METHODS QUALIFYING EXAM

January 2009

INSTRUCTIONS:

a.) DO NOT put your NAME on the exam. Place the NUMBER assigned to you on the UPPER LEFT HAND CORNER of EACH PAGE of your exam.

b.) Please start your answer to EACH QUESTION on a SEPARATE sheet of paper.

c.) Answer all the questions.

d.) Be sure to attempt all parts of every question. It may be possible to answer a later part of a question without having solved the earlier parts.

e.) Be sure to hand in all of your exam. No additional material will be accepted once the exam has ended and you have left the exam room.

f.) You may use only a calculator, pencil or pen, and blank paper for this examination. No other materials are allowed.

Problem I.

Seven mice are subjected to a treatment and researchers are interested in the survival in days of these mice. Here are the survival times in days of a random sample of 7 mice.

16, 23, 38, 94, 99, 141, 197

a.) Give a 95 percent confidence interval (CI) for mean survival of all mice subjected to the treatment using the assumption that $t = \sqrt{n}(\overline{x} - \mu)/s$ has a t-distribution. [the 95th percentile of the t-distribution with 6 d.f. is 1.943, the 97.5th percentile is 2.447].

b.) The researchers feel the interval in a) is too imprecise, approximately how many mice would they need in a new experiment to get an interval of width=30 days?

c.) The researchers suspect that a) is not appropriate. It now assumed that the distribution of survival times is exponential, that is $f(x) = \lambda exp(-\lambda x)$ for some λ. Explain how to simulate to find the distribution of t. In particular, explain precisely how we find t_α, the αth percentile of the distribution of t for any $\alpha \in (0, 1)$?

d.) If from this simulation, $t_{.025} = -3.75$, $t_{.05} = -2.05$, $t_{.95} = 1.45$ and $t_{.975} = 1.75$. What is the appropriate confidence interval for mean survival?

Problem II.

An entomologist was studying the effect of two insecticides (labelled insecticides A and B) on four species of ants (labelled species 1, 2, 3 and 4). The entomologist anticipated that approximately 50% of each species would survive twenty-four hours' exposure to either of the insecticides.

A total of 2000 ants were obtained from each of the four species. For species 1, the ants were randomly divided into groups of 200. Each group was placed in a separate flask, resulting in a total of ten flasks filled with 200 ants each. Five of the ten flasks were randomly selected and then exposed to insecticide A. The other five flasks were exposed to insecticide B.

The same general procedure also was applied independently to each of species 2, 3 and 4.

After twenty-four hours of exposure to the selected pesticide, the entomologist examined each flask and recorded

$$y_{ijk} = \text{Number of ants (out of 200) that are still alive after}$$
$$\text{24 hours from species } i\text{, insecticide } j\text{, flask } k$$

a.) Our entomologist provides you with the 40 values y_{ijk}, $i = 1, 2, 3, 4$; $j = 1, 2$; $k = 1, 2, 3, 4, 5$. Write down a model for y_{ijk} associated with the above description of the experiment.

b.) Display an ANOVA table for this experiment, showing sources of variation and degrees of freedom. Indicate for each source whether the source is a fixed or random effect. Find the expected mean square for each of source of variation in your ANOVA table. Indicate the appropriate denominator of the F statistic for each relevant test.

c.) Give a brief interpretation of the *interaction* term in the model in part b). Be sure to include: i) an algebraic explanation in terms of cell means; and ii) a graphical illustration of two cases involving "zero interaction" and "nonzero interaction," respectively.

d.) Describe briefly two important diagnostic checks you would want to carry out before reporting analysis results from a) and b).

e.) Now our entomologist wants to carry out a similar study involving the same two insecticides. The design is identical to the design described above, except that instead of having four species of ants, the entomologist will use four species of termites. Based on previous studies, the entomologist anticipates that about 0.1% of the termites will survive 24 hours' exposure to either insecticide, while the other 99.9% of the termites will not. Given this new information about this new experiment, would you consider it appropriate to use the methods in a) and b) to analyze the new termite data? Explain why or why not.

Problem III.

