STATISTICS 642 - FINAL EXAMINATION

May 5, 2014

Student's Name ____

Student's Email Address ______________________________________

INSTRUCTIONS FOR STUDENTS:

(1) The Exam consists of 10 pages including this cover page.

(2) You have exactly 2 hours and 15 minutes to complete the exam.

(3) Do not discuss or provide any information to anyone concerning any of the questions on this exam or your solutions until I post the solutions on Friday May 9.

(4) You may use the following:

- Tables - You may use **ONLY the Tables** from the textbook or **STAT642E2_Tables** from Dostat
- Calculator - Your device cannot facilitate a connection to the internet or emailing or texting.
- Summary Sheets - (**7-pages,** 8.5"**x11", you may write on both sides of the 7 pages**)

(6) Do not use your textbook other than tables, class notes, or any other written material except for the summary sheets.

(7) Do not use a computer, cell phone, or any other electronic device.

I attest that I spent no more than 2 hours and 15 minutes to complete the exam. I used only the materials described above. I did not receive assistance from anyone during the taking of this exam and I did not provide assistance to another student in the completion of this exam.

Student's Signature________________________________

INSTRUCTIONS FOR PROCTOR:

Immediately after the student completes the exam:

- Collect all portions of this exam. **DO NOT** allow the student to take any portion of the exam with her/him. Please keep these materials until Friday May 9.
- Please send all 10 pages of the exam to me in the following manner:
  - Scan the exam to a pdf file and then have the student upload the pdf file into WebAssign

(1) I certify that the time at which the student started the exam was ____________________ and the time at which the student completed the exam was

____________________.

(2) I certify that the student has followed all the **INSTRUCTIONS FOR STUDENTS** listed above:

Proctor's Signature________________________________

Problem I. (20 points) For the following experiment, provide the requested information:

An entomologist is studying the effect of three types of insecticides (T1, T2, T3) on the yield of soybean crops. Five varieties of soybean (V1, V2, V3, V4, V5) were randomly selected from the 97 most widely used soybean varieties. Four large fields were randomly selected for the study. Within each field, five similar 1-acre areas were identified and a single variety of soybeans was randomly assigned to each of these areas resulting in each field having all five varieties present. Each of the 20 areas was then divided into 6 regions with 2 regions randomly assigned to each of the three types of insecticide. The total yield of soybeans was obtained for each of the 120 regions. A measure of soil permeability and soil pH was measured for each region prior to planting the soybeans. The entomologist's major goal for the experiment was to determine the interrelationships between varieties of soybeans and types of insecticides on the yield of soybeans.

1. Type of Randomization: CRD, RCBD, LSD, Split-Plot, Crossover, etc.

2. Type of Treatment Structure: single factor, crossed, nested, fractional, etc.

3. Identify each of the Factors as being Fixed or Random:

4. Describe the Experimental Units and Measurement Units:

5. Describe the Measurement Process: Response Variable, Covariates, SubSampling, Repeated Measures

Problem II. (35 points) A food scientist designed a study to evaluate the effect of the age of a consumer on how they rate the bitterness of the four most popular brands of cheddar cheese. Two containers of each of the four Types of cheddar cheese (T) are obtained. Ten professional taste raters (R) are randomly selected from each of three age groups (A): 20-29, 30-49, and 50-70 years. Each of the 30 raters will taste eight portions of cheese presented in random order. The eight portions consist of one portion from each of the 8 containers with the raters unaware of the type of cheese within each portion. Each rater records a bitterness score after each bite of cheese: B_{ijkm}. There is sufficient waiting time between the 8 ratings so that the ratings are not biased by a previous rating of the taste. The food scientist is interested in estimating whether the difference in the average bitterness score for the four types of cheeses is consistent across the three age groups. She is also interested in determining the size of the variation in the ratings within the three age groups.

1. Complete the following AOV table by entering values for degrees of freedom and expected mean squares.

SOURCE	DF	**MS**	Expected MS .
T		3.79	
A		13.27	
T*A		2.78	
R(A)		2.58	
T*R(A)		1.06	
ERROR		0.91	

2. The following model was fit to the data where B_{ijkm} is the bitterness rating of a cheese portion m of cheese Type i from the kth Rater in Age group j :

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

Provide all conditions that must be placed on the terms in your model in order to be able to conduct normal based statistical inference procedures.

