STATISTICS 642 - FINAL EXAM

May 5, 2022

Student's Name _____

Student's Email Address _____

INSTRUCTIONS FOR STUDENTS:

(1) The exam consists of 9 pages of questions including the cover page.

(2) You may start the exam at 6pm (Texas Time) on May 5, 2022.

(3) You must complete your exam before 8pm (Texas Time) and submit your solutions prior to 8:20pm (Texas Time) on May 5, 2022.

(3) Show *ALL* your work on the exam pages. Select the BEST answer for multiple choice unless multiple answers has to be entered to be able to answer thorougly.

(4) You have two options for composing your solutions. For both methods, make sure to write using a black pen or write using a pencil in such a manner that I can clearly read your solutions.

   i. Print out the exam, put your solutions on the exam, then scan into a single pdf and upload to Canvas.

   ii. Write/Type your solutions on paper, scan the solutions into a single pdf, then upload to Canvas.

(5) Do not discuss or provide information to anyone concerning the questions on this exam or your solutions until I post the solutions to the exam.

(6) You may use the following:

- Calculator - Your device cannot facilitate a connection to the internet or to send text messages
- Summary Sheets - **10-pages,** 8.5"**x**11" - (you may write/type/cut-paste anything on both sides of the ten sheets)
- The attached Tables or downloaded **Tables for 642 Exams** from CANVAS
- R or Excel as handcalculator

(7) Do not use any other written material except for your summary sheets and tables.

(8) Do not use a computer for google search, cell phone, or any other electronic device (other than a calculator) except to download the exam and upload your solutions.

I attest that I spent no more than 2 Hours to complete the exam. I used only the materials described above. I did not receive assistance from anyone during the taking of this exam.

Student's Signature_____

Problem I. (30 points) For the following experiment, provide the requested information:

A study was designed to evaluate the impact of chemical plant discharges into Michigan lakes on the uptake of carcinogens by game fish. The four most widely caught species (S1, S2, S3, S4) of fish were selected for this study. The biologist decided to assess the fish's uptake of carcinogens by feeding mice portions of fish known to have been contaminated with the carcinogens. The biologist were also interested in assessing three different methods of determining the amount of carcinogen in the bladder of mice. To control for the variation in bladder density from mouse to mouse, the biologists randomly selected four mice from each of five litters of mice for use in the study. Within each litter, the four mice were randomly assigned to the four species of fish, one mouse to each species. The mice were fed a diet of fish for six weeks and then were sacrificed. The bladder of each mouse was divided into three sections and the sections randomly assigned to one of the three assessment techniques (A1, A2, A3). The amount of carcinogen in the bladder sections (ppb) are displayed in the following table:

		Fish Species			
Liters	Assessment	S1	S2	S3	S4
L1	A1	26.97	26.12	27.83	27.47
	A2	22.60	22.91	19.85	21.63
	A3	30.71	29.53	27.51	28.62
L2	A1	17.47	18.13	18.01	17.97
	A2	16.90	16.31	16.52	15.93
	A3	23.95	22.84	23.84	23.45
L3	A1	20.72	20.41	21.01	21.34
	A2	24.32	25.06	25.92	25.33
	A3	28.31	29.02	29.13	29.36
L4	A1	20.93	22.64	20.64	19.47
	A2	24.53	23.54	18.99	20.33
	A3	21.41	22.84	24.13	21.49
L5	A1	26.98	23.67	25.64	20.67
	A2	21.77	20.67	23.57	22.22
	A3	30.98	30.82	29.48	19.62

1. Circle to select only one letter to indicate the correct Type of Randomization:
   - A. CRD with factorial assignment
   - B. CRD with split plot treatment assignment
   - C. RCBD with factorial assignment
   - D. RCBD with split plot treatment assignment
   - E. RCBD with split split plot treatment assignment
   - E. BIBD with factorial assignment
   - F. BIBD with split plot treatment assignment
   - G. LSD with split plot treatment assignment
2. Circle to select only one letter to indicate the correct Type of Treatment Structure:
   - A. two factors crossed
   - B. three factors crossed
   - C. four factors crossed
   - D. block design, two factors crossed
   - E. block design, three factors crossed
   - F. block design, four factors crossed
3. Circle to select and identify the Fixed Factor(s). All fixed factor(s) should be correctly selected :
   - A. Litter
   - B. Assessment techniques
   - C. Species of Fish
   - D. Fish
   - E. Mouse
   - F. Bladder of each mouse
   - G. The amount of carcinogen in the bladder sections

