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

May 26, 2021; 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.
- Please set up your computer and your cellphone by 11:10 am EDT so Prof. Wei Zhu can email you the applied stat problems soon after the environment scan.
- If you have a question during the exam, please send a chat message to the hosts privately.
- 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 computer monitor, your hands and workspace, with the pages of paper being used for the exam.
- The gallery view must be kept off.
- 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.
- Please put down your name and signature on every page of the exam.
- 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: ______________________ Signature: ______________________

1. A research group conducted a study to estimate the number of passengers (other than the driver) per car passing through a busy intersection. The following data give the number of passengers collected between 8 AM and 3PM on May 07, 2021.

Number of passengers	0	1	2	3	4	≥ 5
Frequency	443	314	148	60	24	11

The research group was unsure whether a Poisson or a Geometric distribution fit the data better.

a) Using only proportion of zeroes and ones (i.e, 443/1000 and 314/1000), can you propose an estimate for λ assuming the data follows a Poisson distribution $Po(\lambda)$?

b) Based on the estimate obtained from (a), test whether the number of passengers per car passing through a busy intersection follow a Poisson distribution on the entire dataset.

c) Consider a geometric distribution *Geo(p)* for the number of occupants (defined as number of passengers +1). Can you propose an estimate of p?

d) Based on the estimates obtained from (c), test whether the number of occupants per car passing through a busy intersection follow a Geometric distribution on the entire dataset.

e) Which distribution fits the data better?

Name: ___________________________ Signature: ___________________________

2. Individual players $\{P_i, i = 1, 2, \ldots, N * C\}$ were organized into N races, with each race competed by C players ($C > 2$). Each player participates in one and only one race, and each race has one and only one winner. The data look like below:

Race	Player 1	Player 2	…	Player C	Winner
1	P_1	P_2		P_C	y_1
2	P_{C+1}	P_{C+2}		P_{2*C}	y_2
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
N	$P_{(N-1)*C+1}$	$P_{(N-1)*C+2}$		P_{N*C}	y_N

where $y_i \in \{P_{(i-1)*C+1}, P_{(i-1)*C+2}, \ldots, P_{i*C}\}$. Each player also has K features $\{X_i, i = 1, 2, \ldots, K\}$, representing their own characteristics that are associated with their race performance, e.g. past record, height, weight etc.

Player	$\boldsymbol{X_1}$	$\boldsymbol{X_2}$	…	$\boldsymbol{X_K}$
1	x_{11}	x_{12}		x_{1K}
2	x_{21}	x_{22}		x_{2K}
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
N*C	$x_{N*C,1}$	$x_{N*C,2}$		$x_{N*C,K}$

a) Construct a logistic regression model that uses these features to model the probability of winning in a race. Clearly write out the logistic model, the likelihood function, and how to solve the MLEs.

b) If C random players were chosen to form a new race, how would you predict the winning probability of each player?

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 $\sigma^2 I_{n\times n}$, $0 < \sigma^2 < \infty$, σ^2 not known. Let $\hat{\beta}$ be the OLS estimate of β. What is $E\{\hat{\beta}^T\hat{\beta}\}$?

Name: ______________________________ Signature: ______________________________

4. A research team has seven factors *A*, *B*, *C*, *D, E*, *F,* and *G* that they believe may affect the value of a dependent variable *Y*. They seek to minimize $E(Y)$. Each factor has two settings: high (+) and low (-). They used a 2_{III}^{7-4} design using the design generators recommended in Montgomery; namely, *D=AB, E=AC, F=BC,* and *G=ABC.* They obtained the results in Table 1 below.

a) Give the analysis of variance table describing these results.
b) What should the research team conclude? Use level of significance 0.01.
c) What is the optimal setting of the seven factors based on these results?
d) What are the advantages and limitations of this design?

Table 1
Experimental Results for 2_{III}^{7-4} Design

A	*B*	*C*	*D*	*E*	*F*	*G*	*Y*
-	-	-	+	+	+	-	682
+	-	-	-	-	+	+	1194
-	+	-	-	+	-	+	1302
+	+	-	+	-	-	-	546
-	-	+	+	-	-	+	1244
+	-	+	-	+	-	-	1486
-	+	+	-	-	+	-	1268
+	+	+	+	+	+	+	1096