A study was carried out on how general daily stress level (low or high) affects one's opinions (favorable or unfavorable) of a proposed new health policy. Interviews were made of random samples of subjects from both rural and urban areas where each was classified according to stress level and opinion of the proposed new health policy. Use the accompanying SAS output to help you answer this question.

a.) Was the residence area associated with either of the two variables of primary interest, daily stress level and opinion on the proposed new health policy? Explain your reasoning.

b.) Carry out a test of independence of the opinion on health policy with stress level, ignoring the area. If there is association, describe the nature of the association.

c.) For each area (rural and urban), carry out a test of independence of the opinion on health policy with stress level. If there is association, describe the nature of the association.

d.) Describe the difference in the results for b.) and c.). Explain how they can differ.

e.) Describe how a similar phenomenom can occur in regression analysis where one fits a model relating two continuous variables for two treatment groups. Give a concrete example.

```
The FREQ Procedure
Table of stress by opinion

stress     opinion

Frequency|favor   |unfavor |  Total
---------+--------+--------+
low      |    103 |    147 |    250
---------+--------+--------+
high     |    133 |    147 |    280
---------+--------+--------+
Total         236      294      530
            44.53    55.47   100.00


Statistics for Table of stress by opinion
Statistic                    DF       Value      Prob
------------------------------------------------------
Chi-Square                    1      2.1222    0.1452


           Estimates of the Relative Risk (Row1/Row2)
Type of Study                   Value       95% Confidence Limits
-----------------------------------------------------------------
Case-Control (Odds Ratio)      0.7744        0.5489        1.0926
Sample Size = 530
--------------------------------------------------------------------------
```

```
Table of residence by stress

residence     stress

Frequency|low     |high    |  Total
---------+--------+--------+
urban    |     60 |    220 |    280
---------+--------+--------+
rural    |    190 |     60 |    250
---------+--------+--------+
Total         250      280      530

Statistics for Table of residence by stress
Statistic                    DF       Value      Prob
------------------------------------------------------
Chi-Square                    1    157.8362    <.0001
Sample Size = 530
---------------------------------------------------------------------------
```

```
Table of residence by opinion


residence     opinion

Frequency|favor   |unfavor |  Total
---------+--------+--------+
urban    |    174 |    106 |    280
---------+--------+--------+
rural    |     62 |    188 |    250
---------+--------+--------+
Total         236      294      530

Statistics for Table of residence by opinion
Statistic                    DF       Value      Prob
------------------------------------------------------
Chi-Square                    1     74.5641    <.0001
Sample Size = 530
```

```
Table 1 of stress by opinion
Controlling for residence=urban

stress     opinion                          Statistics for Table 1 of stress by opinion

                                               Controlling for residence=urban
Frequency|favor   |unfavor |  Total
---------+--------+--------+                   Statistic      DF       Value      Prob
low      |     48 |     12 |     60            ---------------------------------------
---------+--------+--------+                   Chi-Square     1      10.3507    0.0013
high     |    126 |     94 |    220
---------+--------+--------+
Total         174      106      280

           Estimates of the Relative Risk (Row1/Row2)

Type of Study                   Value       95% Confidence Limits
-----------------------------------------------------------------
Case-Control (Odds Ratio)      2.9841        1.5018        5.9297

Sample Size = 280
```

```
Table 2 of stress by opinion
Controlling for residence=rural

stress     opinion                           Statistics for Table 2 of stress by opinion

                                                Controlling for residence=rural
Frequency|favor   |unfavor |  Total
---------+--------+--------+                    Statistic       DF       Value      Prob
low      |     55 |    135 |    190             ---------------------------------------
---------+--------+--------+                    Chi-Square      1       7.3016    0.0069
high     |      7 |     53 |     60
---------+--------+--------+
Total          62      188      250

           Estimates of the Relative Risk (Row1/Row2)

Type of Study                   Value       95% Confidence Limits
-----------------------------------------------------------------
Case-Control (Odds Ratio)      3.0847        1.3207        7.2045

Sample Size = 250
```

```
Summary Statistics for stress by opinion
Controlling for residence

  Cochran-Mantel-Haenszel Statistics (Based on Table Scores)
Statistic    Alternative Hypothesis    DF       Value      Prob
---------------------------------------------------------------
    1        Nonzero Correlation        1     17.5785    <.0001

            Estimates of the Common Relative Risk (Row1/Row2)
Type of Study     Method                    Value     95% Confidence Limits
-------------------------------------------------------------------------
Case-Control      Mantel-Haenszel          3.0255        1.7738        5.1607
  (Odds Ratio)    Logit                    3.0235        1.7731        5.1559

Total Sample Size = 530
```

Problem IV.

