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

May 25, 2022; 11:15AM-1:15PM ET

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.

(4**)** ***This is a 2-hour exam*** **11:15am- 1:15 PM – Please submit your exam by 1:15pm**. Please be sure to turn in this cover page. Good luck!

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)

Name: ______________________   Signature: ______________________

1. A research group is interested in analyzing caffeine content in a particular coffee brand. The following data summarize the caffeine content of 175 cups of coffee.

Caffeine content	Number of cups of coffee
$52 \leq x < 56$	7
$56 \leq x < 60$	22
$60 \leq x < 64$	36
$64 \leq x < 68$	45
$68 \leq x < 72$	33
$72 \leq x < 76$	28
$76 \leq x < 80$	4

(a) Based on the data above, estimate the mean and variance of caffeine content.

(b) Test whether the caffeine content of this coffee brand is normally distributed at $\alpha = 0.10$. Show your work and justify your answer.

Name: ____________________ Signature: ____________________

2. A survey asked a very large group of students in their final year of high school whether they had ever used alcohol, cigarettes, or marijuana. The data can be summarized into a 2x2x2 table by A for alcohol use, C for cigarettes, and M for marijuana, where for each variable category 1 is 'yes' and category 2 is 'no'. Results of using SAS to fit the log-linear model symbolized by (AC, AM, CM) are given below.

(a) From the SAS output, report both the marginal and conditional odds ratios between A and M. Prove or disapprove that the marginal associations between A and M need to equal their conditional associations for this model.

(b) Suppose we view M as a response variable. State the corresponding logit model. Can you use the SAS output below to figure out the ML parameter estimate and standard error of its coefficient? If so, find their values.

(c) What conclusions can you draw from this analysis?

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Deviance	1	0.3740	0.3740
Pearson Chi-Square	1	0.4011	0.4011

Parameter			Estimate	Standard Error	Wald Chi-Square	Pr>ChiSq
Intercept			5.6334	0.0597	8903.96	<.0001
a	1		0.4877	0.0758	41.44	<.0001
c	1		-1.8867	0.1627	134.47	<.0001
m	1		-5.3090	0.4752	124.82	<.0001
a*m	1	1	2.9860	0.4647	41.29	<.0001
a*c	1	1	2.0545	0.1741	139.32	<.0001
c*m	1	1	2.8479	0.1638	302.14	<.0001

LR Statistics			
Source	DF	Chi-Square	Pr>ChiSq
a*m	1	91.64	<.0001
a*c	1	187.38	<.0001
c*m	1	497.00	<.0001

Name: ____________________________ Signature: ____________________________

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. The $n \times n$ identity matrix is denoted by $I_{n\times n}$.

What is $E(Y^T(I_{n\times n} - X(X^T X)^{-1} X^T) Y)$?

Name: ____________________________ Signature: ____________________________

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 5 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 ($i = 1, \cdots, 3$; $j = 1, \cdots, 6$; $r = 1, 2, \cdots, 5$). Here, A_i is the random effect associated with the *i*th classroom instructor ($i = 1, \cdots, 3$), 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 ($j = 1, \cdots, 6$), 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 $\sigma_B^2 = 60$, the variance $\sigma_{AB}^2 = 30$, and the variance $\sigma_E^2 = 250$.

Display the analysis of variance table for this study, including the expected mean squares. What is the correct test statistic for the null hypothesis that $\sigma_A^2 = 0$? What is the probability of a Type II error s 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?