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

January 24, 2018; 11:00AM-1:00PM

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, or cell phone is allowed in the exam.

(4) ***This is a 2-hour exam due by*** **1:00 PM**.

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:

This is to certify that I have taken the applied statistics qualifier and have used no other person as a resource nor have I seen any other student violating this rule.

_____
(Signature)

_____
(Name)

Name: _____

1. It is known that patients with HIV infection have low CD4T immune cell counts. In addition, CD4T also decreases with several other factors, including aging process. A new drug was developed to enhance the immune system of HIV positive patients. 30 male patients from the Stony Brook Hospital were recruited in the study and assigned to one of the two treatment groups, the first group receiving the new drug and the second group receiving the standard drug. The changes in CD4T cell counts (after treatment – before treatment) were recorded and the data are shown below. Additional information including age is also provided.

New Drug			Standard Drug	
CD4T	Age		CD4T	Age
100	56		101	69
101	64		101	20
99	58		95	56
98	64		101	29
101	43		99	64
100	28		98	57
101	29		102	43
100	45		100	56
99	56		101	45
98	69		104	28
101	22		100	22
102	28		99	64
97	57		102	40
100	20		104	58
100	40		101	28

(a) Carry out a formal statistical test to conclude if the new drug is effective. Justify your approach by providing details on assumptions, derivations and limitations.

(b) The research group aims to expand the study to evaluate the effectiveness of this new drug on young HIV positive men of age <30. They have resources to recruit 100 patients. How would you allocate these 100 patients to the two treatment groups to maximize the power of detecting a mean difference D_1? Justify your approach by providing details on assumptions, derivations, limitations, etc (You may use the information from Stony Brook Hospital as your preliminary data)

Name: _____

2. Suppose there are n subjects, and each subject i $(i = 1, \dots, n)$ has a pair of matched responses (y_{i1}, y_{i2}), where $y_{it} \in \{0, 1\}$ for $t = 1, 2$ representing two experimental conditions. Consider the following logistic model:

$$logit[P(y_{it} = 1)] = \alpha_i + \beta I(t)$$

Where I is an indicator function: $I(t) = \begin{cases} 1 & t = 2 \\ 0 & t = 1 \end{cases}$

a. Write out the maximal likelihood function and find the sufficient statistics for $\{\alpha_i\}$
b. Conditioning on sufficient statistics for $\{\alpha_i\}$, derive the conditional maximum likelihood estimation of β and its asymptotic variance.
c. Comment briefly on whether all subjects have equal importance to the model and why.

Name: _____

3. A research team will take J observations at each of the seven settings of x (for a total of 7J observation): $x = -3,-2,-1,0,1,2,3$ At each setting x, the dependent variable Y is given by the model: $Y = \beta_0 + \beta_1 x + \sigma Z$, where Z is $N(0,1)$. Each error term is independent of any other error term. Find the expected value of the sum of squares due to regression.

Name: _____

4. A research team sought to compare three curricula to teach introductory calculus. These curricula are called *A*, *B*, and *C*. The outcome measure *Y* of the quality of a student's performance is the result of a standardized examination. The research team taught each curriculum at two colleges (so that 6 colleges were studied). In each college, a random sample of $R = 8$ students were tested. The research team observed the results in the table below. Test the null hypothesis that the effectiveness of each curriculum is the same. Also test the other hypotheses of interest. State each test statistic, its null distribution, and its alternative. State the usual assumptions. Assume that these assumptions are valid in your analysis.

Summary Statistics of the Study by Curriculum and Class

Curriculum	College	Number of students tested	Mean outcome measure	Variance outcome measure
A	1	8	151.6	432.3
A	2	8	269.3	389.4
B	3	8	399.7	346.8
B	4	8	255.4	502.4
C	5	8	68.9	467.2
C	6	8	171.2	383.1