Advanced Biostatistics **Second Exam** **Name** ___________________________________
April 19, 2006

Directions: ***Thoroughly, clearly, neatly and correctly*** answer the following four exercises in the space given, showing all relevant calculations. Use α = 5% throughout unless otherwise noted. **<u>Total points = 25 (UG) 27 (G)</u>**.

1. (1 + 2.5 + 2.5 = 6 points) Agronomists are interested in using the following nonlinear model function to predict expected girth of rubber trees (Y) as a function of the rate of fertilizer application (X) for five rubber trees:

$$\eta = \delta + \frac{\gamma - \delta}{1 + (x/\phi)}$$

(γ = gamma, δ = delta, ϕ = phi). These data are analyzed using two Proc NLIN's in SAS in the *Appendix*.

(a) For the SAS analysis given in the *Appendix*, is normality (implicitly) assumed? _______________

(b) Using the best test available, test whether or not we can accept whether that the true lower asymptote (the value at x = 0) is equal to 20 and that the true upper asymptote equals 24. Write your null and alternative hypotheses in terms of the symbols γ, δ and ϕ. Show your calculations.

Null ___________________________________ Alternative ______________________________________

Calculated test statistic ___________________________ df _________

p-value _________________________________

Give your detailed and clear conclusion

(c) Using the first NLIN, test whether the true LD_{50} exceeds unity (one), showing your calculations. Note that this is a one-tailed test. Write your null and alternative hypotheses in terms of the symbols γ, δ and ϕ.

Null ___________________________________ Alternative ______________________________________

Calculated test statistic ___________________________ df _________

p-value _________________________________

Give your detailed and clear conclusion

2. (2 + 1 + 2 = 5 points) The counts in the following table come from a study of mental health for a random sample of 40 adult residents of Alachua Count, Florida. The measured variables are the degree of mental impairment (well, mild symptom formation, moderate symptom formation, impaired) and a life events index, which is a composite measure of the number and severity of important life events such as birth of a child, new job, divorce, or death in family that occurred to the subject within the past 3 years. These data are analyzed using SAS with the corresponding program/output given in the *Appendix* (page 2), where we wish to predict mental impairment (denoted **MI**) based on life events (denoted **LE**).

	Mental Impairment			
Life Events	Well (1)	Mild (2)	Moderate (3)	Severe (4)
Low (1)	**6**	**3**	**1**	**1**
Moderate (4)	**4**	**6**	**4**	**3**
High (7.5)	**2**	**3**	**2**	**5**

(a) (***Undergraduate Students only***) Identify the model that is being used here and list the necessary assumptions. Assess whether any of the assumptions have been met if this can be determined (give corresponding p-value).

Model name ____

Model function or equation(s) ____

Assumptions ____

Assessment of key assumption ____

(b) (***All students***) For the SAS analysis given in the *Appendix*, is normality (implicitly) assumed? __________

(c) (***All students***) Interpret the odds ratio in the context of this exercise. Be clear and specific.

(d) (***Graduate Students only***) Predict the mental impairment ***counts*** for individuals with Moderate LE. Show all work, and keep 2 decimal places in your final answer.

3. (8 points) Martin (1942) and Finney (1952:69) present data that can be used to test the relative potency of and interaction between **rotenone** and **duguelin**. In the study, various concentrations of these insecticides were sprayed on *Macrosiphoniella sanborni*, the chrysanthemum aphis, in batches of about fifty and the number dead was noted. These data are analyzed in the attached *Appendix*, which fits eight Proc NLMixed's. Note that "drug" is a proxy for insecticide, the levels of drug are "dd" for duguelin and "rr" for rotenone, and logistic regression is used in the program where π corresponds to the percentage dead and where the independent variable is the concentration of the respective insecticide. The SAS program/output labeled "for Exercise 3AB" is useful to assess relative potency and parallelism, and the programs/outputs labeled "for Exercise 3C" and "for Exercise 3D" can be used to assess interaction between these insecticides.

(a) Test for ***parallelism*** of the two dose-response curves giving your hypotheses (in terms of the model parameters), test statistic and its distribution (including degrees of freedom), and your specific conclusion.

Null ________________________________ Alternative ___________________________________

Calculated test statistic _______________________ df ________ p-value ___________________

Give your detailed and clear conclusion

(b) Test whether the insecticides are ***equally potent***, giving your hypotheses, test statistic (with distribution and degrees of freedom), p-value, and your clear conclusion. Use the best test available.