4. Circle to select and identify the Random Factor(s). All random factor(s) should be correctly selected:

   A. Litter
   B. Assessment techniques
   C. Species of Fish
   D. Fish
   E. Mouse
   F. Bladder of each mouse
   G. The amount of carcinogen in the bladder sections

5. Select one letter to identify the Experimental Units and Measurement Units:

   A. EU=MU=mouse
   B. EU in the whole plot is mouse and splitplot has EU=MU=bladder of each mouse
   C. EU=MU=bladder of each mouse
   D. EU in the whole plot is fish and splitplot has EU=MU=mouse
   E. EU=MU=fish
   F. EU in the whole plot is fish, splitplot has EU=mouse, and split split plot has EU=MU=bladder of each mouse

6. Which of the following best describe the Measurement Process? Select to identify and also state for example the name of Response Variable, Covariates, SubSampling, Repeated Measures variable whichever applies:

   A. Response variable _____

   B. Covariate(s) _____

   C. Subsampling _____

   D. Repeated measure _____

7. Select one letter to best describe the assumption made for the fixed factor with a levels:

   A. $\tau_i = 0$ for i=1,2,...,a
   B. $b_j \sim NIID(0, \sigma_b^2)$ for j=1,2,...,a
   C. $b_j \sim NIID(\mu, \sigma_b^2)$ for j=1,2,...,a
   D. $\tau_1 = \tau_2 = ... = \tau_a$

8. Select one letter to best describe the assumption made for the random factor with a levels:

   A. $\tau_i = 0$ for i=1,2,...,a
   B. $b_j \sim NIID(0, \sigma_b^2)$ for j=1,2,...,a
   C. $b_j \sim NIID(\mu, \sigma_b^2)$ for j=1,2,...,a
   D. $\tau_1 = \tau_2 = ... = \tau_a$

9. In a hypothetical study, assume you have two factors, one is random and the other one fixed, select one letter to indicate if the interaction of these two factors become random or fixed.

   A. Fixed
   B. Random

10. In a hypothetical study, you have two random factors and one fixed factor. Select one letter to indicate the number of variances in the full factorial model including the error variance.

    A. 2
    B. 3
    C. 4
    D. 7
    E. 8
    F. 10

Problem II (31 points.) An industrial engineer designs a study to evaluate two factors which may reduce the time needed to insert electronic components on circuit boards. She has decided on three types of assembly fixtures (F1, F2, F3) and two types of workplace layouts (L1, L2) for investigation. Human operators perform the assembly. Because the two workplace layouts are in two different locations, the engineer decides to randomly select four operators at each of the two workplace layouts for a total of 8 operators in the study. Each operator assembles six circuit boards, two boards for each of the three Fixture types. The assembly times (minutes) are given below.

	Layout 1				Layout 2				
	Operator				Operator				
	1	2	3	4	5	6	7	8	Fixture Means
Fixture 1	22	23	28	25	26	27	28	24	25.25
	24	24	29	23	28	25	25	23	
Fixture 2	30	29	30	27	29	30	24	28	27.94
	27	28	32	25	28	27	23	30	
Fixture 3	25	24	27	26	27	26	24	28	25.06
	21	22	25	23	25	24	27	27	
Layout Means	25.79				26.38				

1. In the AOV table place the values of the degrees of freedom, **DF** and the expected mean squares, **EMS** for each of the sources of variance in the following AOV table. The **Mean Squares (MS)** are provided in the AOV table. The notation F = Fixture, L = Assembly Layout, O = Operator is used in the AOV table.

SOURCE	DF	**MS**	EMS
F		41.40	
L		4.08	
O(L)		11.99	
F*L		9.52	
F*O(L)		5.49	
ERROR		2.33	
TOTAL			

2. At the $\alpha = 0.05$ level, identify the significant effects of Fixture and Assembly Layout on the time to complete a circuit board. Justify your answer.

   For Interaction,

   $H0$:

   Test statistics: $F =$

   Decision with justification:

   For Fixture,

   $H0$:

   Test statistics: $F =$

   Decision with justification:

   For Assembly Layout,

   $H0$:

   Test statistics: $F =$

   Decision with justification:

3. The following model was fit to the data where $y_{ijk\ell}$ is the time needed to assemble

   circuit board ℓ by Operator k using Fixture i in Layout j :

$$y_{ijk\ell} = \mu \ + \ \tau_i \ + \ \gamma_j \ + \ b_{k(j)} \ + \ (\tau\gamma)_{ij} \ + \ d_{ik(j)} \ + \ e_{ijk\ell}$$

   Using the numeric values of the MS's and EMS's given in the AOV table, compute a 95% confidence interval for the mean time to assemble a circuit board using Layout 2. Justify your answer clearly writing the estimated mean, SE of the mean and also critical value.

