Department of Mathematics
University of Toledo

Master of Science Degree
Comprehensive Examination
Applied Statistics

May 31, 1997

Instructions:
Do all four problems.
They will be weighted equally -- 25 points each.
Show all of your computations.
Books, notes, and calculators *may be used.*
This is a three hour test.

1. Adult-onset diabetes is known to be highly genetically determined. A study was done comparing probabilities of a particular allele in a sample of such diabetics and a sample of nondiabetics. The data are shown in the following table:

	Diabetic	Normal
Bb or *bb*	12	4
BB	39	49

(a) Suppose we wish to test whether the probabilities of the alleles are significantly different in the diabetic and normal groups. Set up the null hypothesis H_0 and compute the maximum likelihood estimates of m_{ij} for $i, j = 1, 2$ when H_0 is true. Find the value of an appropriate test statistic and explain how to find the degrees of freedom for the test statistic. What is your conclusion?

(b) Find the sample odds ratio $\hat{\theta}$ and interpret the result. Does the sample odds ratio change its value when the orientation of the above table is reversed so that the rows become the columns and the columns become the rows? Explain your reasoning.

(c) Define the relative risk RR for 2×2 tables and find a sample version $\widehat{RR}$ of RR on the basis of the table above. Which of $\hat{\theta}$ and $\widehat{RR}$ is closer to 1? Explain your result.

2. The following are the weight gains (in pounds) of two random samples of young turkeys fed two different diets but otherwise kept under identical conditions:

Diet 1 (X)	16.3	10.1	10.7	13.5	14.9	11.8	14.3	10.2
	12.0	14.7	23.6	15.1	14.5	18.4	13.2	14.0
Diet 2 (Y)	21.3	23.8	15.4	19.6	12.0	13.9	18.8	19.2
	15.3	20.1	14.8	18.9	20.7	21.1	15.8	16.2

(a) Use normal theory to test the null hypothesis that the two populations sampled are identical against the alternative hypothesis that on the average the second diet produces a greater gain in weight. Let $\alpha = 0.01$. (You may the following facts: $\bar{X} = 14.20625$ $\bar{Y} = 17.93125$, $S_X^2 = \frac{1}{15}\sum_{i=1}^{16}(X_i - \bar{X})^2 = 11.28329$, $S_Y^2 = \frac{1}{15}\sum_{j=1}^{16}(Y_j - \bar{Y})^2 = 10.42629$.)

(b) Test the same hypotheses in part (a) using a nonparametric method. Let $\alpha = 0.01$. (You may use the normal approximation).

(c) Discuss the merits and faults of the two methods—that of part (a) and that of part (b).

3. The "PERU" data set in MINITAB is described on the attached, page 1.The data set itself is also given. In the following pages, I have developed a multiple regression model for the variable SYSTOL as a (linear) function of the previous 8 variables (not DIASTOL, which is another outcome variable). Answer the following questions regarding the analysis which was performed.

a. Based upon the correlation matrix alone, what would you do for an initial regression model? Be sure to consider the order that the variables are placed in the model. State why. Assume that we will *not* use any stepwise routines.

b. I chose to do the first regression shown, using all eight variables in the order given. Which variables does this regression indicate definltively (without question) should be *in*cluded in the final model? Why?

c. Whlch variables does this regression indicate definitively should be excluded in the final model? Why?

d. Use this regression output to test the null hypothesis that all variables except WEIGHT should be excluded from the model, i.e., that all of their parameters are zero. Use level of significance = .10. In this and all tests to follow, give the test statistic, its null disribution, either the P-value or critical value, and your decision.

e. Use this regression output to test the null hypothesis that WEIGHT should be excluded from the model given that *no other* variables are in the model. Use level of significance = .10.

f. Based upon this regression output, what would you use as the next model. State why.

g. I chose to do the next regression shown, on the four variables given. Use this regression output to test the null hypothesis that AGE and CHIN can be excluded from the model which contains WEIGHT and YEARS.

h. Based upon this regression output, what would you use as the next model. State why.

i. I chose to do the next regression analysis using only the two variables, WEIGHT and YEARS. Explain how it could be that YEARS is statistically significant (strongly, P=.004) while the correlation of YEARS with SYSTOL is so small (-.087).

j. Use this regression output to test the null hypothesis that YEARS should be excluded from the model given that WEIGHT is in the model. Use level of significance = .10.

k. Several diagnostic procedures are given for this regression. State which ones indicate a possible problem and remedial measures which could be used to overcome those problems.

l. State how you might have searched for the best multiple regression model differently. Do you think that your search would arrive at a different conclusion?

4. A repeated measures anaslysis at three time points and at four locations yielded the foiiowing data. These should be regarded as four independent multivariate observations on the three dimensional vector.

Observations:
Location 1 -- (1,3,6)'
Location 2 -- (3,5,7)'
Location 3 -- (2,5,3)'
Location 4 -- (2,3,4)'

a. Show that the sample mean vector is $\underset{\sim}{\bar{X}} = (2,4,5)'$ and the sample covariance matrix is

$$S = 1/3\begin{pmatrix} 2 & 2 & 1 \\ 2 & 4 & 0 \\ 1 & 0 & 10 \end{pmatrix}.$$

b. Assume that these observations arise from a muitivariate normal distribution with mean vector $\underset{\sim}{\mu}$. Test H_0: $\mu_1=\mu_2=\mu_3$ using Hotelling's T^2 test. Use a level of significance $\alpha=.05$.

c. Whether you reject H_0 or not, for the six contrasts: $\mu_1-(\mu_2+\mu_3)/2$, $\mu_2-(\mu_1+\mu_3)/2$, $\mu_3-(\mu_1+\mu_2)/2$, $\mu_1-\mu_2$, $\mu_1-\mu_3$, and $\mu_2-\mu_3$, find out whether Scheffe or Bonferroni intervals are shorter.

d. Evaluate the *first* of the intervals (not all six). Use the shorter type from part c.