Advanced Biostatistics **First Exam** **Name ______________________________________**
Spring Semester 2005 **25 or 28 Total Points** **13 April 2005**

Directions: Neatly and thoroughly answer each of the following six problems in the space given, clearly showing all work. Unless otherwise noted, please use $\alpha = 5\%$. Keep four decimal places in your calculations.

1. (2 + 2 + 2 = 6 points) Pawitan (2001:154) presents data related to 40 individuals who underwent an experimental surgery, which are analyzed on p.1 of the *Appendix*. The dataset includes the age of the patient at the time of the surgery and whether (Y = 1) or not (Y = 0) the patient died within the first 30 days immediately following the surgery.

   (a) Clearly interpret the "Point Estimate" of the odds ratio.

   (b) Clearly give the ramifications/implications of the 95% (Wald) confidence interval for the odds ratio.

   (b) Give the predicted probability of survival (not dying) in the first 30 days post surgery for a 65-year-old.

2. (1 + 2.5 + 1 = 4.5 points) In a genetic study of an insect population which damages crops, a random sample of four male insects were each mated with two (different) females chosen at random. The weights (in g) of three eggs produced by each of the eight mated females were measured. The data for this nested design are analyzed in Minitab on the bottom of p.1 of the *Appendix*.

   (a) Briefly indicate why this is a nested design.

   (b) Test whether there is a difference between the Males in terms of the average egg weights.

   Null hypothesis ____

   Alternative hypothesis ____

   Test stat, df and p-value ____

   Detailed conclusion

(c) Give the point estimate of the Males variance component, s_M^2.

3. (2.5 + 2.5 = 5 points) Finney (1978:48) presents data related to the growth response of *Lactobacillus leichmannii* to one of eight different doses of vitamin B_{12} each replicated in 6 independent test tubes. These data are analyzed on pp.2-3 of the *Appendix*, first using PROC NLIN and the resulting residual plot, and then using two PROC NLMixed's.

(a) Write out the assumed model function in terms of x = dose used in the NLIN and identify the roles of the parameters.

Model Function ____

Role of θ_1 ("th1") ____

Role of θ_2 ("th2") ____

Role of θ_3 ("th3") ____

(b) After viewing the residual plot, the two NLMixed's were then run. Using the corresponding output, perform a test of homoskedasticity clearly writing out your hypotheses, the calculated test statistic, degrees of freedom (df), p-value, and your detailed conclusion.

Null hypothesis ____

Alternative hypothesis ____

Test stat, df and p-value ____

Clear conclusion

4. (1 + 1 + 2.5 = 4.5 points) A study was conducted involving 8 subjects randomized to an experimental treatment and 8 subjects randomized to a control treatment, and in which respiration (liters of air/min/m^2 of body area) was measured both at rest and after 30 minutes of work. The data are plotted and analyzed on pp. 4-5 of the *Appendix*. Two separate lines (one for each treatment) are fit in the PROC REG at the top of p.5.

(a) Write down the fitted line for the experimental group

____

(b) Write down the fitted line for the control group

____

(c) Using a ***= 1% here***, test whether ***one line can be fit to both the experimental and control groups***.

Null hypothesis ____

Alternative hypothesis ____

Test stat, df and p-value ____

Clear conclusion

5. (2.5 + 2.5 = 5 points) Schwenke & Milliken (1991) report that in meat science research, a pH level of 6.0 in the steer muscle is desired before processing begins, that at the time of slaughter pH is usually in the range of 7.0-7.2, and that pH decreases over time. In general, however, it is not practical to monitor the pH decline to determine precisely when pH of 6.0 is attained, and they conducted a study to determine the time when pH 6.0 is reached for two treatments, the ***conventional treatment*** (denoted '**conv**') and an ***electrically simulated hot boned treatment*** (denoted '**eshb**'). The study data corresponding to the postmortem pH for the *longissimus dorsi* muscle of 24 steer carcasses, 12 steers randomized to each of the treatments and measured only at 1, 2, 4, 6, 8 or 24 hours postmortem, are graphed on the bottom of p.5 of the *Appendix*. The suggested model for these data is y = $\eta(x,\theta) + \varepsilon$, where x = time postmortem, and

$$h(x,q) = q_2 + (q_1 - q_2)\exp\{-q_3 x\} \qquad (1.1)$$

Note that θ_1 is then the value of the expected pH at time x = 0 and θ_2 is the lower asymptote (at infinity).