Problem III. (12 points) A simulation model of an inventory system involved 7 factors: A, B, C, D, E, F, G, each having two levels. The evaluation of the model used a single replication of a 2^{7-3} fractional factorial design. The generators $ABDG = -1$, $ACDF = +1$, and $BCDE = -1$ were used to select the treatments to be included in the study.

1. How many treatments would be observed in this experiment?

2. Would the following treatment $(A, B, C, D, E, F, G) = (H, L, H, H, L, H, H)$ be used in the experiment? Justify your answer.

3. What is the resolution of this design? Justify your answer.

4. Which of the interactions would need to be negligible in order to be able to obtain an estimate of the main effect of Factor A?

Problem IV. (27 points) **CIRCLE ONE** of (A, B, C, D, E) corresponding to the **BEST**

(1.) After conducting a CRD with the factors A and B having 3 fixed levels each, the researcher conducts a residual analysis and determines there are a large number of outliers. Because a transformation of the data often leads to a situation where the conclusions are hard to interpret, the research would like you to suggest an alternative approach. Which one of the following methods would be a valid method to determine the existence of an **interaction** between the factors A and B?

A. Use the results from the ANOVA F-test because the F-test is robust to deviations from the normality condition.

B. Apply the multiple comparison procedure associated with the Kruskal-Wallis rank based procedure to the pairs of levels of factor A at each level of factor B.

C. Use the results from the transformed data because you can just invert the transformation to obtain results in the original scale.

D. Use the Friedman rank based procedure with the blocking factor designated as factor B.

E. None of the above would be valid.

(2.) A three factor experiment is run with Factor A-fixed, Factor B-fixed nested within Factor A, Factor C-fixed crossed with Factor A. The following effects were significant: nested effect of factor B in factor A, and the interaction between A and C. The following effects were not significant: main effect of A, main effect of C and the interaction between B(A) and C.

A. A comparison of the differences in the levels of Factor B can be conducted using Tukey's HSD on the means for the levels of Factor B averaged over the levels of Factors A and C.

B. A comparison of the differences in the levels of Factor B can be conducted using Tukey's HSD on the means for the levels of Factor B computed separately at each level of factors A and C.

C. A comparison of the differences in the levels of Factor B can be conducted using Tukey's HSD on the means for the levels of Factor B computed separately at each level of factor A.

D. A comparison of the differences in the levels of Factor B can be conducted using Tukey's HSD on the means for the levels of Factor B computed separately at each level of factor C.

E. None of the above

(3.) In a CRD with a quantitative factor A at 4 levels and a qualitative factor B at 2 levels, the researcher wants to know if there is a linear trend in the mean responses across the levels of A. The AOV table reveals a non-significant $A * B$ interaction. Which of the following contrasts would address this question?

A. $L = -3\mu_{11} - \mu_{21} + \mu_{31} + 3\mu_{41} - 3\mu_{12} - \mu_{22} + \mu_{32} + 3\mu_{42}$

B. $L = 3\mu_{11} + \mu_{21} - \mu_{31} - 3\mu_{41} - 3\mu_{12} - \mu_{22} + \mu_{32} + 3\mu_{42}$

C. $L = \mu_{11} + \mu_{21} + \mu_{31} + \mu_{41} - \mu_{12} - \mu_{22} - \mu_{32} - \mu_{42}$

D. $L = \mu_{11} - \mu_{21} - \mu_{31} + \mu_{41} - \mu_{12} + \mu_{22} + \mu_{32} - \mu_{42}$

E. $L = -3\mu_{11} - \mu_{21} + \mu_{31} + 3\mu_{41}$

(4.) In a RCBD repeated measures design, the researcher ignored the repeated measures aspect of the design and analyzed the data as a RCBD with a factorial treatment structure. A residual analysis indicated a strong positive correlation in the residuals. Your assistant tells the project director that this is great because the actual power of an $\alpha = .05$ AOV F-test will now be greater than the power of an $\alpha = .05$ AOV F-test when the residuals are independent. Although your assistant is correct, the negative impact of positive correlation on the testing procedure is

A. the probability of a Type II error will be higher than expected under no correlation.

B. the probability of a Type II error will be less than expected under no correlation.

C. the probability of a Type I error will be less than .05.

D. the probability of a Type I error will be higher than .05.

E. In fact your assistant's statement is incorrect, the power will in fact be decreased because the assumptions of the AOV F-test have been violated.

