Applied Statistics Qualifier Examination
(Part II of the STAT AREA EXAM)

August 19, 2020; 11:15AM-1:20PM EDT

General Instructions:

(1) The examination contains 4 Questions. You are to **answer 3 out of 4** of them. *** Please only turn in solutions to 3 questions ***

(2) You may use up to 4 books and 4 class notes, plus your calculator and the statistical tables.

(3) NO computer, internet, cell phone, or smart watch is allowed (other than serving as Zoom connections & receiving/sending your exam to Prof. Zhu).

(4) ***This is a 2-hour exam*** **11:15am- 1:15 PM – Please scan your solutions as a PDF file & email to Prof. Wei Zhu wei.zhu@stonybrook.edu by 1:20pm.**

Please be sure to fill in the appropriate information below:

I am submitting solutions to QUESTIONS _____, _____, and _____ of the applied statistics qualifier examination. Please put your name on every page of your exam solutions, and add page number for solutions to each question individually.

There are _____ pages of written solutions.

Please read the following statement and sign below:

Academic integrity is expected of all students at all times, whether in the presence or absence of members of the faculty. Understanding this, I declare that I shall not give, use, or receive unauthorized aid in this examination.

_____
(Signature)

_____
(Name)

_____
(SBU ID)

Special Zoom Based Exam Instructions:

- This exam is conducted via Zoom from 11:15 am to 1:15 pm EDT with exam scanned as a PDF file and emailed to Prof. Wei Zhu by 1:20pm EDT.
- If you have a question during the exam, please send a chat message to the hosts privately. Both Prof. Pei-Fen Kuan and Prof. Song Wu will be proctoring the exam too, so you can send chat messages to each one privately if you have questions on their problems.
- The entire Zoom meeting and chat messages are being recorded.
- You should join the Zoom meeting from two devices: Your computer/laptop/tablet (with webcam), and your smartphone (with camera).
- Audio should be muted and video must be kept on during the exam.
- Your computer webcam must fully show your face; your smartphone camera should show your hands and workspace, with the pages of paper being used for the exam.
- At the very beginning of the exam, during set up, you will be asked to do a brief "environment scan", showing the workspace where your computer is and the desk/table/floor where you will be writing your work.
- You are required to bring enough blank pieces of paper to use for the exam. You will show the blank pages at the beginning, during the "environment scan" on Zoom.
- You are not allowed to use the internet for any searching or communication with others, with the sole exception of communication privately with the proctors via Zoom chat (which is set so that your chats only go privately to hosts, not to others).
- It is recommended that you print the exam and write your answers on it. However, you can write your answers on your blank papers if you do not have a printer with you.
- After you finish the exam, scan your pages, ordered, oriented and numbered appropriately, into a single PDF file. Please email the PDF file to Prof. Wei Zhu at: wei.zhu@stonybrook.edu no later than 5 minutes after completion of the exam (i.e., by 1:20 pm EDT).
- No students are allowed to leave the Zoom meeting until the exam is over.
- If you finish the exam early, then submit your exam and remain in the Zoom meeting until the conclusion of the exam at 1:15 pm EDT.
- If the answers are not submitted by 1:20 pm EDT, the exam will not be graded, and a score of zero will be given.
- Now please take a deep breath, and do your best in your exam – all your hard work will pay off, and we wish you the best of luck!

Name: _____

1. An experiment was conducted to compare the cooking time (in minutes) of five brands of rice, with four replicates per brand. The data is given in the table below.

Brand of Rice				
A	B	C	D	E
40	38	44	41	34
45	40	42	43	35
46	38	40	40	34
49	44	34	40	33

a) Estimate the parameters in the one-way ANOVA model

$$y_{ij} = \mu + t_i + \epsilon_{ij}, \qquad i = A, B, \dots, E; \quad j = 1, \dots, 4$$

b) Construct 90% simultaneous confidence intervals for the differences between all pairs of mean cooking time among the five brands.

c) Define $C_1 = 4t_B - t_A - t_C - t_D - t_E$, $C_2 = t_A + t_C - t_D - t_E$, $C_3 = t_A - t_C$, and $C_4 = t_D - t_E$.