(a) Using the appropriate outputs from the first three NLIN runs on pp. 6-7 of the *Appendix*, test whether (only) the lower asymptotes are the same for the two treatments, giving your hypotheses, test statistic, p-value and conclusion.

Null hypothesis ____

Alternative hypothesis ____

Test stat, df and p-value ____

Clear conclusion

(b) Using the appropriate outputs from the first three NLIN runs on pp. 6-7 of the *Appendix*, test whether **both** the lower and upper asymptotes are the same for the two treatments, giving your hypotheses, test stat., p-value and conclusion.

Null hypothesis ____

Alternative hypothesis ____

Test stat, df and p-value ____

Clear conclusion

The rest of the exam for G students only!!

(c) (1.5 points) Write out the model function being fit in the fourth NLIN program and output (p.8) in the *Appendix*, and indicate the role of the new (ϕ = phi) parameter.

(d) (1.5 points) For each treatment, how many hours postmortem should the meat scientists wait for processing in order that the pH be nearest to 6.0? Give approximate 95% confidence intervals.

For Conventional treatment ____

For ESHB treatment ____

First Problem –SAS Program and Output ***AB Exam 1 Appendix***

```
data onea;
  y=0; input age @@; datalines;
50 50 51 51 53 54 54 54 55 55 56 56 56 57 57 58 60 61 62 63 63 64 65 68
69 71
;
data oneb;
  y=1; input age @@; datalines;
57 57 59 61 61 62 62 62 63 64 67 67 68 70
;
data one;
  set onea oneb;
proc logistic descending;
  model y=age;
run;
```

```
                          The LOGISTIC Procedure
                             Response Profile
                     Ordered                      Total
                       Value            y     Frequency
                           1            1            14
                           2            0            26
                      Probability modeled is y=1.
                          Model Fit Statistics
                                                Intercept
                                   Intercept          and
                   Criterion            Only   Covariates
                   AIC                53.796       49.301
                   SC                 55.485       52.679
                   -2 Log L           51.796       45.301

                 Testing Global Null Hypothesis: BETA=0
          Test                 Chi-Square       DF     Pr > ChiSq
          Likelihood Ratio         6.4949        1         0.0108
          Score                    6.1508        1         0.0131
          Wald                     5.3899        1         0.0203

                Analysis of Maximum Likelihood Estimates
                                     Standard          Wald
    Parameter    DF     Estimate        Error    Chi-Square     Pr > ChiSq
    Intercept     1     -10.4814       4.3042        5.9301         0.0149
    age           1       0.1629       0.0702        5.3899         0.0203

                          Odds Ratio Estimates
                              Point          95% Wald
                Effect     Estimate       Confidence Limits
                age           1.177        1.026        1.351
```

