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

May 29, 2019; 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, cell phone, or smart watch is allowed.

(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 experiment was conducted to compare five fluids that are supposed to prevent the build-up of lactic acid in long distance runners. For reasons unrelated to the fluids, there were an unequal number of runners assigned to the treatment. Fluid #1 was plain water. Fluids #2 and #3 were common sports drinks, A, at low and high concentrations. Fluids #4 and #5 was another sports drink, B, at low and high concentrations.

The sample means, sample variances and sample sizes are shown in Table 1.

			Fluid		
	1	2	3	4	5
Mean	33.3	32.6	30.9	29	26.1
Variance	13.1	14.2	12.2	13.9	14.2
Sample size	10	7	10	8	6

a) Construct an ANOVA for testing the hypothesis of equal fluid means.
b) Consider the following contrasts:

C1: $\mu_1 - \frac{\sum_{j=2}^{5} \mu_j}{4}$
C2: $\frac{\mu_2+\mu_3}{2} - \frac{\mu_4+\mu_5}{2}$
C3: $\mu_2 - \mu_3$
C4: $\mu_4 - \mu_5$

where μ_j denotes the mean for fluid j (j=1, 2, …, 5).

i) Give a verbal description of each contrast.
ii) Are these set of contrasts orthogonal? Justify your answer.
iii) Construct a 95% simultaneous confidence interval for these contrasts.

Name: ____________________________

2. Consider a two-class logistic regression problem with $Y \in (0,1)$ and $x \in R$.

(a) Characterize the maximum-likelihood estimates of the slope and intercept parameter if the sample $\{x_i, i = 1, \dots N\}$ for the two classes are separated by a point $x_0 \in R$.

(b) Generalize this result to $x \in R^p$.

Name: ____________________________

3. A research team sought to find the dosage that minimized the expected response variable. They randomly assigned 25 animals to each of six settings of dosages of a supplemental diet and observed the response *Y*, with the results shown in the table below.

Table. Experimental Results

Dose *i*	Animals with dose *i* (J_i)	Mean outcome ($y_{i\bullet}$)	Within treatment variance (s_i^2)
1	25	313.6	9,432.4
2	25	278.8	10,387.1
3	25	235.9	9,452.7
4	25	211.9	11,526.4
5	25	176.6	8,408.5
6	25	146.4	9,498.9

The grand average was $y_{\bullet\bullet} = 227.2$, and the total sum of squares was 1,897,101.5. With regard to the orthogonal polynomials, the estimated linear contrast was -1166.6, and its coefficients were $-5,-3,-1,1,3,5$. The estimated quadratic contrast was 53.4, and its coefficients were $5,-1,-4,-4,-1,5$. The estimated cubic contrast was -24.6, and its coefficients were $-5,7,4,-4,-7,5$. The estimated fourth degree contrast was -10.6, and its coefficients were $1,-3,2,2,-3,1$. The estimated fifth degree contrast was 103.8, and its coefficients were $-1,5,-10,10,-5,1$.

What is the degree of the polynomial in dosage that has adequate fit for this data? Show your work, including the lack of fit test and analysis of variance table for your answer. Is there an optimal setting of the dosage? If so, what is it?

Name: ___________________________

4. A research team had four factors *A*, *B*, *C*, and *D* that it believed may affect the value of a dependent variable *Y*. Each factor had two settings: high (+) and low (-). The research team sought that setting of the four factors that maximized the expected outcome variable.

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? Use the 0.01 level of significance.

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

A	*B*	*C*	*D*=*ABC*	*Y*
-	-	-	-	433
+	-	-	+	857
-	+	-	+	462
+	+	-	-	834
-	-	+	+	426
+	-	+	-	879
-	+	+	-	467
+	+	+	+	865