Advanced Biostatistics **First Exam** **Name** _____
March 1, 2006

Directions: ***Thoroughly, clearly, neatly and correctly*** answer the following 3 exercises and 3 short answer questions in the space given (or on the back), showing all relevant calculations. Use $\alpha = 5\%$ throughout.

Exercises

1. (1.5 + 3 + 1.5 = 6 points) Amphetamine is a drug that it is felt suppresses appetite. To test this effect, a pharmacologist randomly allocated 24 rats to receive one of three treatment groups (to receive an injection of amphetamine at one of two dosage levels or an injection of saline solution, i.e., amphetamine at zero dose level). She then measured the amount of food consumed by each animal in the 3-hour period following injection, and these data are reproduced, graphed and analyzed on p.1 of the *Appendix*. Her goal is to test whether there is a significant linear relationship between amphetamine dose and food consumption.

   (a) State the assumptions that must be made for this SLR analysis ***in the context of this study***. Be specific and clear.

   (b) Test whether there is a significant linear relationship between amphetamine dose and food consumption. Be sure to write out the statistical model function, the null and alternative hypotheses, value of the relevant test statistic (TS) and distribution (including df), p-value, and your detailed and clear conclusion.

   Statistical Model Function _____

   Null _____ Alternative _____

   TS _____ df _____ p-value _____

   Detailed and clear conclusion

   (c) Clearly interpret the estimate of the slope parameter in this SLR model in the context of this study. Give the units.

2. (4 + 5 = 9) The new cholesterol-lowering supplement, Fibralo, was studied in a double-blind study against the marketed reference supplement, Gemfibrozil, in 34 non-insulin dependent diabetic patients. The study's objective was to compare the mean decrease in triglyceride levels (denoted "trichg" in the dataset) between the two treatment groups. The degree glycemic control, measured by hemoglobin A_{1c} levels (denoted "hgba1c" in the dataset), was thought to be an important factor as well. This covariate was measured at the start of the study and is shown in the listing on p.2 in the *Appendix* with the percent changes in triglycerides from pre-treatment to the end of the 10-week trial. The data are graphed and analyzed using Minitab on pp.3-4 of the *Appendix*.

   (a) After removing the hemoglobin covariate, test whether there is a difference in mean responses between supplements. ***For this part only, assume that the respective lines are parallel***. Give the relevant Minitab output # to use in the analysis, null and alternative hypothesis, observed test statistic and distribution (including df), p-value and detailed conclusion.

   Use Minitab output # _____ Write out model function: E(Y) = _____

   Null _____ Alternative _____

   TS _____ df _____ p-value _____

   Detailed conclusion

   (b) Dropping (no longer making) the parallelism assumption used in part (a), test whether a single regression line could be used for the two groups for the data graphed at the top of p.3 in the *Appendix*, clearly writing out your statistical model, your new hypotheses, calculated test statistic and its distribution (with df), p-value, and your detailed and clear conclusion. Identify the relevant output #(s).

   Write out the model function: E(Y) = _____

   Null hypothesis _____

   Alternative hypothesis _____

   Showing your calculations, give the calculated test statistic _____

   Degrees of freedom _____

   p-value _____

   Detailed and clear conclusion

3. (0.75 + 0.75 + 1 + 2 + 2 + 1 = 7.5 points) Cardiac researchers at a major U.S. research hospital conducted a ***crossover*** study of the effects of three drug treatments (labeled A, B and C) on heart rates of randomly chosen elderly individuals. The data are analyzed in SAS, and the SAS programs and outputs are given on pages 5 and 6 of the *Appendix*. Note that each of these programs includes a term for the carryover effect of the previous treatment (labeled "residue" here).

   (a) How many patients were used in this study? _____

   (b) How many sequences of the three drugs were used in this study? _____

   (c) Do you feel that the washout period used in this study (two weeks) was sufficient to remove the effects of the previous drug before the next one was given? Why or why not? Be specific (give relevant p-value).

   (d) Using only the **First SAS Output**, do you feel that the treatment averages differ? Be specific, giving your hypotheses, value of the test statistic, its distribution (including degrees of freedom), p-value, and a detailed conclusion.

   (e) Using only the **Second SAS Output**, do you feel that the treatment averages differ? Be specific, giving your hypotheses, value of the test statistic, its distribution (including degrees of freedom), p-value, and a detailed conclusion.