This a Simple Linear Regression problem with several parts.

1.) Assume that the model is valid with respect to the mean function, that is,

$$E(Y) = \beta_o + \beta_1 * X$$

2.) The errors may not be iid.

3.) There are three parts to this question; with each part standing on its own merits.

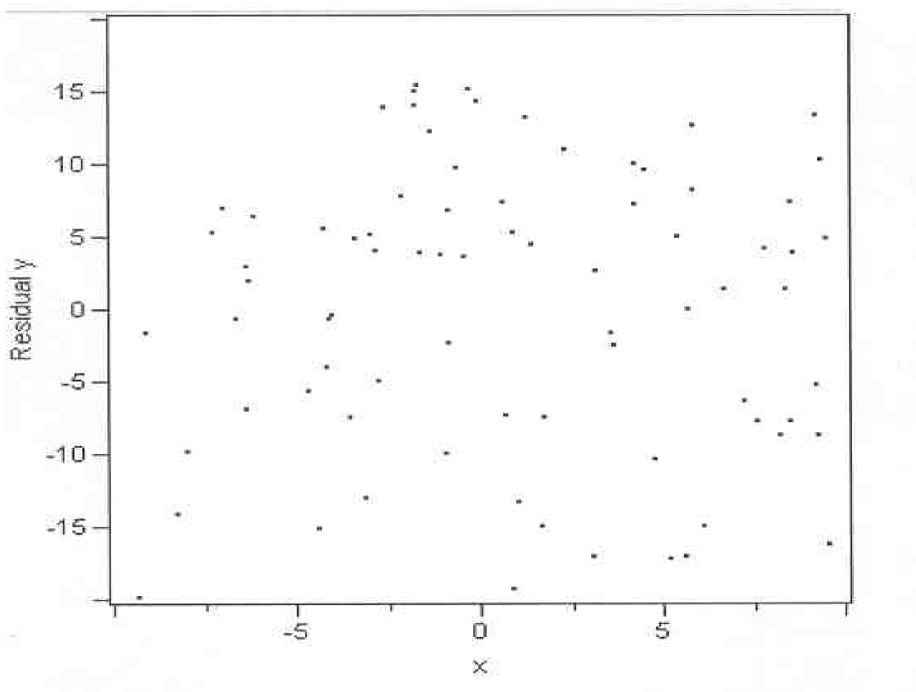
4.) Explain your answers. No long answers are needed.

Part 1.

a.) Given Graph 1 below, is there any reason to suggest that the errors are not iid?

b.) Can you tell from Graph 1 if the residuals are normally distributed?

c.) Are there any remedial measures you need to take based on Graph 1. If yes, please indicate what they are.