3. At the $\alpha = 0.05$ level, evaluate the effect of the Rater's Age on the average bitterness rating of cheese. Note that the numbers given in the AOV table are the Mean Squares (MS) NOT the Sum of Squares (SS).

4. Estimate the standard error of the estimated difference in the average ratings between age groups 20-29 and 50-70.

Problem III. (9 points) Six factors, A, B, C, D, E, F, are selected to be included in a study to determine which of these factors have an impact on the level of defective products in the company's production line. It was decided to use 2 levels (L or H) of each of these factors. The study design was a fractional factorial with 16 runs using the generators

$I_1 = BCD = +$ and $I_2 = ACE = +$ to select the 16 treatments to be used in the study.

1. For each of the following treatments, check YES if the treatment will appear in the experiment, otherwise check NO.

   i. $(A, B, C, D, E, F) = (H, L, H, H, L, H)$ ______YES ______NO

   ii. $(A, B, C, D, E, F) = (H, L, H, L, H, H)$ ______YES ______NO

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

3. What effects which must be assumed to be negligible in order to determine whether there is an interaction between Factors A and B using the data from this experiment.

Problem IV. (36 points) **Place** one of (A, B, C, D, E) corresponding to the **BEST** answer on the **ANSWER SHEET - Page 10**. Only **ONE LETTER** should be used for each question.

(1) A scientist is designing a study to compare clover accumulation for five different planting densities. A field is divided into 20 plots with 4 plots assigned to each density. Why is it necessary to randomly assign the planting densities to the plots?

- A. To maximize the potential of determining differences.
- B. To minimize the heterogeneity in the variances.
- C. Randomization is necessary since all the plots are in the same field.
- D. To avoid government intervention.
- E. To provide appropriate reference populations for statistical inference.

(2) A veterinarian wants to investigate $t = 5$ treatments for controlling heartworms in puppies. Her consulting statistician determines that she will need $r = 9$ replications per treatment. There is enormous variation in the effectiveness of the treatment so the veterinarian wants to use groups of homogeneous puppies and decides to use 9 litters of puppies as her blocking variable with 5 puppies per litter. From the experiment, the following results were obtained: $MS_{Block} = 34.2$ and $MS_{Error} = 11.4$. The veterinarian wants to conduct a second experiment without blocking, a CRD with the same 5 treatments. What is the minimum number of replications (puppies) per treatment that would be needed to achieve the same level of accuracy in estimating the treatment means in the CRD as she obtained using the RCBD?

- A. 5
- B. 9
- C. 13
- D. 62
- E. It would be impossible to determine with the given information.

(3) If a Bonferroni F-test is used to test the significance of 25 orthogonal contrasts constructed from 40 treatment means, at what value must α_{PC} be set in order to obtain $\alpha_F = .05$?

- A. .05000
- B. .00125
- C. .00128
- D. .00200
- E. .00205

(4) In a study of the differences in the mean potassium content of five varieties of bananas, a nutritionist decides to randomly select the same number of banana plants from each of the varieties. Why is it a good design principal to have equal replication in the study.

A. obtain an exact F distribution for the test statistics.

B. simplify the form of the expected mean squares

C. increase the power of the F-test when the treatments have unequal variances.

D. protect the level of the F-test for the cases when the treatments have unequal variances.

E. all of the above

(5) In a study to determine if there is a difference in the mean strength of cotton fibers produced by the thousands of fiber manufacturers located in North America, the researcher randomly selects 10 manufacturers to be included in the study. From each manufacturer, 20 samples of cotton fiber are randomly selected from their warehouse and a tensile strength measurement is determined for each sample. The researcher used Tukey's HSD procedure to determine which pairs of manufacturers had significantly different mean responses. What is a major criticism of the researcher's methodology?