Null ________________________________ Alternative ___________________________________

Calculated test statistic _______________________ df ________ p-value ___________________

Give your detailed and clear conclusion

(c) Using the "Exercise 3C" program/output, test whether these insecticides exhibit significant antagonism/synergism, again giving your hypotheses, test statistic (with distribution and degrees of freedom), p-value, and your clear conclusion. Use the best test available.

Null ________________________________ Alternative ___________________________________

Calculated test statistic _______________________ df ________ p-value ___________________

Give your detailed and clear conclusion

(d) Does the model fit in program/output "for Exercise 3D" indicate these compounds exhibit significant antagonism/synergy? Support your claim, using the best test available.

Null ________________________________ Alternative ___________________________________

Calculated test statistic _______________________ df ________ p-value ___________________

Give your detailed and clear conclusion

(e) (***<u>Graduate students only</u>*** – 2 additional points) Give detailed reasons why the Separate Ray model is preferred to model these data instead of the Finney model. (Give your answer below and/or over.)

4. (1 + 1.5 + 1.5 + 2 = 6 points) A study was conducted involving 45 randomly chosen individuals in which it was noted whether the individual had an accident in the last year (denoted '**accident**' below), **age** (in years), vision status (denoted '**vision**', 0 = no problems, 1 = some problems) and whether or not the individual took a course in drivers education (denoted '**drive_ed**', 0 = no, 1 = yes). Our goal is to predict whether or not the individual has had an accident using logistic regression using some or all of the three explanatory variables, and the data are analyzed in SAS in the *Appendix*. Note also that in the SAS program, a dummy variable has been created called '**agegroup**' which is equal to unity (1) for younger (under 20) and older (over 65) drivers, and which is equal to zero for middle-aged drivers.

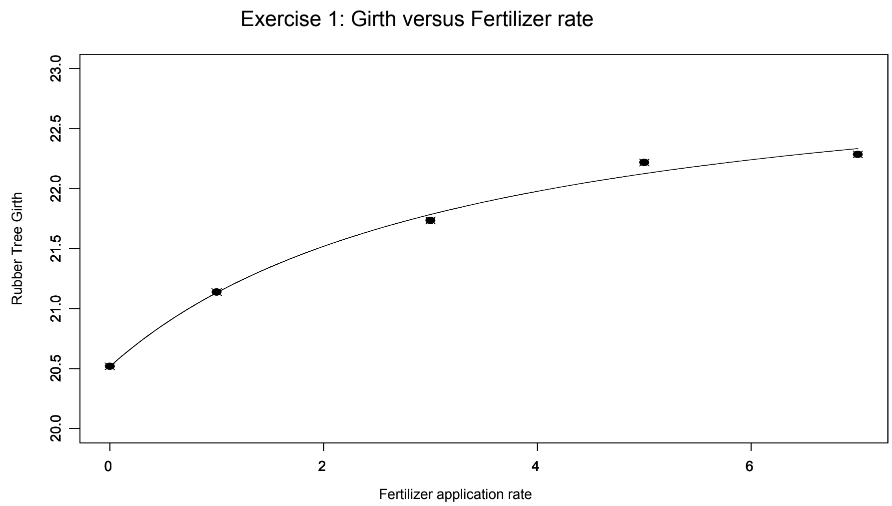
(a) For the SAS analysis given in the *Appendix*, is normality (implicitly) assumed? ________________

(b) Using the output, calculate and clearly interpret the point estimate for the odds ratio for **agegroup**.

(c) Using the output, calculate and clearly interpret the point estimate for the odds ratio for **drive_ed**.

(d) Using the output, obtain and clearly interpret the 90% confidence interval for the odds ratio for **vision**. Also, point out the ***ramifications*** of this confidence interval.

