Stat 5810 Applied Spatial Statistics **Fall Semester 2000**

Homework Assignment 1 (8/28/2000)

30 Points - Due 2nd Week of class

STAT 3000

FINAL EXAM

NAME: _____

The total number of points for this exam is 300. Your grade on this exam will account for 30% of your total course grade.

The exam consists of 7 problems. The maximum number of points for each problem and for each partial question is given in parentheses.

Write in the spaces provided below each question. If necessary, write also on the back. Write neatly.

Show as much work as possible in order to get the full number of points. Write down each step you are taking**. If you use a calculator, write down the correct formula you are using to get the full number of points**. **Just the end result, even if it is entirely correct, is not enough.**

You have one hour and 50 minutes to complete this exam. Read the questions carefully. Start with the question which is easiest for you and then move on to more difficult problems.

Some of the questions are multi-part questions, in which an answer to part a) is needed to answer the following parts. If you cannot answer part a) and you feel that you cannot proceed without knowing this answer, make up a (plausible) answer and proceed to part b) etc.

Problem I (40 points)

About 1% of the people in a large population carries the AIDS virus. ELISA tests are used to detect presence of the AIDS virus in blood. When AIDS virus is present, the ELISA test is positive with probability of 0.997. When the AIDS virus is not present, the ELISA test gives a false positive result with probability of 0.015.

a) (10 pts) What is the probability that the ELISA test for AIDS is positive for a randomly chosen person from this population?

b) (15 pts) What is the probability that a person carries the AIDS virus given that the ELISA test is positive?

c) (15 pts) What is the probability that a person does not carry the AIDS virus given that the ELISA test is negative?

Problem II (40 points)

Five samples of a ferrous-type substance were used to determine if there is a difference between a laboratory chemical analysis and an X-ray fluorescence analysis on the iron content. Each sample was split into two subsamples and the two types of analysis were applied. The results of both analyses are given in Table 1.

Table 1: Iron content (%) obtained from 5 samples using two types of analysis.

Analysis	Sample				
	1	2	3	4	5
X-ray	2.0	2.0	2.3	2.1	2.4
Chemical	2.2	1.9	2.5	2.3	2.4

a) (25 pts) Do the two methods of analysis give different results? State the appropriate hypotheses, perform an appropriate statistical test, obtain the approximate P-value (use one of the tables provided!), and state the conclusions. Explain briefly your results to a lab technician who does not have any knowledge of statistics.

b) (15 pts) Why is this a paired experiment? Why did the researcher decide to use paired design for this experiment? Explain briefly what factors could have caused bias in the results of the experiment if independent samples had been used.

Problem III (40 points)

For their out-of-class STAT project, Nicole, Troy, and Terry from STAT 3000 - Section 002 decided to analyze hardness of water in the Widstoe Chemistry Building at the USU. They analyzed nine (9) 50ml samples of tap water and obtained the average hardness of 205.22 ppm $CaCO_3$ and the standard deviation of 2.17 ppm $CaCO_3$.

a) (15 pts) Construct a two-sided 95% confidence interval for the mean hardness of water in the Widstoe Chemistry Building.

b) (5 pts) Would a 99% two-sided confidence interval for mean water hardness be shorter or longer that the 95% confidence interval found in a)? You do not have to construct this interval - just answer the question.

c) (20 pts) Reported water hardness in Logan and surrounding is 190 ppm $CaCO_3$. Based on the results of our students' experiment, is there any evidence that hardness of water in the Widstoe Chemistry Building is **greater** than that reported for Logan and surrounding?

Problem IV (35 points)

Suppose that X_1 and X_2 are independent random variables with $E(X_1) = \mu$, $Var(X_1) = 4$, $E(X_2) = \mu$, and $Var(X_2) = 6$. Consider the point estimate of μ

$$\hat{\mu}_1 = \frac{X_1}{2} + \frac{X_2}{2}.$$

a) (10 pts) What is the bias of $\hat{\mu}_1$?

b) (10 pts) What is the variance of $\hat{\mu}_1$?

c) (15 pts) What value of p minimizes the variance of

$$\hat{\mu} = pX_1 + (1-p)X_2 \quad ?$$

Problem V (45 points)

An agronomist believes that a newly developed plant food will increase the mean yield of tomato plants by more than 5 pounds. Thirteen (13) plants were treated with the food and had a mean yield of 28.4 pounds and a standard deviation of 2.8 pounds. Fifteen (15) identical plants were untreated and had a mean yield of 22.3 pounds and a standard deviation of 2.3 pounds.

a) (25 pts) Do the data provide sufficient evidence to support the agronomist's theory? Test at the 5% significance level assuming **unequal** population variances. Explain your results to the agronomist (who does not have too much knowledge of statistics).

b) (10 pts) Construct a 95% one-sided confidence interval for **lower bound** of the difference between mean yields of treated and untreated plants. How does this confidence interval support your conclusions from part a)?

c) (10 pts) Use an F-test for variances to test whether the yields of treated plants have the same variance as the yields of untreated plants.

[1]**Problem VI (50 points)**

A test consists of a series of short questions. A student has a probability p of correctly answering a question. A strong student has a higher p than a weak student. The correctness of answers to different questions is independent. All questions have equal level of difficulty, so that p for a given student is the same for all questions.

a) (10 pts) Jane is a good student for whom $p = 0.75$. Use normal approximation to find a probability that Jane scores **70% or lower** on a 100-question test.

b) (10 pts) Laura is a weaker student for whom $p = 0.6$. Use normal approximation to find a probability that Laura scores **70% or higher** on a 100-question test.

[1] This problem addresses normal approximation to binomial distribution. **You do not have to use continuity corrections** for questions a) - d).

c) (15 pts) Jane and Laura are taking the same test. The test has 100 questions and each question is worth one point. What is the probability that Jane scores at least 20 points more than Laura on this test? Assume that the number of correct answers can be well approximated by a normal distribution with mean $m = 100p$ and variance $s^2 = 100p(1-p)$.

d) (15 pts) What is the probability that Laura's score on the test described in c) is higher than Jane's score? Assume that the number of correct answers can be well approximated by a normal distribution with mean $m = 100p$ and variance $s^2 = 100p(1-p)$.

Problem VII (50 points)

A biologist has a culture consisting of 13 cells. Each of these cells divides into two cells independently of other cells. The time (in hours) until cell division of each cell has an exponential distribution with $\lambda = 0.5$.

a) (10 pts) What is the expected time (in hours) until cell division of a given cell?

b) (10 pts) What is the probability that a given cell divides within one hour?

c) (15 pts) Assuming that none of the cells in the culture has died and that each cell can divide into two cells only once within one hour, what is the expected number of cells in the culture after one hour?

d) (15 pts) Assuming that none of the cells in the culture has died and that each cell can divide into two cells only once within one hour, what is the probability that after one hour the biologist has no more than 16 cells in the culture?