Department of Mathematics
University of Toledo

Master of Science Degree
Comprehensive Examination
Applied Statistics

November 23, 1996

Instructions:

Choose 4 problems from the 5 given.
They will be weighted equally -- 25 points each.
You may attempt the fifth problem for up to 10 points extra credit. You select the fifth problem.
Show all of your computations.
Books, notes, and calculators *may be used.*
This is a three hour test.

1. Exploring City Populations

The United States conducts a census of its entire population once every ten years. The last such census was conducted in 1990. This data for this problem is only a very small portion of the data which is available (http://www.census.gov/cdrom/lookup).

The Census Bureau defines certain categories of "places" in a state including: cities, towns, boroughs, and the rather ubiquitous CDPs (Census Designated Places). This problem explores the demographics of entities which the Census Bureau has classified as cities in the state of Massachusetts. The aims of this exploration are primarily descriptive, not inferential, since the data does not come from a sample of cities in the state, but are measurements on all the cities in the state.

These questions below are primarily intended to be answered without the aid of a calculator. If you are using calculator more than to check a couple of numbers, then you are probably not doing the problem correctly. A straight-edge or ruler may be useful.

In 1990 the population of the state of Massachusetts was 6,016,425. There are 39 cities in Massachusetts. These cities and their 1990 populations are listed in the table below:

City	Population	City	Population
Attleboro	38383	Medford	57407
Beverly	38195	Melrose	28150
Boston	574283	New Bedford	99922
Brockton	92788	Newburyport	16351
Cambridge	95802	Newton	82585
Chelsea	28710	North Adams	16797
Chicopee	56632	Northampton	29242
Everett	35701	Peabody	47039
Fall River	92703	Pittsfield	48622
Fitchburg	41194	Quincy	84985
Gardner	20125	Revere	42786
Gloucester	28716	Salem	38091
Haverhill	51418	Somerville	76210
Holyoke	43704	Springfield	156983
Lawrence	70207	Taunton	49832
Leominster	38145	Waltham	57878
Lowell	103439	Westfield	38372
Lynn	81245	Woburn	35943
Malden	53844	Worcester	169759
Marlborough	31813		

a. Produce a stem-and-leaf plot for the city populations of Massachusetts.
b. Describe the distribution of city populations for Massachusetts.
c. Give a typical value for a city population in Massachusetts. Justify your choice.
d. Produce the five-number summary for city populations in Massachusetts.

2. Experimental Designs

a. Give a simple randomized block design for six treatments (A, B, C, D, E, and F) and four blocks (I, II, III, and IV) which is both orthogonal and balanced. You may choose the number of units per block, but indicate which treatment will be applied to each unit.

b. Does an orthogonal design exist if there are only three nnits per block? If not, why? If so, produce it.

c. Does a balanced design exist if there are only three units per block? If not, why? If so, produce it.

d. One possible design for the case when there are only three units per block is:

I	II	III	IV
A	A	B	C
B	D	D	E
C	E	F	F

In comparing treatment means, is the precision of the estimate of tlie A–D comparison greater than, less than, or the same as the precision of the estimate comparing B–F? Or is there enough evidence to decide? What about the precision of the estimate of C–D compared with the precision of the estimate of D–E? (Remember that precision is inversely related to variance.)

3. Below is a section of MINITAB output where some data was simulated and basic statistics were obtained.

```
MTB > Random 50 'U';
SUBC>   Uniform 0.0 1.0.
MTB > Let 'X' = 10*(1-sqrt(1-'U'))
MTB > Sort 'U' 'X' 'U' 'X';
SUBC>   By 'X'.
MTB > Describe 'X'.

               N     MEAN   MEDIAN   TRMEAN    STDEV   SEMEAN
X             50    2.910    2.087    2.721    2.517    0.356

              MIN      MAX       Q1       Q3
X           0.065    8.691    0.913    4.888

MTB > Print 'X'.
X
  0.06536   0.20875   0.26086   0.41053   0.43836   0.49264   0.60594
  0.61236   0.78932   0.86965   0.90008   0.90442   0.91543   0.99962
  1.03099   1.11052   1.21146   1.30302   1.36422   1.59804   1.84212
  1.88736   1.99084   2.01458   2.08260   2.09194   2.20438   2.23097
  2.29522   2.46273   2.61030   2.80950   2.84774   2.89225   3.04345
  3.39173   4.24951   4.85234   4.99683   5.19885   5.25416   5.32906
  5.85091   5.99004   7.35760   8.17465   8.20768   8.23257   8.33161
  8.69103
```

a. Identify the true distribution of X.

b. Perform a chi-square test of H_0: X comes from an exponential distribution. Use $\alpha = .05$.

c. Find a 90% confidence interval for the mean of the distribution of X. Whatever method you use, identify the important assumptions and defend your use of it in this situation.

4. Two methods, A and B, were used to measure the latent heat of fusion of ice in several different (independent) experiments. The outcomes are given below for m=13 experiments with method A and n=8 experiments with method B:

Method A		Method B
79.98	80.05	80.02
80.04	80.03	79.94
80.02	80.02	79.98
80.04	80.00	79.97
80.03	80.02	79.97
80.03		80.03
80.04		79.95
79.97		79.97

Basic Stats:	A	B
Mean	80.02	79.98
St Dev	.024	.031

a. Do a parametric test to determine whether or not these distributions are centered in the same location. Check into (and critically evaluate) the assumptions.

b. Find a 99% confidence interval for the difference between the method A and B means. Don't forget to comment on assumptions.

5 . Say that the following data is from 10 paired observations:

Group 1	Group 2
10	8
12	9
13	9
13	8
15	13
16	18
16	16
19	20
21	18
29	25

a. Perform a paired difference t-test to test the null hypothesis that the two group means are equal versus the alternative that the group 1 mean is larger. Find and state the P-value. Check the necessary assumptions.

b. Repeat part a with the most appropriate nonparametric test.

c. This type of problem can be solved in the multivariate context. Assume that the observation pairs are i.i.d. bivariate normal. In this context, define the appropriate parameters, state the hypotheses, write the test statistic and give its distribution (clue: it's Hotelling T-square). Do this test and comment on the similarities and differences between the two approaches (parts a and c; for example, compare the necessary assumptions) and the two answers.