Department of Mathematics
University of Toledo

Master of Science Degree
Comprehensive Examination
Applied Statistics

April 13, 2002

Instructions:

Do six problems out of the seven given
Show all of your computations in your blue book
Prove all of your assertions or quote appropriate theorems
Books, notes, and calculators may be used
Problems are equally weighted
This is a three hour test

Problem I

The quiz scores for the 9 students in Section 1 of the course "Introductory Statistics" have a mean of 35.22 and standard deviation of 4.94. The quiz scores for the 9 students in Section 2 have a mean of 31.56 and standard deviation of 4.48. Assuming normality of the quiz scores for both Section 1 and 2 students, and assuming equal variances, answer the following questions:

1. Obtain a point estimate of the difference in mean quiz scores.

2. Obtain a point estimate of the assumed common variance of quiz scores.

3. Obtain a 95% confidence interval for the difference in mean quiz scores.

4. Obtain the P-value for the test of the hypothesis that the mean quiz score for the Section 1 students is higher than that for the Section 2 students.

Problem II

A manufacturer wishes to compare the wearing qualities of two different types of automobile tires, A and B. For the comparison, one tire of Type A and one tire of type B are randomly assigned and mounted on the rear wheels of five automobiles. The automobiles are then operated for a specified number of miles, and the amount of wear is recorded for each tire. The data obtained was as follows:

Automobile	1	2	3	4	5
Tire A	10.6	9.8	12.3	9.7	8.8
Tire B	10.2	9.4	11.8	9.1	8.3

1. Obtain a point estimate of the difference in mean wear measurements.

2. Obtain an 80% confidence interval for the difference in mean wear measurements.

3. Obtain the P-value for the test of the null hypothesis that the mean wear measurements for the two types of tires are equal (using a two-sided alternative hypothesis).

Problem III

One explanation for the widespread incidence of the hereditary condition known as sickle-cell trait is that the trait confers some protection against malarial infection. In one investigation, a random sample of 543 African children were checked for the trait and for malaria. The results obtained were as follows:

	Heavy malarial infection	Light or no malarial infection	Total
Has sickle-cell trait	36	100	136
Does not have trait	152	255	407
Total	188	355	543

Let p_1 denote the incidence of heavy malarial infection among children with the sickle-cell trait, and let p_2 denote the incidence of heavy malarial infection among children without the sickle-cell trait.

1. Obtain a point estimate for $p_1 - p_2$.

2. Obtain a 90% confidence interval for $p_1 - p_2$.

3. Obtain the P-value for the test of H_0: $p_1 - p_2 = 0$ vs. H_1: $p_1 - p_2 \neq 0$, based on a "Z-test" that uses an estimated standard error computed under the assumption that the null hypothesis is true.

4. Obtain the P-value for the test of H_0: $p_1 - p_2 = 0$ vs. H_1: $p_1 - p_2 \neq 0$, based on a "Z-test" that uses an estimated standard error computed without making the assumption that the null hypothesis is true.

Problem IV

Consider again the data of Problem III.

1. Obtain the P-value for both the Pearson chi-square test and likelihood ratio chi-square test of the hypothesis that the variables "Sickle-cell trait status" and "Malarial status" are independent. Are either of these tests equivalent to either of the "Z-tests" conducted in Problem V?

2. Obtain an 85% confidence interval for the odds ratio.

3. Was the study design <u>cross-sectional</u>, <u>cohort</u>, or. <u>case-control</u>? (Circle one)

Problem V

Consider the following data set giving the final calculus grades and math achievement test scores for a random sample of 10 students:

Final calculus grade	65	78	52	82	92	89	73	98	56	75
Math achievement test score	39	43	21	64	57	47	28	75	34	52

We will consider the regression of "final calculus grade" on "math achievement test score", and will assume that all 3 assumptions of the simple linear regression model (linear regression function, constant standard deviation, and normality) are applicable.

The following descriptive statistics have been obtained:

$$\bar{X} = 46, \quad \bar{Y} = 76, \quad S_x = 16.57977, \quad S_y = 15.11438, \quad r = .83979$$

1. Obtain a point estimate of the regression slope coefficient.
2. Obtain a point estimate of the regression intercept coefficient.
3. Obtain a point estimate of the assumed constant standard deviation.
4. Obtain the P-value for a test of the null hypothesis that the regression slope coefficient is 0, against the two-sided alternative hypothesis.
5. Obtain a point estimate and 95% confidence interval for the mean final calculus grade for students with a math achievement test score of 50.

Problem VI

The miles obtained per gallon of gasoline for a particular make of tractor on a testing machine is known to be normally distributed with a mean of 10.5 mpg (miles per gallon) and standard deviation of 1.8 mpg. Company engineers have redesigned the carburetor in an effort to increase the mean mpg. A random sample of 9 tractors were equipped with the new carburetor and tested on the machine. The sample mean and sample standard deviation obtained were 12.0 mpg and 2.7 mpg, respectively. Assuming that the distribution of mpg is also normal for the new carburetor, answer the following questions:

1. Obtain the P-value for the test of the hypothesis that the standard deviation for the variable "mpg" is larger for the new carburetor than for the old.

2. Obtain the P-value for the test of the hypothesis that the mean "mpg" is larger for the new carburetor than for the old.

3. Obtain a 99% lower confidence bound for the mean "mpg" of the new carburetor.

Problem VII

Three methods of determining blood serum amylase concentrations in patients with pancreatitis are to be compared. The concentration measurements using all three methods on each of nine patients are given below.

1. What criteria should be satisfied for us to apply the methods of ANOVA to this comparison?
2. What nonparametric method would be most appropriate for testing whether or not the three methods yield data with distributions whose measures of center are equal?
3. Perform the test you suggest in #2 using level of significance $\alpha = .05$. You may use a large sample approximation if you wish.

	Method		
Specimen	A	B	C
1	4000	3210	6120
2	1600	1040	2410
3	1600	647	2210
4	1200	570	2060
5	840	445	1400
6	352	156	249
7	224	155	224
8	200	99	208
9	184	70	227