Graph, SAS Program and Output for Exercise 1 -




```
data one;
  do fert=0,1,3,5,7;
    input girth @@; output; end; datalines;
20.518 21.138 21.734 22.218 22.286
;
proc nlin;
  parms gamma=20 delta=23 phi=2;
  mean=delta+(gamma-delta)/(1+fert/phi);
  model girth=mean;    run;
proc nlin;
  parms phi=2; gamma=21; delta=24;
  mean=delta+(gamma-delta)/(1+fert/phi);
  model girth=mean;    run;
```

First NLIN output

Source	DF	Sum of Squares	Mean Square	F Value	Approx Pr > F
Model	2	2.2388	1.1194	165.19	0.0060
Error	2	0.0136	0.00678		
Corrected Total	4	2.2524			

Parameter	Estimate	Approx Std Error	Approximate 95% Confidence Limits	
gamma	20.5170	0.0802	20.1720	20.8619
delta	23.2117	0.3256	21.8109	24.6124
phi	3.3775	1.0413	-1.1028	7.8578

Second NLIN output

Source	DF	Sum of Squares	Mean Square	F Value	Approx Pr > F
Model	1	2330.2	2330.2	33236.8	<.0001
Error	4	0.2804	0.0701		
Uncorrected Total	5	2330.5			

Parameter	Estimate	Approx Std Error	Approximate 95% Confidence Limits	
phi	8.8221	1.9565	3.3900	14.2541

SAS Program and Output for Exercise 2 -

```
data two;
  do LE=1,4,7.5;
  do MI=1,2,3,4;
    input count @@; output;
  end; end; datalines;
6 3 1 1 4 6 4 3 2 3 2 5
;
proc logistic;
  title 'Mental Impairment ex. from Agresti p.279';
  weight count;
  model MI=LE;
run;
```

```
         Mental Impairment ex. from Agresti p.279

                The LOGISTIC Procedure
    Data Set                      WORK.TWO
    Response Variable             MI
    Number of Response Levels     4
    Number of Observations        12
    Weight Variable               count
    Sum of Weights                40
    Model                         cumulative logit

                      Response Profile
    Ordered                    Total            Total
      Value           MI     Frequency           Weight
          1            1             3        12.000000
          2            2             3        12.000000
          3            3             3         7.000000
          4            4             3         9.000000

    Score Test for the Proportional Odds Assumption
      Chi-Square       DF     Pr > ChiSq
          0.1368        2         0.9339

                 Testing Global Null Hypothesis: BETA=0
         Test                     Chi-Square       DF     Pr > ChiSq
         Likelihood Ratio             5.9827        1         0.0144
         Score                        5.4463        1         0.0196
         Wald                         5.8180        1         0.0159

                Analysis of Maximum Likelihood Estimates
                                  Standard          Wald
 Parameter       DF     Estimate       Error     Chi-Square     Pr > ChiSq
 Intercept 1      1       0.3115      0.5897         0.2790         0.5974
 Intercept 2      1       1.7105      0.6514         6.8953         0.0086
 Intercept 3      1       2.6315      0.7234        13.2339         0.0003
 LE               1      -0.3001      0.1244         5.8180         0.0159

                           Odds Ratio Estimates
                             Point          95% Wald
                Effect    Estimate       Confidence Limits
                LE           0.741        0.580        0.945
```

SAS Program for Exercise 3AB -

```
data one;
  input rr dd num dead @@; rr=rr/10; dd=dd/10; ratio=dead/num;
  datalines;
102   0 50 44 77   0 49 42 51   0 46 24 38   0 48 16 26   0 50  6
  0 505 48 48   0 404 50 47   0 303 49 47   0 202 48 34   0 101 48 18
 51 203 50 48 40 163 46 43 30 122 48 38 20  81 46 27 10  41 46 22 5 20 47 7
;
data onerr; set one;
  if dd=0; dose=rr; drug='rr'; drop rr dd;
data onedd; set one;
  if rr=0; dose=dd; drug='dd'; drop rr dd;
data two; set onerr onedd;
  dumr=(drug='rr'); dumd=(drug='dd');
proc nlmixed data=two;
  parms th2r=10 th2d=10 th3r=2 th3d=2;
  th2=th2r*dumr+th2d*dumd; th3=th3r*dumr+th3d*dumd;
  t=(dose/th2)**th3; p=t/(1+t);
  model dead~binomial(num,p);run;
proc nlmixed data=two;
  parms th2r=10 th2d=10 th3=2;
  th2=th2r*dumr+th2d*dumd;
  t=(dose/th2)**th3; p=t/(1+t);
  model dead~binomial(num,p);run;
proc nlmixed data=two;
  parms th2r=10 rho=1 th3=2;
  th2d=rho*th2r; th2=th2r*dumr+th2d*dumd;
  t=(dose/th2)**th3; p=t/(1+t);
  model dead~binomial(num,p);run;
```

