Name ______________________

Student Number ______________________

University of Toronto in Mississauga

December Examinations 2005
STA 442F
Duration - 3 hours

Aids allowed: Calculator. Printouts will be supplied.

$F = (\frac{n-p}{s})(\frac{a}{1-a}) \qquad a = \frac{R_F^2 - R_R^2}{1-R_R^2} = \frac{sF}{n-p+sF}$

1. (8 Points) Make up an original study that would employ a one-way multivariate analysis of covariance with one quantitative covariate and two dependent variables. Give a paragraph briefly describing the study, followed by list of the variables; classify each variable as independent or dependent, and quantitative or categorical.

Continued on page 2

2. (8 Points) Answer each question below True or False. Write "T" or "F" on the line. You must get at least 9 out of 10 correct in order to get credit for this question.

____ In an observational study, a statistically significant relationship between the independent variable and the dependent variable can provide some evidence of a causal relationship if the study is well controlled.

____ If a subject (case) provides data for more than one value of an independent variable, we call that independent variable a *within-subjects* variable.

____ We observe $r = -0.70$, $p = .009$. We conclude that that high values of X tend to go with low values of Y and low values of X tend to go with high values of Y.

____ If $p < .05$ we say the results are statistically significant at the .05 level, and we conclude that the independent variable and the dependent variable are unrelated.

____ In a study attempting to predict income from education and race, there is a significant interaction between education and race. This means that income and race are related.

____ When you add another independent variable in multiple regression, R^2 cannot go down.

____ We observe $r = 0.50$, $p = .002$. This means that 50% of the variation in the dependent variable is explained by a linear relationship with the independent variable.

____ An experimental study is one in which cases are randomly assigned to the different values of an independent variable.

____ When a relationship between the independent variable and the dependent variable is *not* statistically significant, we conclude there is no relationship between the two variables in the population.

____ A multivariate analysis is one with multiple independent variables.

Continued on page 3

3. (7 Points) Recall the **cars data**, in which you want to test for differences in fuel efficiency as a function of country (actually, region) of origin, while controlling for weight (X_1) and length (X_2) of vehicle. You need to set up a regression model using *cell means coding*; there will be no β_0 in the model.

   (a) List all the independent variables in your model. Say how the dummy variables are defined.

   (b) Make a table with three rows, one for each country. Make a column for each dummy variable, and show the value of the dummy variable for each country. Make one more column, a wider one. In that column, give the expected value of Y. Your expected value formulas must have only numbers, betas (β) and the symbols X_1 and X_2.

Continued on page 4

4. Now please look at the printout for the cars data, and answer these questions:

   (a) (5 Points) We want to test whether fuel efficiency is related to *any* of weight, length and country of origin. The null hypothesis is that fuel efficiency is not related to any of these variables.

      i. What is the value of the test statistic? The answer is a single number.

      ii. What is the p-value? The answer is a number or possibly a range of numbers, if $p < .0001$.

      iii. Are the results statistically significant at the 0.05 level? Answer Yes or No.

   (b) (3 Points) For any country of origin, and holding weight constant at a fixed level, is length positively related to fuel efficiency, or is it negatively related? How do you know? Your answer to this question should be the words "Positively Related" or "Negatively related," and a single number — a parameter estimate.

   (c) (5 Points) Controlling for country of origin, we want to test whether fuel efficiency is related to weight, length, or both.

      i. What is the value of the test statistic? The answer is a single number.

      ii. What is the p-value? The answer is a number or possibly a range of numbers, if $p < .0001$.

      iii. Are the results statistically significant at the 0.05 level? Answer Yes or No.

Continued on page 5

(d) (6 Points) Controlling for weight and length, we want to test whether fuel efficiency is related to country of origin.

i. What is the value of the test statistic? The answer is a single number.

ii. What is the p-value? The answer is a number or possibly a range of numbers, if $p < .0001$.

iii. Are the results statistically significant at the 0.05 level? Answer Yes or No.

iv. After allowing for the reduced model, what proportion of the remaining variation is explained by this effect?

(e) (14 Points) Now we will look at *one* of the pairwise comparisons.

i. We want to test the difference between cars of U.S. versus Japanese origin, controlling for weight and length.

A. What is the value of the test statistic? The answer is a single number.

B. What is the p-value? The answer is a number or possibly a range of numbers, if $p < .0001$.

C. Are the results statistically significant at the 0.05 level? Answer Yes or No.

D. Are the results still statistically significant with a Bonferroni correction for the fact that we are making three pairwise comparisons? Answer Yes or No.

