of the row averages (column averages).

Solution

Now, $\mu_{\cdot\cdot} \overset{def}{=} \frac{1}{ab} \sum_{i=1}^{a} \sum_{j=1}^{b} \mu_{ij}$

Show $\mu_{\cdot\cdot} = \frac{1}{a} \sum_{i=1}^{a} \mu_{i\cdot}$

~~Proof~~ Now $\mu_{i\cdot} = \frac{1}{b} \sum_{j=1}^{b} \mu_{ij}$

So $\cancel{\mu_{\cdot\cdot} = 1}$

$$\frac{1}{a} \sum_{i=1}^{a} \mu_{i\cdot} = \frac{1}{a} \sum_{i=1}^{a} \frac{1}{b} \sum_{j=1}^{b} \mu_{ij}$$

$$= \frac{1}{ab} \sum_{i} \sum_{j} \mu_{ij} = \mu_{\cdot\cdot}$$

DONE