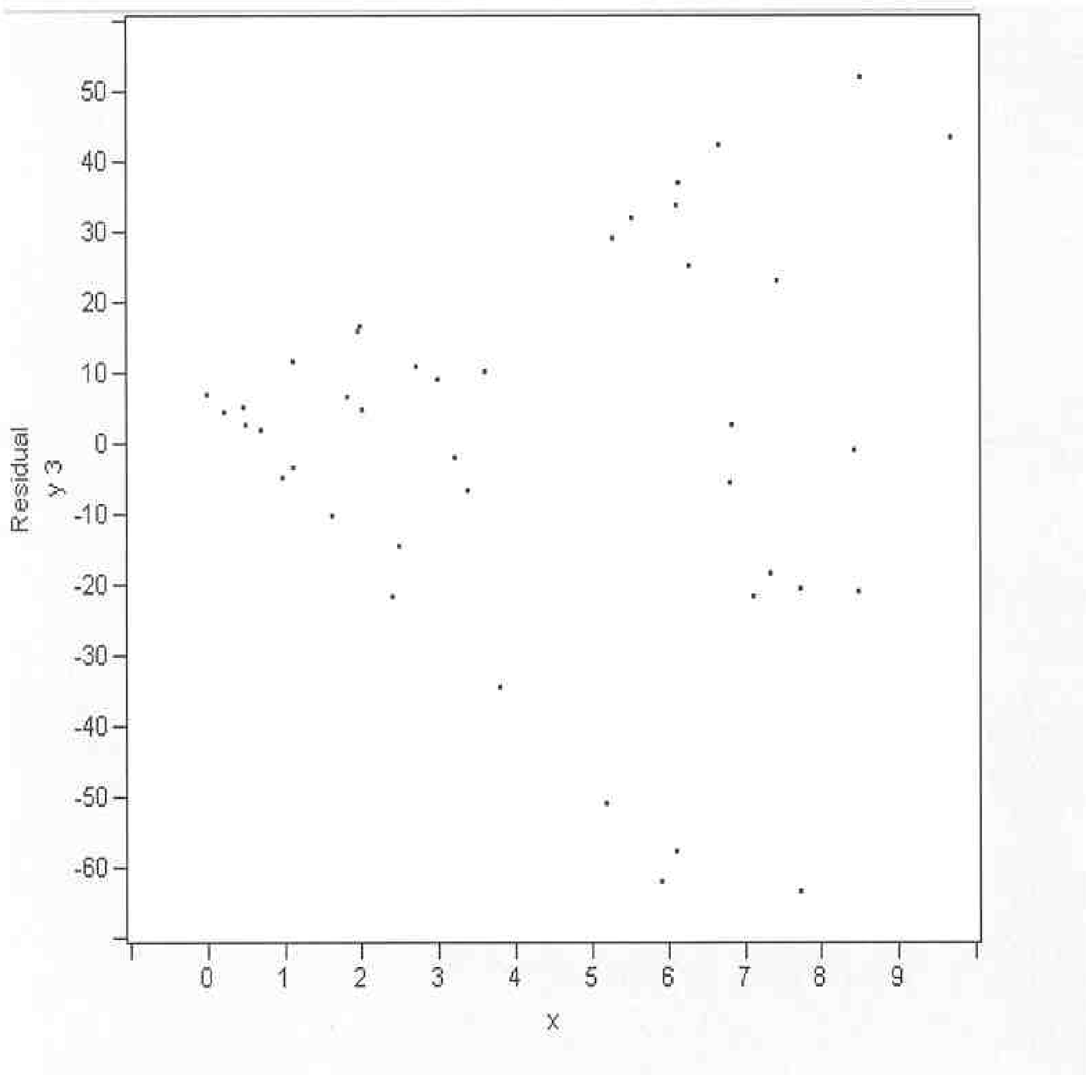
GRAPH 1

Part 2.

a.) Given Graph 2 below, is there any reason to suggest that the errors are not iid?

b.) Can you tell from Graph 2 if the residuals are normally distributed?

c.) Are there any remedial measures you need to take based on Graph 2. If yes, please indicate what they are.