A. Tukey's HSD procedure is too conservative to detect small differences in the means.

B. Tukey's HSD procedure will result in an inflated Type II error rate in the selection of pairs of differences.

C. Tukey's HSD procedure should only be used when the treatment levels have fixed effects.

D. Tukey's HSD procedure will result in an inflated Type I error rate in the selection of pairs of differences.

E. None of the above, the use of Tukey's HSD is an appropriate procedure.

(6) After conducting a CRD with the factors F_1 and F_2 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 F_1 and F_2?

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

B. Use the transformation, because you can just invert the transformation to obtain results in the original scale.

C. Use the Kruskal-Wallis rank based procedure with its multiple comparison procedure to compare the pairs of levels of F_1 at each level of F_2.

D. Use the Friedman rank based procedure with the blocking factor designated as F_2.

E. None of the above would be valid.

(7) In a CRD with a quantitative factor F_1 at 4 levels and a qualitative factor F_2 at 2 levels, the researcher wants to know if there is a linear trend in the mean responses across the levels of F_1. The AOV table reveals a non-significant $F_1 * F_2$ 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. none of the above

(8) The following model was fit to the experimental data from a RCBD with repeated measures. There were b blocks, t fixed treatment levels having r reps each. Each EU was measured at p time points. Let y_{ijkl} be the measurement of lth EU receiving treatment j in block i during time period k.

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

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

A. b_i's are independent, c_{ij}'s are independent, and e_{ijkl}'s are independent

B. b_i's are independent, c_{ij}'s are independent, and e_{ijkl}'s are correlated

C. b_i's are independent, c_{ij}'s are correlated, and e_{ijkl}'s are independent

D. b_i's are correlated, c_{ij}'s are independent, and e_{ijkl}'s are independent

E. b_i's are correlated, c_{ij}'s are correlated, and e_{ijkl}'s are correlated

(9) A researcher is designing a CRD experiment having 2 factors: Factor A with 2 fixed levels and Factor B with 3 fixed levels and 10 reps per treatment. The researcher wants to determine if this would be a sufficient number of reps in order that an $\alpha = .05$ F-test would have power of at least .80 to detect a difference of 19 units in the mean responses of at least two of the treatments. From past experiments, an estimate $\hat{\sigma}_e^2 = 100$ is provided. What would be the power of the F-test?

A. $0 \leq Power \leq .5$

B. $.5 < Power \leq .65$

C. $.65 < Power \leq .8$

D. $.8 < Power \leq 1.0$

E. Cannot be determined with the given information

(10) In a Crossover Design, both the sequence effect and the carryover effect were highly significant ($p < .0001$). The treatment effect

A. cannot be tested because it is completely confounded with the carryover effect.

B. cannot be tested because it is completely confounded with the sequence effect.

C. cannot be tested because it is completely confounded with both the sequence and carryover effect.

D. should be tested using only of the data observed during the first time period of each sequence.

E. all of the above

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

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

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

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

E. none of the above

(12) A chemical company wishes to study the difference in response times (in milliseconds) for a number of different types of circuits used in an automatic value shutoff mechanism. From past studies, the value of σ_e is taken to be 2 milliseconds. The researcher has a list of over 100 Types of Circuits that are of interest to the company. The company wants to determine if there is a significant variation in the performance of the 100 Types of circuits. In order to control for the variation within each Type of Circuit, she decides it is necessary to evaluate 4 circuits of each Type selected for the study. What is the smallest number of Types of Circuits that must be selected for use in the experiment in order to obtain an α=.01 test having power of at least 0.85 whenever the standard deviation in the Types of circuits is greater than 2.5 milliseconds.

A. 5

B. 6

C. 7

D. 9

E. cannot be determined from the given information

ANSWER SHEET FOR MULTIPLE CHOICE QUESTIONS

Name ____

Place your answer to each of the MULTIPLE CHOICE questions in the space provided.
Use upper case letters (A, B, C, D, E) with only **ONE** answer per question:

(1.)_________

(2.)_________

(3.)_________

(4.)_________

(5.)_________

(6.)_________

(7.)_________

(8.)_________

(9.)_________

(10.)_________

(11.)_________

(12.)_________