   (f) Summarize your findings for this study in terms of whether the treatments differ, and give the logical "next step" to help us make our final conclusion regarding these treatments.

Short Answer Questions (2.5 points each) – Be specific and clear in your responses

4. Returning to Exercise 2(a), if the covariate (hgba1c) was ignored and we wanted to compare the treatment means, what would our conclusion be and why? The correct output here is output # _____

5. A medical researcher wanted to study the efficacy of six drugs (labeled A – F), found 3 patients to whom she would randomize the drugs, but was told that a patient could only take four of the six drugs. She decided to give drugs A, B, D and E to the first patient, drugs B, C, E and F to the second patient, and drugs A, C, D and F to the third patient.

   Why is this block design "incomplete"? _____

   Why is this incomplete block design not "balanced"? _____

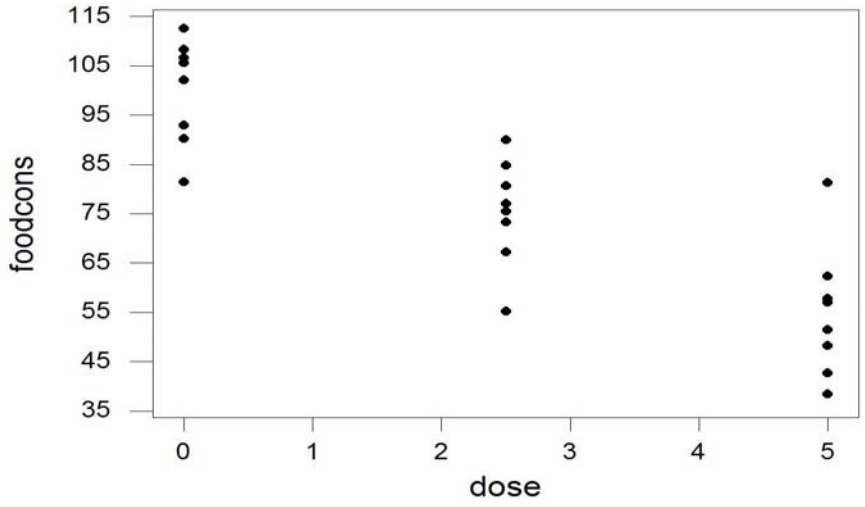
6. You collect count data for males and females and survival rates from a type of surgery, and you arrange the data into a two by two table with males on the first row and females on the second row, survival counts in the first column and deceased counts in the second column.

   Your calculated estimated odds ratio is 2.8239. Clearly interpret this estimate in the context of this setting.

   Your calculated 90% confidence interval for the true odds ratio is (1.2997,6.1356). Clearly interpret this interval in the context of this setting and give its ***ramifications***.

Exercise 1 output

	Dose = 0.0 mg/kg	Dose = 2.5 mg/kg	Dose = 5.0 mg/kg
	112.6	73.3	38.5
	102.1	84.8	81.3
	90.2	67.3	57.1
	81.5	55.3	62.3
	105.6	80.7	51.5
	93.0	90.0	48.3
	106.6	75.5	42.7
	108.3	77.1	57.9
Mean (g/kg)	100.0	75.5	55.0
SD (g/kg)	10.7	10.7	13.3
No. of animals	8	8	8



