Department of Mathematics
University of Toledo

Master of Science Degree
Comprehensive Examination
Applied Statistics -- In-Class Portion

April 17, 1999

<u>*Instructions:*</u>
Do all four problems.

Show all of your computations.
Books, notes, and calculators *may be used.*
This is a three hour test.

The examination committee tries to proofread the exams as carefully as possible. Nevertheless, the exam may contain misprints. If you are convinced that a problem has been stated incorrectly, mention this to the proctor and indicate your interpretation in your solution. In such cases, do not interpret the problem in such a way that it becomes trivial.

1. Along with this exam, you received four attachments. These all relate to a project I (Don White) did with Ken Bachmann and others in the College of Pharmacy. The topic is "allometry" where an attempt is made to relate the properties of a chemical as it acts in animals (here rats) to how it acts in humans. The data set (part of which is displayed) includes information on "volume of distribution" (VofD) and "half-life" (HL). VofD measures the apparent volume into which the chemical immediately distributes itself in the body and relates to the early concentrations of the chemical in the blood stream. HL measures how fast the chemical moves through the body and is eliminated. As the name implies, it is the time required for tbe concentration (usually in the blood stream) to reduce to half of the original value. Ultimately, in the project, we only considered HL.

The data set consists of 10 variables, as can be seen in the input statement in the SAS program provided. First is the chemical name. Next come the rat VofD and human VofD, each followed by a number indicating the reference where the information was found (Look at the number of references in the first -- 1996 -- paper!). Then come the rat HL and human HL, also with reference numbers. Finally, we report logP which measures how "lipophyllic" the compound is, i.e., how much it "sticks" to fat as opposed to water. This seemed important to consider (we considered it in the second, 1997, paper) since 1) humans have more fat than rats and 2) many environmental toxins with extremely long half-lives have high logP.

Our goal is to relate human half-life to rat half-life and logP. To that end, I have provided the SAS program already mentioned, its SASLOG, the (partial) data, the (partial) LISTING, and the two publications (check out the COLOR graphs). Please answer the following questions about this SAS analysis.

a) The first DATA step and the UNIVARIATE and PLOT procedures which follow it basically communicate why we transform to X1 and Y1. Discuss the results which lead us to use these transformations.

b) The REG procedures all have Y1 as the dependent variable. The last REG model uses X1 and X2 only. There are other reasonable models, but this is the one I selected for completing the analysis for this exam. In addition to what was done before running this final model, what other analyses would you have done to help find the best model? Explain why. Please use SAS code wherever possible to indicate what else you would have done.

c) Discuss the analysis of the residuals performed. Include any conclusions you can draw about the adequacy of the model. If anything is missing that you would have included, mention that as well.

d) Tell what the last portion of the program (from the second DATA step on down) accomplishes and how. Finally, tell why you think we would want this information (Notice Table 3 in the 1996 paper).

2. On each of ten rats, two incisions were made. With random assignments, one of the incisions was closed with tape and the other with sutures (thread). The strength of each closure was measured 40 days later. The measured strengths are:

Rat	1	2	3	4	5	6	7	8	9	10
Tape	659	984	397	574	447	479	676	761	647	577
Suture	452	587	460	787	351	277	234	516	577	513

Our goal is to compare the strengths of the two closure methods.

a) Tell what is the usual parametric test and the assumptions which must be satisfied for it to be valid. Use descriptive statistics only to indicate whether you believe these assumptions are satisfied.

b) Whether or not the assumptions are satisfied, use the parametric test to decide whether or not there is evidence for a difference between the two groups.

c) As a back-up method, to be sure that there is not a problem relative to the assumptions listed in part a, we often do nonparametric tests as well. Identify whatever nonparametric tests apply to this problem. State any assumptions that go with this (these) test(s).

d) Whatever tests you identified in part c, do them.

e) Based on parts a-d, draw a final conclusion about this problem.

f) Finally, use parametric methods to give a point and interval estimate of the difference between the two closure methods. Use a confidence level of 95%.

3. Assume that f is a density function symmetric about 0, that is, satisfies the condition $f(x) = f(-x)$ for all x. Let $X_1, \ldots, X_n$ be a random sample from a population with density function f, and suppose that we observe the values $Y_i = \theta + X_i$, $i = 1, \ldots, n$. Let the observed values of $n = 20$ observations of Y_i be 315, 493, 366, 291, 501, 503, 388, 526, 308, 410, 418, 540, 285, 360, 426, 475, 336, 455, 301, 359. Use the Wilcoxon signed rank test with $\alpha = 0.05$ to test the hypothesis that θ satisfies the inequality $\theta \leq 350$ against the alternative $\theta > 350$.

4. In a famous sociological study called *Middletown*, two teachers administered questionnaires to 784 high school students. The students were asked which two of ten given attributes were most desirable in their fathers. The following table shows how the desirability of the attribute "being a college graduate" was rated by male and female students

	Mentioned	Not Mentioned
Male	86	283
Female	55	360

(a) Compute the Pearson chi-squared statistic X^2 and the likelihood-ratio statistic G^2.

(b) Did the males and females value this attribute differently? Use $\alpha = 0.05$.

(c) Compute and interpret the sample odds ratio.

(d) Construct an approximate 95% confidence interval for the population odds ratio. Interpret the direction and strength of the association between gender and the desirability of this attribute.