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

January 25, 2017; 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. An experimenter is interested in the effects of two factors and their (potential) interaction on the quality of bread made in a bread machine. A bread machine is a popular kitchen appliance. To use it, you pour in the ingredients for bread, turn it on, and wait for approximately 4 hours. The result is a loaf of baked bread. This experimenter is interested in the effects of the quantity of yeast used (1 Tablespoon or 2) and the order in which the ingredients are put into the machine: wet first then dry or dry first then wet. He bakes 12 loaves of bread, 3 for each combination, in random order. He has two tasters for each loaf who score the loaf from 1 to 10 according to how it tastes (1 is awful and 10 is great).

a) Write down a model for this experiment (i.e., provide an equation giving the model for the responses, the range of all subscripts that appear in the equation and distributional assumptions of the variables).

b) Interpret all the parameters in your model (including the parameters in the distributional assumptions).

c) Derive the statistical tests to answer the questions of interest to the experimenter.

Name: ___________________________

2. Data from two clinical trials are provided in Table below.

Clinical Trial	Intervention Group	Number of Strokes	Follow-up (Person-Years)
1	Anticoagulant (A)	$Y_{1A} = 4$	$P_{1A} = 456$
	Control (C)	$Y_{1C} = 19$	$P_{1C} = 440$
2	Anticoagulant (A)	$Y_{2A} = 6$	$P_{2A} = 237$
	Control (C)	$Y_{2C} = 9$	$P_{2C} = 242$
Total	Anticoagulant (A)	$Y_{.A} = 10$	$P_{.A} = 693$
	Control (C)	$Y_{.C} = 28$	$P_{.C} = 682$

These trials examined the effect of anticoagulant treatment for the prevention of stroke in patients with valvular atrial fibrillation

(a) An estimate of the effect of anticoagulant treatment for the prevention of stroke is given by the observed stroke rate ratio $(Y_{.C}/P_{.C})/(Y_{.A}/P_{.A}) = (28/682)/(10/693) = 2.85$. Calculate a 95% confidence interval for the observed stroke rate ratio. Provide a brief summary of your analysis.

(b) Construct a test of the null hypothesis that the true effect of intervention is the same for both clinical trials (i.e. H_0: No interaction between treatment group and clinical trial). Clearly state any assumptions you have made. Briefly summarize the results of your analysis.

Name: ____________________________

3. Let (Y_i, X_i), $i = 1, \ldots, n$, denote the values of the response and $p - 1$ dimensional covariate vectors for a random sample of n subjects. We want to fit the regression model

$$Y = X\beta + \varepsilon,$$

where β is a p-dimensional vector of unknown regression coefficients to be estimated. Note that β includes the intercept term. Let **X** be an n × p matrix whose ith row is $(1, X_i)$ and **Y** be an n-dimensional vector whose ith entry is Y_i, $i = 1, \ldots, n$. Consider a ridge regression estimate

$$\hat{\beta}(\lambda) = (\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}^T \mathbf{Y},$$

where **I** is an n × n identity matrix, λ > 0 is a regularization parameter.

(a) Calculate $E(\mathbf{Y} - \mathbf{X}\hat{\beta}(\lambda))$. Under what condition on λ is this equal to zero?
(b) Calculate $\mathrm{Var}(\mathbf{Y} - \mathbf{X}\hat{\beta}(\lambda))$.
(c) Find the matrix such that the fitted value from fitting the ridge regression estimators is written as **HY**. Calculate **H**(**I** − **H**)**Y**.

Name: ___________________________

4. An educational program to improve student performance in a mathematical subject by using individual tutors to supplement standard classroom instruction. The research team used a balanced two-way layout with 12 observations per instructor-tutor combination to study the components of variance associated with two random factors *A* (classroom instructor) and *B* (student tutor). 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 rth student's improvement ($r = 1, \ldots, 12$; the dependent variable was improvement in a student's score in a mathematics curriculum). Here, A_i is the random effect associated with the *i*th classroom instructor ($i = 1,2$), where A_i is 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 ($j = 1,2,3$), where B_j is normally and independently distributed with mean 0 and variance σ_B^2 . Finally, $(AB)_{ij}$ is the random effect associated with the interaction of the *i*th classroom instructor with the *j*th tutor, where $(AB)_{ij}$ is normally and independently distributed with mean 0 and variance σ_{AB}^2. The error random variables are normally and independently distributed as well. They observed the results given in Table 1 below.

a. Display the analysis of variance table.
b. What statistics did you use to test each null hypothesis.
c. What conclusions do you reach about the importance of the factors? Use the 0.10, 0.05, and 0.01 levels of significance.

Table 1
Improvement in Student Performance
By Professor and Tutor
(12 replicates at each setting)

Instructor	Tutor	Mean	Variance
1	1	384.0	400.8
2	1	335.7	454.3
1	2	372.7	484.5
2	2	328.9	384.9
1	3	351.3	418.7
2	3	357.2	370.6