Department of Mathematics
University of Toledo

Master of Science Degree
Comprehensive Examination
Applied Statistics

May 20, 1995

Instructions:

You will do a total of five problems.

Three should be chosen from the *Detailed Knowledge* Questions, numbered 1-4.

Two should be chosen from the *General Knowledge* Questions, numberd 5-7.

List the problems to be graded on the cover of your Blue Book.

Show all of your computations in your Blue Book.

Books, notes, and calculators *may be used.*

This is a three hour test.

DETAILED KNOWLEDGE PORTION

CHOOSE THREE OF FOUR

1. Heart Muscle Physiology

This problem deals with the physiology of heart muscle in rats. Laboratory experiments were conducted to examine the properties of heart muscle as a function of the amount of time the muscle was exposed to certain conditions. The relationship of muscle function to time was also examined at two different temperatures.

Two sets of experiments were conducted, one set (21 observations) at 22°C and the other set (8 observations) at 30°C. Each experiment consisted of placing a rat's heart into a solution, waiting for a specified number of minutes and then making a measurement of the fraction of a particular cellular substance that had been exchanged from solution. For the analysis, X is the waiting time in minutes, and Y is the log base 10 of the fraction exchanged. A regression of Y on X was performed for the 22°C experiments, and then another regression was performed for the 30°C experiments. Some minitab output for these analyses is provided below.

a. Give a 90% prediction interval for a new Y for the 30°C experiments and measured at 14 minutes.
b. There are theoretical reasons for believing that the slopes for the two regression lines would be different. Perform an $\alpha = .05$ test for the equality of slopes for these two models. Indicate clearly the assumptions you are making and any approximations that were used.
c. Exhibit a single linear model for all the data (all 29 experiments) which could be used to test the hypothesis indicated in part b above. For this model, what is the test statistic you would use and its sampling distribution.

```
MTB > # This is the analysis of the 22 degree C experiments
MTB > regress c5 1 c4
The regression equation is
C5 = -0.771 - 0.0507 C4
Predictor       Coef        Stdev     t-ratio
Constant    -0.77116      0.03591      -21.48
C4         -0.050675     0.003834      -13.22
s = 0.05348       R-sq = 90.2
MTB > # This is the analysis of the 30 degree C experiments
MTB > regress c7 1 c6
The regression equation is
C7 = -0.835 - 0.0812 C6
Predictor       Coef        Stdev     t-ratio
Constant    -0.83488      0.05302      -15.75
C6         -0.081190     0.006023      -13.48
s = 0.03903       R-sq = 96.8
MTB > mean c4
   MEAN   =  8.86
MTB > mean c5
   MEAN   =  -1.220
MTB > mean c6
   MEAN   =  8.50
MTB > mean c7
   MEAN   =  -1.525
MTB > stdev c4
   ST.DEV. =   194.6
MTB > stdev c6
   ST.DEV. =    42.0
```

2. A civil engineer has measured the deflection of beams made of three types of metals, labeled A,B, and C. The deflections are measured in 1/1000 of an inch and the data, with some useful statistics, are listed below:

Type	Observations	Means	Standard Deviations
A	82 86 79 83 85 84 86 87	84	2.62
B	74 82 78 75 76 77	77	2.83
C	79 79 77 78 82 79	79	1.67

Our task is to decide if there is sufficient information to conclude that the three metals produce a different response, and, also, if they are different, to find out which one (or ones) yields the smallest deflection.

a. Draw a simple plot which displays the three data distributions as compared to each other.

b. The plot shows some possible differences between the metals. In order to find out if the differences are statistically significant, perform the ANOVA by finding first the sum of squares and mean squares for treatment (metal) and error. Test at level of significance $\alpha = .03$. Assume that the three variances are equal.

c. If there is a statistically significant difference in part b, we might wish to find simultaneous confidence intervals for differences between pairs of means. Again assuming equality of variances, tell how you would find 97% simultaneous confidence intervals for these differences. Then, whether or not you find a statistically significant difference in part b, find the interval for metals B & C *ONLY*. State clearly how you are taking into account the need for simultaneous intervals. For A-B and A-C, I found the appropriate intervals to be (2.5,11.5) and (1.3,8.7), respectively.

d. Based on the test in b and the intervals in c, what is your conclusion -- are there any differences, and if so, which metals are better than which other metals?