SAS Output for Exercise 3AB -

Exercise 3 First NLMixed

-2 Log Likelihood = 40.2

Parameter Estimates

Parameter	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
th2r	4.8289	0.2496	10	19.35	<.0001	0.05	4.2728	5.3850
th2d	12.7509	1.0567	10	12.07	<.0001	0.05	10.3963	15.1054
th3r	3.1035	0.3877	10	8.00	<.0001	0.05	2.2397	3.9674
th3d	2.7784	0.3780	10	7.35	<.0001	0.05	1.9362	3.6206

Exercise 3 Second NLMixed

-2 Log Likelihood = 40.5

Parameter Estimates

Parameter	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
th2r	4.8203	0.2581	10	18.67	<.0001	0.05	4.2452	5.3954
th2d	12.9701	0.9515	10	13.63	<.0001	0.05	10.8500	15.0902
th3	2.9427	0.2712	10	10.85	<.0001	0.05	2.3384	3.5471

Exercise 3 Third NLMixed

-2 Log Likelihood = 40.5

Parameter Estimates

Parameter	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
th2r	4.8203	0.2581	10	18.67	<.0001	0.05	4.2452	5.3954
rho	2.6907	0.2419	10	11.13	<.0001	0.05	2.1518	3.2296
th3	2.9427	0.2712	10	10.85	<.0001	0.05	2.3384	3.5471

SAS Program for Exercise 3C -

```
data three; set one;
  eps=0.000000000000000000000000000000000000001;
  if rr=0 then rr=eps; if dd=0 then dd=eps; dose=rr+dd;
  ray=2*(_n_ le 5)+1*(_n_ ge 6 and _n_ le 10)+3*(_n_ ge 11);
proc nlmixed data=three;
  title 'Quiz4 Exercise B Fourth NLMixed';
  parms th2=20 th3=2 th4=1 th5=0;
  z=dd+th4*rr+th5*sqrt(th4*rr*dd);
  t=(z/th2)**th3; den=1+t; p=t/den;
  model dead~binomial(num,p);
run;
proc nlmixed data=three;
  title 'Quiz4 Exercise B Fifth NLMixed';
  parms th2=20 th3=2 th4=1;
  z=dd+th4*rr;
  t=(z/th2)**th3; den=1+t; p=t/den;
  model dead~binomial(num,p);
run;
```

SAS Output for Exercise 3C -

Exercise 3 Fourth NLMixed

-2 Log Likelihood = 79.0

Parameter Estimates

Parameter	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
th2	11.9659	1.0285	16	11.63	<.0001	0.05	9.7855	14.1462
th3	2.3352	0.1742	16	13.40	<.0001	0.05	1.9659	2.7045
th4	2.5083	0.2649	16	9.47	<.0001	0.05	1.9468	3.0699
th5	0.7411	0.2548	16	2.91	0.0103	0.05	0.2009	1.2812

Exercise 3 Fifth NLMixed

-2 Log Likelihood = 89.6

Parameter Estimates

Parameter	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
th2	9.9553	0.8098	16	12.29	<.0001	0.05	8.2386	11.6719
th3	2.2104	0.1712	16	12.91	<.0001	0.05	1.8475	2.5732
th4	2.2036	0.2508	16	8.79	<.0001	0.05	1.6719	2.7353

<u>**SAS Program for Exercise 3D**</u> -

```
proc nlmixed data=three;
  parms th21=15 th22=5 k3=1 th31=2 th32=2 th33=2;
  c3=1/4.05; th23=k3*th21*th22*(1+c3)/(th22+c3*th21);
  th2=th21*(ray=1)+th22*(ray=2)+th23*(ray=3);
  th3=th31*(ray=1)+th32*(ray=2)+th33*(ray=3);
  t=(dose/th2)**th3; den=1+t; p=t/den;
  model dead~binomial(num,p);
run;
proc nlmixed data=three;
  parms th21=15 th22=5 th31=2 th32=2 th33=2; k3=1;
  c3=1/4.05; th23=k3*th21*th22*(1+c3)/(th22+c3*th21);
  th2=th21*(ray=1)+th22*(ray=2)+th23*(ray=3);
  th3=th31*(ray=1)+th32*(ray=2)+th33*(ray=3);
  t=(dose/th2)**th3; den=1+t; p=t/den;
  model dead~binomial(num,p);
run;
proc nlmixed data=three;
  parms th21=15 th22=5 k3=1 th3=2;
  c3=1/4.05; th23=k3*th21*th22*(1+c3)/(th22+c3*th21);
  th2=th21*(ray=1)+th22*(ray=2)+th23*(ray=3);
  t=(dose/th2)**th3; den=1+t; p=t/den;
  model dead~binomial(num,p);
run;
```