i) Do the least-squares estimates of these contrasts form a set of orthogonal contrasts? Justify your answer.

ii) Give a verbal description of each contrast.

iii) Compute the sum of squares for each contrast (SS_{C_k}) and the treatment sum of squares (*SSa*). Can you find out the relationship between SS_{C_k}'s and *SSa*? (Hint: for contrast C_k, $SS_{C_k} = (\sum_i c_{ki}\hat{\mu}_i)^2/(\sum_i {c_{ki}}^2/n_i)$, where $\hat{\mu}_i$ is the estimated mean of brand *i*)

Name: _____

2. Let Y_i = 1 or 0 denote the observed binary response for unit i with $\pi_i = P(Y_i = 1)$, $i = 1, \ldots, n$. Suppose R_i is the unobserved true response for unit i with $\pi_i^* = P(R_i = 1)$ satisfying the linear logistic model

$$logit(\pi_i^*) = \beta^T x_i,$$

and the observed response Y_i is subject to mis-classification as follows:

$$P(Y_i = 0|R_i = 1) = \delta_i$$
$$P(Y_i = 1|R_i = 0) = \epsilon_i.$$

a. Suppose δ_i and ϵ_i are known quantities, please write out the likelihood function and describe how β can be estimated.
b. If δ_i and ϵ_i are unknown, is it possible that under some condition(s), the logistic model $logit(\pi_i) = \beta^T x_i$ can still be valid? If yes, please find the condition(s); if not, please prove it.

Name: _____

3. Let $Y = X\beta + \varepsilon$, where Y is an $n \times 1$ vector of random variables, X is an $n \times p$ matrix of known constants of full column rank p ($p < n$), β is a $p \times 1$ vector of unknown constants, and ε is an $n \times 1$ vector of normally distributed random variables with $E(\varepsilon) = 0$ and variance-covariance matrix $\sigma^2 V,\ 0 < \sigma^2 < \infty$, with V a known positive definite $n \times n$ matrix. Find
$E(Y^T V^{-1} X (X^T V^{-1} X)^{-1} X^T V^{-1} Y)$.

Name: _____

4. An educational program to improve student performance in a mathematical subject will use individual tutors to supplement standard classroom instruction. A research team will conduct a pilot study using a balanced two-way layout with 8 observations per instructor-tutor combination to study the components of variance associated with two random factors *A* (classroom instructor) and *B* (student tutor). The dependent variable *Y* was the improvement in a student's score in a mathematics curriculum. They used the standard model: $Y_{ijr} = \mu + A_i + B_j + (AB)_{ij} + \sigma_E Z_{ijr}$where $Y_{ijr}$was the value of the *r*th student score ($i = 1, \cdots, 4;\ j = 1, \cdots, 5;\ r = 1,2, \cdots, 8$). Here, A_i is the random effect associated with the *i*th classroom instructor ($i = 1, \cdots, 4$), where $\{A_i\}$ are normally and independently distributed with mean 0 and variance σ_A^2. Additionally, B_j is the random effect associated with the *j*th student tutor ($j = 1, \cdots, 5$), where $\{B_j\}$ are normally and independently distributed with mean 0 and variance σ_B^2. The random variable $(AB)_{ij}$ is the random effect associated with the interaction of the *i*th classroom instructor with the *j*th tutor, where $\{(AB_{ij})\}$ are normally and independently distributed with mean 0 and variance σ_{AB}^2.The error random variables $\{Z_{ijr}\}$are standard normal random variables that are independently distributed. Each set of random variables is independent of every other set. The research team hypothesizes that the variance $\sigma_A^2 = 150$, the variance $\sigma_B^2 = 40$, the variance $\sigma_{AB}^2 = 20$, and the variance $\sigma_E^2 = 200$

Display the analysis of variance table for this study, including the expected mean squares. What is the correct test statistic? What is the probability of a Type II error using this test statistic using the hypothesized values of the variances when the team tests the null hypothesis that $\sigma_A^2 = 0$ at the 0.01 level of significance?