Second Problem – Minitab Output

Minitab Output

Nested ANOVA: y versus male, female

Analysis of Variance for y

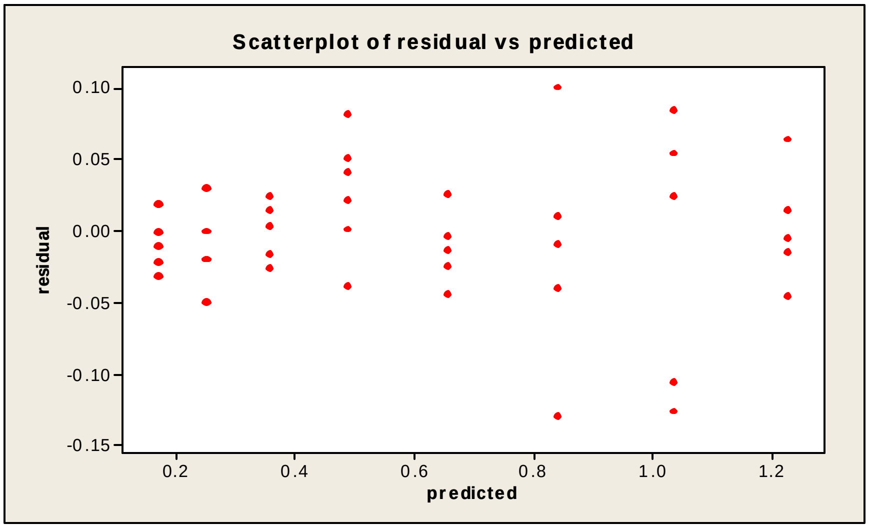
Source	DF	SS	MS	EMS
male	3	16.8913	5.6304	$s^2 + 3s_F^2 + 6s_M^2$
female	4	10.1450	2.5362	$s^2 + 3s_F^2$
Error	16	16.5933	1.0371	s^2
Total	23	43.6296		

Third Problem – SAS Program and Output

The Original NLIN Run

```
data one;
  do dose=0.23,0.35,0.53,0.79,1.19,1.78,2.67,4;
  do rep=1 to 6;
    input y @@; y=y/100; drop rep; output;
  end; end; datalines;
15 14 19 19 17 16 28 20 23 25 23 23 36 36 34 37 33 38 51 53 54 45 57 49
68 63 64 61 65 68 85 80 71 85 94 83 106 91 109 93 109 112 121 122 129 124 118 124
;
proc nlin data=one;
  parms th1=2 th2=2 th3=2;
  t=(dose/th2)**th3; den=1+t; mean=th1*t/den;
  model y=mean;
run;
```

Residual Plot



The First NLMixed

```
proc nlmixed data=one;
  parms th1=2 th2=1.5 th3=2 se=0.1 rho=1;
  t=(dose/th2)**th3; den=1+t; mean=th1*t/den; var=se*se*mean**rho;
  model y~normal(mean,var);
run;
```

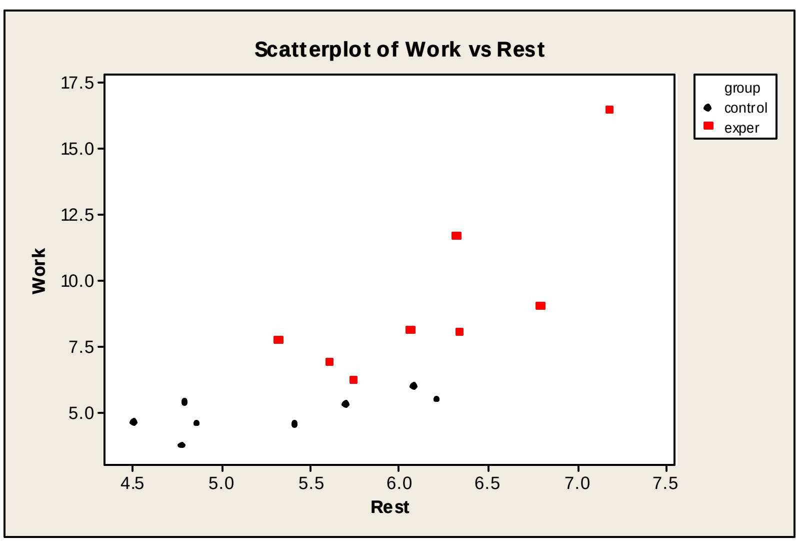
```
                                 The NLMIXED Procedure

                                        Dimensions
                             Observations Used                    48
                             Observations Not Used                 0
                             Total Observations                   48
                             Parameters                            5

                           NOTE: GCONV convergence criterion satisfied.

                                      Fit Statistics
                             -2 Log Likelihood                -174.8
                             AIC (smaller is better)          -164.8
                             AICC (smaller is better)         -163.3
                             BIC (smaller is better)          -155.4


                                   Parameter Estimates
                      Standard
  Parameter  Estimate    Error    DF  t Value  Pr > |t|   Alpha     Lower     Upper
Gradient
  th1          1.8021   0.1480    48    12.18    <.0001    0.05    1.5045    2.0996  -
0.00142
  th2          1.9995   0.3263    48     6.13    <.0001    0.05    1.3435    2.6556
0.000703
  th3          1.0597  0.05145    48    20.60    <.0001    0.05    0.9562    1.1631
0.000553
  se          0.06153 0.008886    48     6.92    <.0001    0.05   0.04366   0.07939
0.001657
  rho          1.3608   0.3076    48     4.42    <.0001    0.05    0.7424    1.9792  -
0.00005
```