E. After allowing for the reduced model, what proportion of the remaining variation is explained by this effect?

Continued on page 6

ii. What is the predicted fuel efficiency in Kilometers per litre for a Japanese car of average weight and length? By average, I mean set weight and length to their sample mean values. The answer to this question is a single number. Please "show your work" by writing down the expression you are trying to calculate, and then doing it with your calculator.

iii. What is the predicted fuel efficiency in Kilometers per litre for a U.S. car of average weight and length? By average, I mean set weight and length to their sample mean values. The answer to this question is a single number. Please "show your work" by writing down the expression you are trying to calculate, and then doing it with your calculator.

iv. In plain, non-statistical language, what do you conclude about the relative fuel efficiency of U.S. and Japanese cars *once you allow for length and weight*? The answer is one short sentence; you have more room than you need.

Continued on page 7

5. (8 Points) Consider the Titanic data. For just adults, we want to know if survival was related to sex when one allows for class.

   (a) Give the value of the test statistic. Your answer is a single number.

   (b) The test is significant beyond any doubt. What degrees of freedom would you use to calculate the p-value? Your answer is a single number.

   (c) Are the Bonferroni-corrected tests on the sub-tables all significant at the 0.05 level? Answer Yes or No. If the answer is No, specify which ones are not significant.

   (d) In plain, non-statistical language, what do you conclude from this analysis?

6. (5 Points) In the analysis of the **shoe data**, I followed up a significant main effect for period (Before, During and After the advertising campaign) with Bonferroni-corrected pairwise comparisons of marginal means. In plain, non-statistical language, what do you conclude from the pairwise comparisons? *Marks will be deducted if you use even a single statistical term.*

Continued on page 8

7. (8 Points) Make up an original study with one categorical within-subjects independent variable, one quantitative between-subjects independent variable, and one categorical dependent variable. Give a paragraph briefly describing the study, followed by list of the variables; classify each variable as independent or dependent, and categorical or quantitative. State whether each independent variable is between-subjects or within-subjects.

Continued on page 9

8. (5 Points) For the **exercise tolerance data**, please focus upon the following question: Does the relationship of Smoking to exercise tolerance depend on Fat level? Answer the question and describe the results in plain, non-statistical language. Your answer should be guided by the results of a statistical test, but you must not mention that test explicitly.

9. (5 Points) For the **farm data**, we want to know whether the effect of irrigation method depends upon the type of fertilizer.

   (a) What is the value of the test statistic? The answer is a single number.

   (b) What is the p-value? The answer is a number or possibly a range of numbers, if $p < .0001$.

   (c) Are the results statistically significant at the 0.05 level? Answer Yes or No.

   (d) In plain, non-statistical language, what do you conclude?

Continued on page 10

10. (8 Points) Give an example of a true experimental study with one independent variable, and the conclusions of the study are invalid because of a confounding variable. Hint: consider a drug study without a double blind.

11. (5 Points) In the **wine study**, 6 judges rated 4 wines on a scale from 0 to 40, presented in random order and with a double blind. Based on the Bonferroni-corrected tests for pairwise differences between means, sort the wines into groups such that the members of each group of wines are *not* significantly different, and arrange the groups in order from highest to lowest quality. Here is an example of what I mean (of course this answer is incorrect):

    "Wine 3 is rated highest, followed by Wines 1 and 4, followed by Wine 2."

Total marks = 100 points

STA442F05 Final Printout

Table of Contents

```
/********************** cars.sas *************************/
options linesize=79 pagesize=200 noovp formdlim='-';
title 'Metric Cars Data: STA442/1008 Final Exam';

proc format; /* Used to label values of the categorical variables */
     value carfmt    1 = 'US'
                     2 = 'Japanese'
                     3 = 'European' ;
data auto;
     infile 'mcars.dat' firstobs=2; /* Skip the first line */
     input id country kpl weight length;
/* Cell Means Indicator dummy vars */
     if country = 1 then c1=1;  else c1=0;
     if country = 2 then c2=1;  else c2=0;
     if country = 3 then c3=1;  else c3=0;
     label country = 'Country of Origin'
           kpl = 'Kilometers per Litre';
     format country carfmt.;

proc reg simple;
     title 'ANCOVA with cell means coding using proc reg';
     model kpl = weight length c1 c2 c3 / noint;
     TestA: test c1 = c2 = c3 = 0;
     TestB: test c1 = c2 = c3;
     TestC: test c1 = c2;
     TestD: test c1 = c3;
     TestE: test c2 = c3;
     TestF: test c1 = c2 = 0;
     TestG: test c1 = c3 = 0;
     TestH: test c2 = c3 = 0;
     TestI: test weight=length=0;
     TestJ: test weight=length=0, c1 = c2 = c3;
     TestK: test weight=length=0, c1 = c2 = c3 = 0;
```