(5.) In a CRD 3x4 factorial design, the SAS or R output for the LSMEANS of the the Two Factors displayed the expression NON-ESTIMABLE. This means that the main effects of the two factors

A. have a significant interaction and hence we should not estimate main effects.

B. cannot be unbiasedly estimated using a linear combination of the data values.

C. have REML estimates but cannot be obtained in a
nite number of iterations.

D. are confounded and hence cannot be uniquely estimated.

E. there must have been errors in your SAS or R code.

(6.) A researcher is studying two treatments: factor F_1 with three levels and factor F_2 with three levels. The initial design had five replications of each of the nine treatments, but because of problems that occurred during the experiment, only the following number of observations were observed for the nine treatments:
$n_{11} = 0; n_{12} = 5; n_{13} = 5; n_{21} = 4; n_{22} = 0; n_{23} = 5; n_{31} = 0; n_{32} = 3; n_{33} = 3$
which of the following would be an appropriate contrast to evaluate the main effect of F_2?

A. $\frac{1}{3}(\mu_{12} + \mu_{22} + \mu_{32}) - \frac{1}{3}(\mu_{13} + \mu_{23} + \mu_{33})$

B. $\frac{1}{2}(\mu_{12} + \mu_{32}) - \frac{1}{3}(\mu_{13} + \mu_{23} + \mu_{33})$

C. $\frac{1}{2}(\mu_{12} + \mu_{32}) - \frac{1}{2}(\mu_{13} + \mu_{33})$

D. $\mu_{12} - \mu_{33}$

E. $\mu_{32} - \mu_{23}$

(7.) A covariate was measured along with the responses within a completely randomized design. The researcher determines that the slopes of the 5 treatment lines are different. A comparison of the 5 treatments

A. could be made using adjusted treatment means at specified values of the covariate.

B. cannot be conducted because there is an interaction between the covariate and treatment which violates the conditions for analysis of covariance.

C. could be made using Tukey's HSD on the sample treatment means.

D. could be made using Tukey's HSD on the adjusted treatment means.

E. none of the above are valid

(8.) A RCBD with repeated measurements on each EU was conducted producing data which was modeled as follows: there were b blocks, t fixed treatment levels observed on r EU's for each treatment. Each EU was measured at p time points. Let $y_{ijk\ell}$ be the measurement of ℓth EU receiving treatment j in block i during time period k.

$$y_{ijk\ell} = \mu + b_i + \tau_j + d_{ij} + \gamma_k + (\tau\gamma)_{jk} + e_{ijk\ell}, \quad \text{with} \;\; i = 1, \ldots, b; \;\; j = 1, \ldots, t; \;\; k = 1, \ldots, p; \;\; \ell = 1, \ldots, r;$$

where μ, τ_j γ_k, and $(\tau\gamma)_{jk}$ are fixed population parameters and b_i, d_{ij}, and $e_{ijk\ell}$ are random variables with $N(0, \sigma_b^2)$, $N(0, \sigma_d^2)$, and $N(0, \sigma_e^2)$ distributions, respectively. Which one of the follow statements most appropriately describes the correlation structure of the random variables in the model?

A. b_i's are independent, d_{ij}'s are independent, and $e_{ijk\ell}$'s are independent

B. b_i's are correlated, d_{ij}'s are independent, and $e_{ijk\ell}$'s are independent

C. b_i's are independent, d_{ij}'s are correlated, and $e_{ijk\ell}$'s are independent

D. b_i's are independent, d_{ij}'s are independent, and $e_{ijk\ell}$'s are correlated

E. b_i's are correlated, d_{ij}'s are correlated, and $e_{ijk\ell}$'s are correlated

(9.) A plant scientist is studying $t = 4$ chemicals for retarding the growth of rose bushes in commercial nurseries. The project statistician calculates that in order to meet power specifications $r = 10$ rose bushes are needed for each of the four retardants. There is considerable variation in the growth rate of rose bushes depending on the Area in which the bush is planted as a seedling. Thus, the study design is a RCBD with 4 bushes from each of 10 Areas, the blocking variable. From the experiment, the following results were obtained: $MS_{Area} = 36$ and $MS_{Error} = 12$. If a second experiment was conducted as a CRD by randomly selecting rose bushes ignoring the Area but using the same 4 growth retardants, what is the minimum number of replications (rose bushes) per growth retardant that would be needed to achieve the same level of accuracy in estimating the treatment means in the CRD as was obtained in the original design, the RCBD?

A. 5

B. 10

C. 15

D. 20

E. It would be impossible to determine with the given information.