The Second NLMixed

```
proc nlmixed data=one;
  parms th1=2 th2=1.5 th3=2 se=0.1; rho=0;
  t=(dose/th2)**th3; den=1+t; mean=th1*t/den; var=se*se*mean**rho;
  model y~normal(mean,var);
run;
```

```
                                   The NLMIXED Procedure

                                         Dimensions
                             Observations Used                    48
                             Observations Not Used                 0
                             Total Observations                   48
                             Parameters                            4

                           NOTE: GCONV convergence criterion satisfied.

                                       Fit Statistics
                             -2 Log Likelihood                -159.2
                             AIC (smaller is better)          -151.2
                             AICC (smaller is better)         -150.3
                             BIC (smaller is better)          -143.8

                                    Parameter Estimates
                          Standard
  Parameter   Estimate       Error    DF  t Value  Pr > |t|   Alpha      Lower      Upper
Gradient
  th1           1.9203      0.1854    48    10.36    <.0001    0.05     1.5474     2.2931   -
0.00041
  th2           2.2838      0.4627    48     4.94    <.0001    0.05     1.3535     3.2142
0.000232
  th3           1.0131     0.06874    48    14.74    <.0001    0.05     0.8749     1.1513
0.00032
  se           0.04606    0.004701    48     9.80    <.0001    0.05    0.03661    0.05552   -
0.01292
```

Fourth Problem – Graph, SAS Program and Output



```
data one;
  do group='experim','control';
    egroup=(group='experim');
  do rep=1 to 8;
    input rest work @@; egrest=egroup*rest; drop rep; output;
  end; end; datalines;
5.74 6.24 6.79 9.07 5.32 7.77 7.18 16.46 5.60 6.95 6.06 8.14 6.32 11.72 6.34 8.06
6.21 5.50 4.50 4.64 4.86 4.61 4.78 3.78 4.79 5.41 5.70 5.32 5.41 4.54 6.08 5.98
;
proc print; run;
```

Obs	group	egroup	rest	work	egrest
1	experim	1	5.74	6.24	5.74
2	experim	1	6.79	9.07	6.79
3	experim	1	5.32	7.77	5.32
4	experim	1	7.18	16.46	7.18
5	experim	1	5.60	6.95	5.60
6	experim	1	6.06	8.14	6.06
7	experim	1	6.32	11.72	6.32
8	experim	1	6.34	8.06	6.34
9	control	0	6.21	5.50	0.00
10	control	0	4.50	4.64	0.00
11	control	0	4.86	4.61	0.00
12	control	0	4.78	3.78	0.00
13	control	0	4.79	5.41	0.00
14	control	0	5.70	5.32	0.00
15	control	0	5.41	4.54	0.00
16	control	0	6.08	5.98	0.00