<u>**SAS Output for Exercise 3D**</u> -

Exercise 3 Sixth NLMixed

-2 Log Likelihood = 68.0

Parameter Estimates

Parameter	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
th21	12.7509	1.0567	16	12.07	<.0001	0.05	10.5107	14.9911
th22	4.8289	0.2496	16	19.35	<.0001	0.05	4.2998	5.3580
k3	0.6615	0.07038	16	9.40	<.0001	0.05	0.5123	0.8107
th31	2.7784	0.3780	16	7.35	<.0001	0.05	1.9771	3.5797
th32	3.1035	0.3877	16	8.00	<.0001	0.05	2.2816	3.9255
th33	1.8264	0.2130	16	8.58	<.0001	0.05	1.3749	2.2780

Exercise 3 Seventh NLMixed

-2 Log Likelihood = 82.3

Parameter Estimates

Parameter	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
th21	9.3429	1.1423	16	8.18	<.0001	0.05	6.9213	11.7644
th22	4.7052	0.2520	16	18.67	<.0001	0.05	4.1711	5.2393
th31	1.9881	0.3586	16	5.54	<.0001	0.05	1.2280	2.7482
th32	3.0805	0.3907	16	7.88	<.0001	0.05	2.2523	3.9088
th33	1.9350	0.2119	16	9.13	<.0001	0.05	1.4857	2.3843

Exercise 3 Eighth NLMixed

-2 Log Likelihood = 78.9

Parameter Estimates

Parameter	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
th21	11.9681	1.0285	16	11.64	<.0001	0.05	9.7878	14.1484
th22	4.7703	0.3025	16	15.77	<.0001	0.05	4.1290	5.4117
k3	0.7355	0.06751	16	10.90	<.0001	0.05	0.5924	0.8787
th3	2.3357	0.1743	16	13.40	<.0001	0.05	1.9662	2.7052

SAS Program for Exercise 4 -

```
Proc format;
  value agegroup 0 = '>=20 and <=65'
                 1 = '<20 or >65';
  value vision   0 = 'No Problem'
                 1 = 'Some Problem';
  value yes_no   0 = 'No'
                 1 = 'Yes';
run;
data one;
  input accident age vision drive_ed @@;
  agegroup=0; if age < 20 or age > 65 then agegroup=1;
  label accident='Accident in Last Year?'
        vision  ='Vision Problem?'
        drive_ed='Driver Education?';
  format accident drive_ed young old  yes_no.
         agegroup                     agegroup.
         vision                       vision.;
datalines;
1 17 1 1 1 44 0 0 1 48 1 0 1 55 0 0 1 75 1 1 0 35 0 1 0 42 1 1 0 57 0 0 0 28 0 1
0 20 0 1 0 38 1 0 0 45 0 1 0 47 1 1 0 52 0 0 0 55 0 1 1 68 1 0 1 18 1 0 1 68 0 0
1 48 1 1 1 17 0 0 1 70 1 1 1 72 1 0 1 35 0 1 1 19 1 0 1 62 1 0 0 39 1 1 0 40 1 1
0 55 0 0 0 68 0 1 0 25 1 0 0 17 0 0 0 45 0 1 0 44 0 1 0 67 0 0 0 55 0 1 1 61 1 0
1 19 1 0 1 69 0 0 1 23 1 1 1 19 0 0 1 72 1 1 1 74 1 0 1 31 0 1 1 16 1 0 1 61 1 0
;
proc logistic descending;
  model accident=young old vision drive_ed;
run;
```

SAS Output for Exercise 4 -

```
Third Logistic                The LOGISTIC Procedure

                             Model Fit Statistics
                                                   Intercept
                                     Intercept        and
                   Criterion           Only        Covariates
                  -2 Log L            61.827         44.365

Analysis of Maximum Likelihood Estimates
                                     Standard          Wald
Parameter     DF     Estimate          Error     Chi-Square     Pr > ChiSq
Intercept      1      -0.7518         0.7001         1.1532         0.2829
agegroup       1       1.8684         0.8187         5.2075         0.0225
vision         1       1.6676         0.7568         4.8553         0.0276
drive_ed       1      -1.0881         0.7635         2.0307         0.1541
```