GRAPH 2

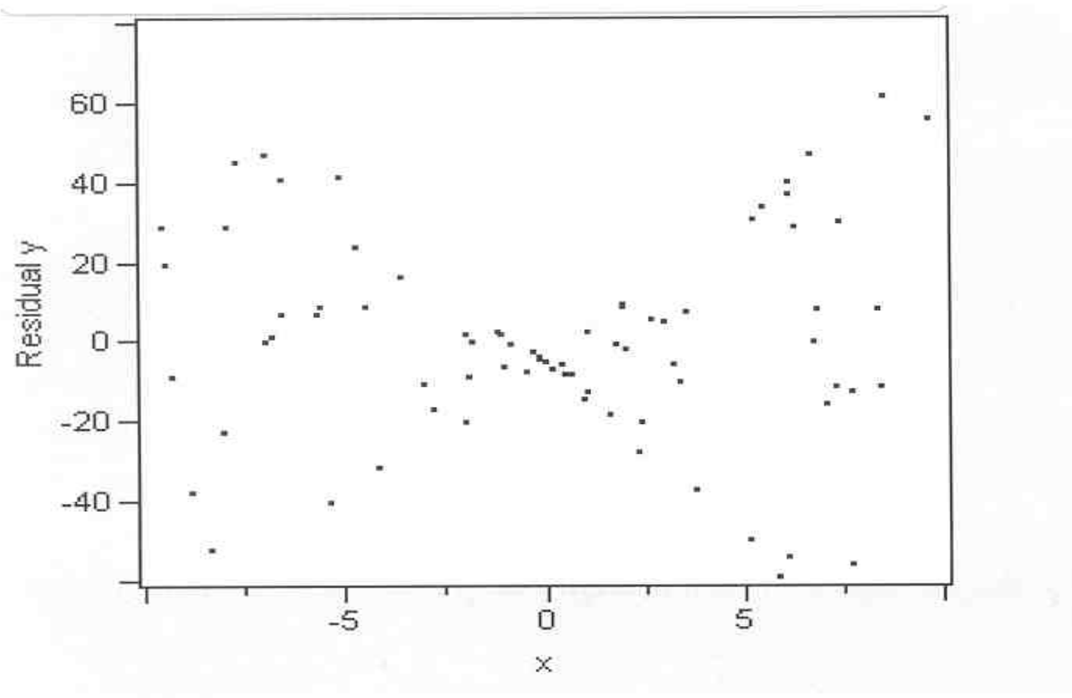
Part 3.

a.) Given Graph 3 below, is there any reason to suggest that the errors are not iid?

b.) Can you tell from Graph 3 if the residuals are normally distributed?

c.) Are there any remedial measures you need to take based on Graph 3. If yes, please indicate what they are.

d.) Given the data in Dataset #3 below, what is the null hypothesis for testing homogeneity of variances? Give $H_o : \cdots$



GRAPH 3

Dataset #3

Id	Y	X
1	-48.906	-9.488
2	-57.390	-9.407
3	-84.613	-9.245
4	-109.74	-8.815
5	-120.01	-8.324
6	-87.559	-7.987
7	-34.971	-7.879
8	-17.096	-7.686

9	-9.396	-6.986
10	-56.142	-6.984
11	-53.162	-6.783
12	-11.369	-6.518
13	-45.517	-6.515
14	-37.812	-5.660
15	-35.617	-5.592
16	-81.905	-5.324
17	1.668	-5.073
18	-12.897	-4.677
19	-25.544	-4.435
20	-62.664	-4.102
21	-10.673	-3.578
22	-32.863	-3.008
23	-36.823	-2.750
24	-32.754	-1.936
25	-10.763	-1.933
26	-21.630	-1.886
27	-12.070	-1.822
28	-11.733	-1.776
29	-3.675	-1.167
30	-3.673	-1.105
31	-11.111	-1.021
32	-4.039	-0.832
33	-7.244	-0.408
34	-1.422	-0.318
35	-1.345	-0.148
36	-1.781	-0.141
37	-1.188	0.007
38	-1.365	0.234
39	1.921	0.471
40	-0.024	0.498
41	1.211	0.692
42	-2.098	0.984
43	0.740	1.114
44	15.724	1.122
45	-0.533	1.626
46	18.468	1.807
47	29.165	1.951
48	30.305	1.988
49	18.895	2.016
50	-3.365	2.398
51	4.689	2.486
52	32.524	2.712
53	34.096	3.005
54	25.316	3.221
55	22.761	3.396
56	41.879	3.610
57	-0.731	3.801
58	-1.605	5.198
59	79.410	5.276
60	84.994	5.539
61	-4.279	5.929
62	1.950	6.119
63	93.363	6.122
64	96.671	6.133
65	86.779	6.301
66	108.040	6.680
67	62.021	6.823
68	70.248	6.840
69	49.236	7.135
70	55.136	7.355
71	97.180	7.429
72	57.155	7.733
73	14.233	7.739
74	84.284	8.431
75	65.244	8.499
76	138.130	8.519
77	142.630	9.690