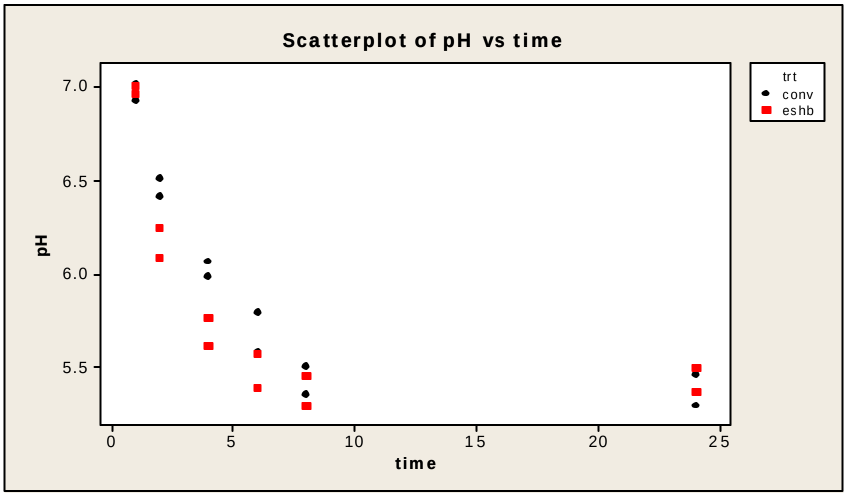
```
proc reg; model work=rest egroup egrest; run;
```

```
                         The REG Procedure
                     Dependent Variable: work

                         Analysis of Variance
                                 Sum of           Mean
      Source                DF        Squares         Square    F Value
Pr > F
      Model                  3      124.04476       41.34825      15.68
0.0002
      Error                 12       31.64918        2.63743
      Corrected Total       15      155.69394

                 Root MSE               1.62402    R-Square     0.7967
                 Dependent Mean         7.13688    Adj R-Sq     0.7459
                 Coeff Var             22.75530

                         Parameter Estimates
                              Parameter       Standard
         Variable      DF      Estimate          Error    t Value    Pr > |t|
         Intercept      1       1.09182        5.01742       0.22      0.8314
         rest           1       0.73342        0.94202       0.78      0.4513
         egroup         1     -17.66140        7.91656      -2.23      0.0455
         egrest         1       3.46044        1.36533       2.53      0.0262
```

```
proc reg; model work=rest; run;
```

```
                         The REG Procedure
                     Dependent Variable: work

                         Analysis of Variance
                                 Sum of           Mean
      Source                DF        Squares         Square    F Value
Pr > F
      Model                  1       94.09904       94.09904      21.39
0.0004
      Error                 14       61.59491        4.39964
      Corrected Total       15      155.69394

                 Root MSE               2.09753    R-Square     0.6044
                 Dependent Mean         7.13688    Adj R-Sq     0.5761
                 Coeff Var             29.39005

                         Parameter Estimates
                              Parameter       Standard
         Variable      DF      Estimate          Error    t Value    Pr > |t|
         Intercept      1     -11.65126        4.09626      -2.84      0.0130
         rest           1       3.27891        0.70900       4.62      0.0004
```

Fifth Problem – Graph, SAS Program and Output



The First NLIN

```
data one;
  do time=1,2,4,6,8,24;
  do trt='conv','eshb';
    trtc=(trt='conv'); trte=(trt='eshb');
  do rep=1,2;
    input ph @@; drop rep; output;
  end; end; end; datalines;
7.02 6.93 7.01 6.96 6.42 6.51 6.25 6.09 6.07 5.99 5.77 5.62
5.59 5.80 5.57 5.39 5.51 5.36 5.46 5.30 5.30 5.47 5.37 5.50
;
proc nlin data=one;
  parms th1c=7 th1e=7 th2c=5 th2e=5 th3c=0.5 th3e=0.5;
  th1=th1c*trtc+th1e*trte; th2=th2c*trtc+th2e*trte; th3=th3c*trtc+th3e*trte;
  model ph=th2+(th1-th2)*exp(-th3*time);
run;
```

```
                          The NLIN Procedure
                        Dependent Variable ph
        NOTE: Convergence criterion met.

                                        Sum of        Mean               Approx
    Source                      DF     Squares      Square    F Value    Pr > F
    Model                        5      7.8707      1.5741     163.09    <.0001
    Error                       18      0.1737     0.00965
    Corrected Total             23      8.0445

                                           Approx
        Parameter      Estimate       Std Error    Approximate 95% Confidence Limits
        th1c             7.5714          0.1394       7.2785       7.8643
        th1e             8.4361          0.3304       7.7421       9.1302
        th2c             5.3530          0.0635       5.2196       5.4865
        th2e             5.4196          0.0453       5.3244       5.5148
        th3c             0.3229          0.0408       0.2371       0.4087
        th3e             0.6658          0.0916       0.4734       0.8581
```