3. An Experimental Design

For the model $y_{ijk} = \mu + \alpha_i + \gamma_j + \delta_k + \epsilon_{ijk}$, where $i = 1,2,3$, $j = 1,2,3$, $k = 1,2,3$ (not all combinations of indices occur) and the ϵ_{ijk} are independent and identically distributed $\mathcal{N}(0, \sigma^2)$, a particular design matrix, $\mathbf{X}$, and parameterization vector, $\boldsymbol{\beta}$, are given below.

$$\mathbf{X} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 \end{pmatrix}$$

$$\boldsymbol{\beta} = \begin{pmatrix} \mu + \alpha_1 + \gamma_1 + \delta_1 \\ \alpha_2 - \alpha_1 \\ \alpha_3 - \alpha_1 \\ \gamma_2 - \gamma_1 \\ \gamma_3 - \gamma_1 \\ \delta_2 - \delta_1 \\ \delta_3 - \delta_1 \end{pmatrix}$$

a. Indicate which of the pairs of factors are orthogonal.

b. If factor A, associated with the α parameters, and factor G, associated with the γ parameters, are both considered blocking factors, then what is this design called (give the name for the particular design).

c. Produce the design matrix, $\mathbf{X}^*$, corresponding to the parameterization,

$$\boldsymbol{\beta}^* = \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \gamma_1 \\ \gamma_2 \\ \delta_1 \\ \delta_2 \end{pmatrix}$$

d. For the design matrix, $\mathbf{X}$, provided above, the $(\mathbf{X}'\mathbf{X})^{-1}$ matrix is given below. Using that information, give the variance for the three contrasts; $\delta_2 - \delta_1, \delta_3 - \delta_1$, and $\delta_3 - \delta_2$.

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{pmatrix} 0.778 & -0.333 & -0.333 & -0.333 & -0.333 & -0.333 & -0.333 \\ -0.333 & 0.667 & 0.333 & 0.000 & 0.000 & 0.000 & 0.000 \\ -0.333 & 0.333 & 0.667 & 0.000 & 0.000 & 0.000 & 0.000 \\ -0.333 & 0.000 & 0.000 & 0.667 & 0.333 & 0.000 & 0.000 \\ -0.333 & 0.000 & 0.000 & 0.333 & 0.667 & 0.000 & 0.000 \\ -0.333 & 0.000 & 0.000 & 0.000 & 0.000 & 0.667 & 0.333 \\ -0.333 & 0.000 & 0.000 & 0.000 & 0.000 & 0.333 & 0.667 \end{pmatrix}$$

4. As in problem 2, a civil engineer has measured the deflection of beams made of three types of metals, labeled A,B, and C. The deflections are measured in 1/1000 of an inch and the data, with some useful statistics, are listed below:

Type	Observations	Means	Standard Deviations
A	82 86 79 83 85 84 86 87	84	2.62
B	74 82 78 75 76 77	77	2.83
C	79 79 77 78 82 79	79	1.67

a. Perform the nonparametric analog to the test done in #2, i.e., the Kruskal-Wallis test, to see if there are any differences between the three groups.

b. We could also ask whether or not the three distributions come from the same underlying distribution using the Wilcoxon test for equality of two distributions on each pair of treatments. Do the Wilcoxon test for B versus C *ONLY*. The other two pairs (A & B and A & C are significantly different). Make these tests into one simultaneous test by taking into account the fact that there are three pairs of treatments. Do this by properly adjusting the individual α for your test so that your overall α will be at most .06 (i.e., do a Bonferroni correction for the tests).

GENERAL KNOWLEDGE PORTION

CHOOSE TWO OF THREE

5. Following is data from a sequence of three baseline observations on six animals.

Obs # -->	1	2	3	Animal #
	1.4	1.4	2.0	1
	3.0	3.0	3.0	2
	4.8	4.9	8.4	3
	10.2	11.0	11.7	4
	5.4	6.7	6.8	5
	6.8	7.9	8.4	6

a. Test the hypothesis that the observations are stable versus the alternative that the means are changing. Use $\alpha = .10$. Hint: In a multivariate context, the null hypothesis of no difference, $H_0: \mu_1=\mu_2=\mu_3$ is equivalent to $H_0: \mu_2-\mu_1=0$ and $\mu_3-\mu_2=0$. Also, to do this problem, you will need to invert a 2 by 2 symmetric matrix. Recall that if the matrix is not singular, then

$$\begin{pmatrix} a & b \\ b & c \end{pmatrix}^{-1} = \frac{1}{ac-b^2}\begin{pmatrix} c & -b \\ -b & a \end{pmatrix}$$

b. If a difference was found in part a, it would be of interest to find a set of simultaneous confidence intervals for the differences between pairs of the three means. Describe how you would find these intervals if they were jointly at least 85% intervals. Then find the interval to compare the 3rd and 1st baselines *ONLY.*

6. As in problem 2, a civil engineer has measured the deflection of beams made of three types of metals, labeled A,B, and C. The deflections are measured in 1/1000 of an inch and the data, with some useful statistics, are listed below:

Type	Observations	Means	Standard Deviations
A	82 86 79 83 85 84 86 87	84	2.62
B	74 82 78 75 76 77	77	2.83
C	79 79 77 78 82 79	79	1.67

