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

January 28, 2015; 11:00AM-1:00PM

<u>Instructions:</u>

(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**.

<u>Please be sure to fill in the appropriate information below:</u>

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.

<u>Please read the following statement and sign below:</u>

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)

____

Print your name here, please.

1. A drug antibiotic manufacturer randomly sampled 12 different locations in the fermentation vat to try and estimate the mean potency for the batch of antibiotic being prepared. Readings were as follows:

8.9, 9.0, 9.1, 8.9, 9.1, 9.0, 9.0, 8.8, 9.1, 8.9, 8.8, 9.2

(To ease the calculation, the sample mean and standard deviation are: $\bar{Y} = 8.983; S = 0.127$)

(a) Set up a 98% confidence interval for the mean potency for the batch and interpret the interval.

(b) A laboratory technician sampled 10 different locations in the same fermentation vat as the manufacturer above and her readings were as follows:

9.0, 9.0, 9.1, 8.9, 9.0, 8.8, 8.9, 9.1, 9.2, 9.0

(The sample mean and standard deviation for this sample are: $\bar{Y} = 9.0; S = 0.115$)

Set up a 98% confidence interval for the mean potency based on this new batch of data.

(c) The manufacturer and the technician are arguing about whose CI is better. They come to you for advice as to what you think they should do to come up with the best interval. What do you recommend they do? Explain your reasoning.

(You may need to use some of the following quantities: $Z_{0.01} = 2.326; Z_{0.02} = 2.053$;
$t_{0.01,8} = 2.896$; $t_{0.01,9} = 2.821$; $t_{0.01,10} = 2.764$; $t_{0.01,11} = 2.718$; $t_{0.01,12} = 2.681$;
$t_{0.02,8} = 2.449$; $t_{0.02,9} = 2.398$; $t_{0.02,10} = 2.359$; $t_{0.02,11} = 2.328$; $t_{0.02,12} = 2.303$.)

2. Suppose that a population of individuals can be partitioned into k sub-populations or groups, $G_1, \dots, G_k$, with relative frequencies $\pi_1, \dots, \pi_k$. Multivariate measurements Z made on individuals have the following distributions for the k groups:

$$G_j: Z \sim N_p(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}), \qquad j = 1, \dots, k.$$

Let $\boldsymbol{z}^*$ be an observation made on an individual drawn at random from the combined population. The prior odds that the individual belongs to G_j are $\pi_j/(1-\pi_j)$. Show that the posterior odds for G_j given $\boldsymbol{z}^*$ are:

$$odds(Y = j|\boldsymbol{z}^*) = \frac{\pi_j}{1-\pi_j} \times e^{\alpha_j + \boldsymbol{\beta}_j^T \boldsymbol{z}^*}.$$

(a) Find the expression for α_j and $\boldsymbol{\beta}_j$ in terms of $\boldsymbol{\mu}_j$ and $\boldsymbol{\Sigma}$.

(b) What simplifications can be made if the k Normal means $\boldsymbol{\mu}_j$ lie on a straight line in R^p?

(c) Comment briefly on the difference between the maximum likelihood estimation of α_j and $\boldsymbol{\beta}_j$ via the normal-theory likelihood and estimation via logistic regression.

3. A research team has four factors *A*, *B*, *C*, and *D* that it believes 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_{IV}^{4-1} design setting *D*=*ABC* and obtained the results in the table below. What should the research team conclude? Be sure to discuss the optimal setting of the independent variables.

Table for Problem

Experimental Results for 2_{IV}^{4-1} Design

A	*B*	*C*	*D*=*ABC*	*Y*
-	-	-	-	341
+	-	-	+	338
-	+	-	+	344
+	+	-	-	336
-	-	+	+	343
+	-	+	-	247
-	+	+	-	333
+	+	+	+	242

4. A research team will examine ten settings of the independent variable *x*: $x = 1, 2, \ldots, 10$ They will observe J independent observations at each setting of *x*. The dependent variable *Y* is related to *x* by the model: $Y = \beta_0 + \beta_1 x + \sigma Z$, where *Z* is *N*(0,1). All random errors are independent of each other. The research team knows that $\sigma^2 = 400$. The research team will test the null hypothesis $H_0 : \beta_1 = 0$ using level of significance $\alpha = 0.01$. The research team knows that $\sigma^2 = 400$ and seeks to set *J* so that β, the probability of a Type II error is 0.01. How many observations *J* at each setting are necessary so that $\beta = 0.01$ when $\beta_1 = 1$?