The Second NLIN

```
proc nlin data=one;
  parms th1c=7 th1e=7 th2=5 th3c=0.5 th3e=0.5;
  th1=th1c*trtc+th1e*trte; th3=th3c*trtc+th3e*trte;
  model ph=th2+(th1-th2)*exp(-th3*time);
run;
```

```
                          The NLIN Procedure
                        Dependent Variable ph
        NOTE: Convergence criterion met.

                                        Sum of        Mean               Approx
    Source                      DF     Squares      Square    F Value    Pr > F
    Model                        4      7.8636      1.9659     206.51    <.0001
    Error                       19      0.1809     0.00952
    Corrected Total             23      8.0445

                                           Approx
        Parameter      Estimate       Std Error    Approximate 95% Confidence Limits
        th1c             7.6072          0.1386       7.3171       7.8974
        th1e             8.3537          0.2958       7.7345       8.9729
        th2              5.3950          0.0369       5.3178       5.4723
        th3c             0.3432          0.0359       0.2680       0.4184
        th3e             0.6339          0.0794       0.4678       0.8000
```

The Third NLIN

```
proc nlin data=one;
  parms th1=7 th2=5 th3c=0.5 th3e=0.5;
  th3=th3c*trtc+th3e*trte;
  model ph=th2+(th1-th2)*exp(-th3*time);
run;
```

```
                                       The NLIN Procedure
                                      Dependent Variable ph
                 NOTE: Convergence criterion met.

                                                 Sum of          Mean                      Approx
           Source                       DF      Squares        Square      F Value      Pr > F
           Model                         3       7.7979        2.5993       210.87      <.0001
           Error                        20       0.2465        0.0123
           Corrected Total              23       8.0445

                                                    Approx
                  Parameter       Estimate      Std Error     Approximate 95% Confidence Limits
                  th1               7.7938         0.1478        7.4856       8.1021
                  th2               5.3806         0.0442        5.2884       5.4729
                  th3c              0.3728         0.0415        0.2862       0.4594
                  th3e              0.5019         0.0560        0.3851       0.6188
```

The Fourth NLIN

```
proc nlin data=one;
  parms th2=5 th3c=0.5 th3e=0.5 phic=3 phie=3;
  th3=th3c*trtc+th3e*trte; phi=phic*trtc+phie*trte;
  model ph=th2+(6-th2)*exp(-th3*(time-phi));
run;
```

```
                                       The NLIN Procedure
                                      Dependent Variable ph
                 NOTE: Convergence criterion met.

                                                 Sum of          Mean                      Approx
           Source                       DF      Squares        Square      F Value      Pr > F
           Model                         4       7.8636        1.9659       206.51      <.0001
           Error                        19       0.1809       0.00952
           Corrected Total              23       8.0445

                                                    Approx
                  Parameter       Estimate      Std Error     Approximate 95% Confidence Limits
                  th2               5.3950         0.0369        5.3178       5.4723
                  th3c              0.3432         0.0359        0.2680       0.4184
                  th3e              0.6339         0.0794        0.4678       0.8000
                  phic              3.7779         0.2141        3.3298       4.2260
                  phie              2.5039         0.1474        2.1954       2.8125
```