We will now use the residuals to determine whether or not it is safe to assume that the error terms are drawn from populations with the same variances and whether or not the populations are normal.

a. Find the residuals for this data, i.e., the observations minus the estimated treatment means. Then draw some plot which displays the distribution of the residuals. Does this data look relatively normal or not?

b. Next for each of the three pairs of residual observations, we could do the Seigel-Tukey test to address the question of whether or not the "spreads" of the three groups are equal. Do this test for comparing metals A and C *ONLY.*

c. Repeat part b (A and C only) using the F-test for variances. What assumption(s) is(are) required for this test to be valid?

d. For residuals, the appropriate mean is, of course, zero. Using all of the residuals, estimate the variance and then do a chi-square goodness of fit test to see whether or not the residuals fit a normal distribution with mean zero. Use $\alpha = .05$. List any possible drawbacks to this test which might adversely affect the results. What is your conclusion?

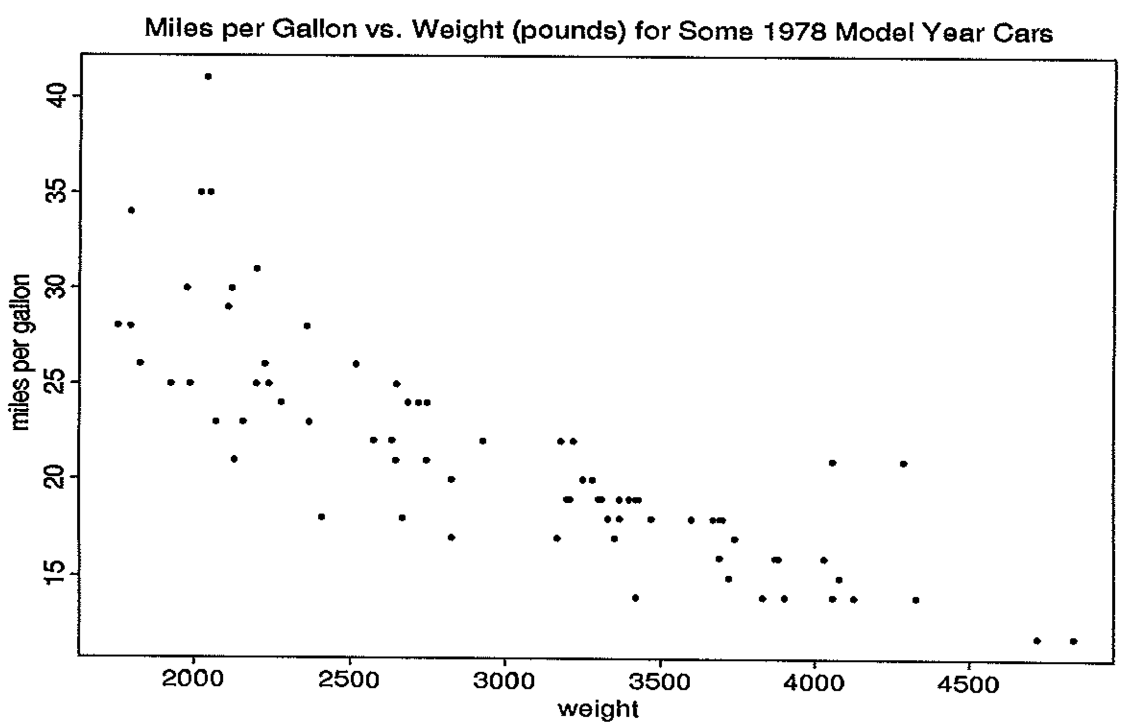
7. 1978 Automobiles

a. Below are the sticker prices for 74 cars from 1978. Produce a stem-and-leaf plot for this data. Then briefly describe the shape of the distribution.

Sticker Prices ($) for Various Models

4099	4749	3799	9690	6295
9735	4816	7827	5788	4453
5189	10372	4082	11385	14500
15906	3299	5705	4504	5104
3667	3955	6229	4589	5079
8129	3984	5010	5886	6342
4296	4389	4187	5799	4499
11497	13594	13466	3995	3829
5379	6303	6165	4516	3291
8814	4733	5172	5890	4181
4195	10371	12990	4647	4425
4482	6486	4060	5798	4934
5222	4723	4424	4172	3895
3798	5899	3748	5719	4697
5397	6850	7140	11995	

b. Below is a plot of miles per gallon versus weight in pounds for 74 cars from 1978. This plot is reproduced on the following page. On the **full page** plot, add a median trace and the hinge traces. Briefly describe the relationship between the two variables. [Be sure to turn in your plot with your blue books.]



Miles per Gallon vs. Weight (pounds) for Some 1978 Model Year Cars