Output 1.1. Regression Analysis: foodcons versus dose

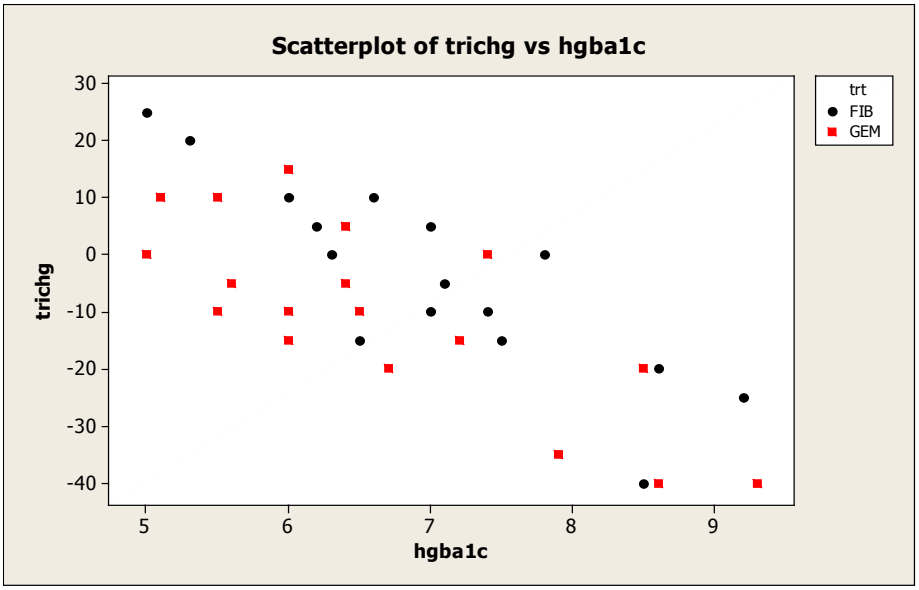
```
The regression equation is
foodcons = 99.3 - 9.01 dose

Predictor        Coef     SE Coef          T        P
Constant       99.331       3.680      26.99    0.000
dose           -9.008       1.140      -7.90    0.000

S = 11.40       R-Sq = 73.9%     R-Sq(adj) = 72.8%

Analysis of Variance
Source            DF          SS          MS         F        P
Regression         1      8113.5      8113.5     62.41    0.000
Residual Error    22      2859.9       130.0
Total             23     10973.4
```

Exercise 2 output

Printout of data

trt	hgba1c	trichg	trtfib	product
FIB	7.0	5	1	7.0
FIB	6.0	10	1	6.0
FIB	7.1	-5	1	7.1
FIB	8.6	-20	1	8.6
FIB	6.3	0	1	6.3
FIB	7.5	-15	1	7.5
FIB	6.6	10	1	6.6
FIB	7.4	-10	1	7.4
FIB	5.3	20	1	5.3
FIB	6.5	-15	1	6.5
FIB	6.2	5	1	6.2
FIB	7.8	0	1	7.8
FIB	8.5	-40	1	8.5
FIB	9.2	-25	1	9.2
FIB	5.0	25	1	5.0
FIB	7.0	-10	1	7.0
GEM	5.1	10	0	0.0
GEM	6.0	15	0	0.0
GEM	7.2	-15	0	0.0
GEM	6.4	5	0	0.0
GEM	5.5	10	0	0.0
GEM	6.0	-15	0	0.0
GEM	5.6	-5	0	0.0
GEM	5.5	-10	0	0.0
GEM	6.7	-20	0	0.0
GEM	8.6	-40	0	0.0
GEM	6.4	-5	0	0.0
GEM	6.0	-10	0	0.0
GEM	9.3	-40	0	0.0
GEM	8.5	-20	0	0.0
GEM	7.9	-35	0	0.0
GEM	7.4	0	0	0.0
GEM	5.0	0	0	0.0
GEM	6.5	-10	0	0.0

Exercise 2 output



Output 2.1. Regression Analysis: trichg versus hgba1c

The regression equation is
trichg = 65.0 - 10.6 hgba1c

Predictor	Coef	SE Coef	T	P
Constant	65.05	10.81	6.02	0.000
hgba1c	-10.629	1.564	-6.80	0.000

S = 10.85 R-Sq = 59.1% R-Sq(adj) = 57.8%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	5442.7	5442.7	46.21	0.000
Residual Error	32	3769.1	117.8		
Total	33	9211.8			

Output 2.2. Regression Analysis: trichg versus trtfib, hgba1c, product

The regression equation is
trichg = 58.0 + 26.0 trtfib - 10.3 hgba1c - 2.30 product

Predictor	Coef	SE Coef	T	P
Constant	58.05	12.66	4.58	0.000
trtfib	26.00	19.92	1.31	0.202
hgba1c	-10.283	1.875	-5.49	0.000
product	-2.304	2.867	-0.80	0.428

S = 9.734 R-Sq = 69.1% R-Sq(adj) = 66.1%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	3	6369.3	2123.1	22.41	0.000
Residual Error	30	2842.5	94.7		
Total	33	9211.8			

