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

May 24, 2023; 11:15AM-1:15PM EST

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 *** Please note that if you choose to do Question 4, you will have the flexibility to choose to do one of the three problems (4a, 4b, 4c) in that category, in this special transition period.

(2) You may use up to 4+2 books and 4+2 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 turn in by 1:15pm**.

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. Suppose we have three independent random samples from three normal populations. $X_1, X_2, \dots, X_{n_1} \sim N(\mu_X, \sigma_X^2)$, $Y_1, Y_2, \dots, Y_{n_2} \sim N(\mu_Y, \sigma_Y^2)$ and $Z_1, Y_2, \dots, Z_{n_3} \sim N(\mu_Z, \sigma_Z^2)$. Two hypotheses are to be tested simultaneously:

$H_0^A: 2\mu_X - \mu_Y - \mu_Z = 0$ vs $H_1^A: 2\mu_X - \mu_Y - \mu_Z \neq 0$

$H_0^B: \mu_X + \mu_Y - 2\mu_Z = 0$ vs $H_1^B: \mu_X + \mu_Y - 2\mu_Z \neq 0$

(a) At an overall significance level $\alpha = 0.05$, derive the test statistics and decision rules for testing these two hypotheses.

(b) Suppose $\sigma_X^2 = 2\sigma_Y^2 = 4\sigma_Z^2$ and the unequal sample allocation $n_1 = n_2 = 2n_3$, where σ_X^2 is known. Derive the sample size formula to ensure a power >80% at an overall $\alpha = 0.05$ for testing the two hypotheses.

Name: ______________________________ Signature: ______________________________

2. For a I × J two-way contingency table with row and column random variables denoted as X and Y, one kind of association measure between X and Y, called *tau*, is calculated based on the so-called variation measure

$$V(Y) = \sum_{j=1}^{J} \pi_{+j}\left(1 - \pi_{+j}\right) = 1 - \sum_{j=1}^{J} \pi_{+j}^2$$

(a) Show $V(Y)$ is the probability that two independent observations on Y fall in different categories.

(b) Show that $V(Y) = 0$ when $\pi_{+j} = 1$ for some j and $V(Y)$ takes maximum value of $(J - 1)/J$ when $\pi_{+j} = 1/J$ for all j.

(c) For the proportional reduction in variation, show that $E[V(Y|X)] = 1 - \sum_i \sum_j \frac{\pi_{ij}^2}{\pi_i}$.

Name: ____________________________ Signature: ____________________________

3. Let X_1 be an $n \times p_1$ matrix of known constants with $0 < p_1 < n$ and $rank(X_1) = p_1$. Let X_2 be an $n \times p_2$ matrix of known constants with $0 < p_2 < n$ and $rank(X_2) = p_2$. Let $rank([X_1\ X_2]) = p_1 + p_2,\ p_1 + p_2 < n$.

Let β_1 be a $p_1 \times 1$ vector of unknown constants. Let β_2 be a $p_2 \times 1$ vector of unknown constants. Let Y be an $n \times 1$ vector of random variables with

$$E(Y) = X_1\beta_1 + X_2\beta_2.$$

The variance-covariance matrix of Y is $vcv(Y) = \sigma^2 V$, where V is a positive definite symmetric $n \times n$ matrix. Find $E[Y^T(I_{n\times n} - X(X^T X)^{-1}X^T)Y]$.

Name: ____________________________ Signature: ____________________________

4a. A research team has been hired by a county government to assess an education program using a three-stage nested design. There is an outcome measure of the quality of a student's work called *Y*. There will be a random sample of *I*=4 school districts. There will be a random sample of *J=6* schools withing each district. There will be a random sample of $K = 2$ classes within each school. Finally, there will be a random sample of $R = 8$ students from each class. The researchers will use the model:

$$Y_{ijkr} = \mu + A_i + B_{j(i)} + C_{k(ij)} + \sigma_E Z_{(ijk)r}$$

The random variables $\{A_i\}$ are normally and independently distributed with expected value 0 and variance σ_A^2. The random variables $\{B_{j(i)}\}$ are normally and independently distributed with expected value 0 and variance σ_B^2. The random variables $\{C_{k(ij)}\}$ are normally and independently distributed with expected value 0 and variance σ_C^2. The random variables $\{Z_{(ijk)r}\}$ are independent and identically distributed normal random variables with mean 0 and variance 1. The sets of random variables $\left\{\{A_i\}, B_{j(i)}, \left\{C_{k(ij)}\right\}, \{Z_{(ijk)r}\}\right\}$ are independent of each other.

The research team believes that $\sigma_A^2 = 30, \sigma_B^2 = 40, \sigma_C^2 = 50, \sigma_E^2 = 80.$ What is the test of the null hypothesis that $\sigma_A^2 = 0$? Specify its null and alternative distributions. Find the probability of a Type II error for this test using level of significance 0.01. under the specification given above.

Name: ____________________________ Signature: ____________________________

4b. Consider the VAR (1) model:

$$\begin{pmatrix} Y_{1,t} \\ Y_{2,t} \end{pmatrix} = \begin{bmatrix} 0.7 & 0.1 \\ 0.1 & 0.7 \end{bmatrix} \begin{pmatrix} Y_{1,t-1} \\ Y_{2,t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix}, \text{ where } \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix} \overset{i.i.d.}{\sim} \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix} \right)$$

(a) Is this model in structural or reduced form? Please justify your answer.
(b) Is this VAR(1) series stationary? Please show the entire derivation.
(c) Please derive the symmetric structural form of this VAR(1).
(d) Please derive the asymmetric structural form of this VAR(1) using the Cholesky decomposition. Please make sure that in this form, $Y_{1,t}$ would depend on $Y_{2,t}$, while the other way is not true.

Name: __________________________ Signature: __________________________

4c. For a feedforward neural network model with no hidden layer and the identity activation function, when would the loss function of "sse" and "ce" yield identical results? Please show the entire derivation for full credit.