Exercise 2 output

Output 2.3. Regression Analysis: trichg versus trtfib

```
The regression equation is
trichg = - 10.3 + 6.22 trtfib

Predictor        Coef     SE Coef          T        P
Constant      -10.278       3.927      -2.62    0.013
trtfib          6.215       5.725       1.09    0.286

S = 16.66       R-Sq = 3.6%      R-Sq(adj) = 0.5%

Analysis of Variance

Source            DF          SS          MS         F        P
Regression         1       327.2       327.2      1.18    0.286
Residual Error    32      8884.5       277.6
Total             33      9211.8
```

Output 2.4. Regression Analysis: trichg versus trtfib, hgba1c

```
The regression equation is
trichg = 64.6 + 10.2 trtfib - 11.3 hgba1c

Predictor        Coef     SE Coef          T        P
Constant       64.593       9.643       6.70    0.000
trtfib         10.222       3.363       3.04    0.005
hgba1c        -11.268       1.410      -7.99    0.000

S = 9.678       R-Sq = 68.5%     R-Sq(adj) = 66.4%

Analysis of Variance
Source            DF          SS          MS         F        P
Regression         2      6308.1      3154.0     33.67    0.000
Residual Error    31      2903.7        93.7
Total             33      9211.8
```

First SAS Program for Exercise 3 -

```
proc glm;
  class pt seq per trt residue;
  model hr=seq pt(seq) per trt residue;
run;
```

First SAS Output for Exercise 3 -

The GLM Procedure

Dependent Variable: hr

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	29	6408.694444	220.989464	3.91	<.0001
Error	42	2372.583333	56.490079		
Corrected Total	71	8781.277778			

R-Square	Coeff Var	Root MSE	hr Mean
0.729813	9.301326	7.515988	80.80556

Source	DF	Type I SS	Mean Square	F Value	Pr > F
seq	5	508.944444	101.788889	1.80	0.1333
pt(seq)	18	4692.333333	260.685185	4.61	<.0001
per	2	146.777778	73.388889	1.30	0.2835
trt	2	668.777778	334.388889	5.92	0.0054
residue	2	391.861111	195.930556	3.47	0.0404

Source	DF	Type III SS	Mean Square	F Value	Pr > F
seq	5	701.183333	140.236667	2.48	0.0466
pt(seq)	18	4692.333333	260.685185	4.61	<.0001
per	1	6.750000	6.750000	0.12	0.7313
trt	2	343.950000	171.975000	3.04	0.0583
residue	2	391.861111	195.930556	3.47	0.0404

Second SAS Program for Exercise 3 -

➔ Note that "*trtb*" is a contrast between treatments A and B, and "*trtc*" is a contrast between treatments A and C

```
data one;
  input pt seq$ per basehr hr trt$ residue @@;
  if trt='A' then do;       trtb=-1; trtc=-1; end;
  else if trt='B' then do; trtb=1;  trtc=0;  end;
  else if trt='C' then do; trtb=0;  trtc=1;  end;
datalines;
  (Datalines deleted)

proc glm;
  class pt seq per residue;
  model hr=seq pt(seq) per trtb trtc residue;
run;
```

Second SAS Output for Exercise 3 -

```
The GLM Procedure

Dependent Variable: hr
                                        Sum of
      Source                    DF       Squares     Mean Square    F Value    Pr > F
      Model                     29     6408.694444      220.989464       3.91    <.0001
      Error                     42     2372.583333       56.490079
      Corrected Total           71     8781.277778

                    R-Square     Coeff Var      Root MSE       hr Mean
                    0.729813      9.301326      7.515988      80.80556

      Source                    DF       Type I SS     Mean Square    F Value    Pr > F
      seq                        5      508.944444      101.788889       1.80    0.1333
      pt(seq)                   18     4692.333333      260.685185       4.61    <.0001
      per                        2      146.777778       73.388889       1.30    0.2835
      trtb                       1      320.333333      320.333333       5.67    0.0219
      trtc                       1      348.444444      348.444444       6.17    0.0171
      residue                    2      391.861111      195.930556       3.47    0.0404

      Source                    DF     Type III SS     Mean Square    F Value    Pr > F
      seq                        5      701.183333      140.236667       2.48    0.0466
      pt(seq)                   18     4692.333333      260.685185       4.61    <.0001
      per                        1        6.750000        6.750000       0.12    0.7313
      trtb                       1      217.800000      217.800000       3.86    0.0562
      trtc                       1      292.612500      292.612500       5.18    0.0280
      residue                    2      391.861111      195.930556       3.47    0.0404
```