cars.lst

```
-------------------------------------------------------------------------------

                  ANCOVA with cell means coding using proc reg                1
                                                 15:56 Sunday, December 4, 2005

                                The REG Procedure

                             Descriptive Statistics

                                                 Uncorrected                 Standard
Variable            Sum           Mean              SS      Variance     Deviation

Intercept     100.00000        1.00000       100.00000             0             0
weight           141345     1413.45000       212754681        131016     361.96142
length            48489      484.88600        23811410    3029.97113      55.04517
c1             73.00000        0.73000        73.00000       0.19909       0.44620
c2             13.00000        0.13000        13.00000       0.11424       0.33800
c3             14.00000        0.14000        14.00000       0.12162       0.34874
kpl           879.48000        8.79480      8440.74000       7.13019       2.67024

-------------------------------------------------------------------------------

                  ANCOVA with cell means coding using proc reg                2
                                                 15:56 Sunday, December 4, 2005

                                The REG Procedure
                                  Model: MODEL1
                  Dependent Variable: kpl Kilometers per Litre

               NOTE: No intercept in model. R-Square is redefined.

                              Analysis of Variance

                                     Sum of           Mean
Source                   DF         Squares         Square    F Value    Pr > F

Model                     5      8195.93850     1639.18770     636.12    <.0001
Error                    95       244.80150        2.57686
Uncorrected Total       100      8440.74000


             Root MSE              1.60526    R-Square     0.9710
             Dependent Mean        8.79480    Adj R-Sq     0.9695
             Coeff Var            18.25237

                               Parameter Estimates

                                             Parameter       Standard
  Variable    Label                   DF      Estimate          Error    t Value

  weight                               1      -0.00207        0.00139      -1.49
  length                               1      -0.02874        0.00925      -3.11
  c1                                   1      25.85295        2.83493       9.12
  c2                                   1      24.49010        2.75828       8.88
  c3                                   1      25.73537        2.69159       9.56
```

Parameter Estimates

Variable	Label	DF	Pr > |t|
weight		1	0.1389
length		1	0.0025
c1		1	<.0001
c2		1	<.0001
c3		1	<.0001

Test TestA Results for Dependent Variable kpl

Source	DF	Mean Square	F Value	Pr > F
Numerator	3	81.87863	31.77	<.0001
Denominator	95	2.57686		

Test TestB Results for Dependent Variable kpl

Source	DF	Mean Square	F Value	Pr > F
Numerator	2	8.50611	3.30	0.0411
Denominator	95	2.57686		

Test TestC Results for Dependent Variable kpl

Source	DF	Mean Square	F Value	Pr > F
Numerator	1	15.74243	6.11	0.0152
Denominator	95	2.57686		

Test TestD Results for Dependent Variable kpl

Source	DF	Mean Square	F Value	Pr > F
Numerator	1	0.12092	0.05	0.8290
Denominator	95	2.57686		

Test TestE Results for Dependent Variable kpl

Source	DF	Mean Square	F Value	Pr > F
Numerator	1	10.32125	4.01	0.0482
Denominator	95	2.57686		

Test TestF Results for Dependent Variable kpl

Source	DF	Mean Square	F Value	Pr > F
Numerator	2	107.30322	41.64	<.0001
Denominator	95	2.57686		

Test TestG Results for Dependent Variable kpl

Source	DF	Mean Square	F Value	Pr > F
Numerator	2	120.43287	46.74	<.0001
Denominator	95	2.57686		

Test TestH Results for Dependent Variable kpl

Source	DF	Mean Square	F Value	Pr > F
Numerator	2	122.60725	47.58	<.0001
Denominator	95	2.57686		

Test TestI Results for Dependent Variable kpl

Source	DF	Mean Square	F Value	Pr > F
Numerator	2	169.74773	65.87	<.0001
Denominator	95	2.57686		

Test TestJ Results for Dependent Variable kpl

Source	DF	Mean Square	F Value	Pr > F
Numerator	4	115.27195	44.73	<.0001
Denominator	95	2.57686		

Test TestK Results for Dependent Variable kpl

Source	DF	Mean Square	F Value	Pr > F
Numerator	5	1639.18770	636.12	<.0001
Denominator	95	2.57686		

```
/* titanic2.sas */
title 'Titanic Data: STTA442 Final Check';
options linesize=79 pagesize=35 noovp formdlim='_';

proc format;
     value clfmt  0 = 'Crew';
     value agefmt 1 = 'Adult' 0 = 'Child';
     value sexfmt 1 = 'Male' 0 = 'Female';
     value ynfmt  0 = 'No'  1 = 'Yes';

data iceberg;
     infile 'titanic.dat';
     input class age sex survived;
     pasclass = class; if pasclass=0 then pasclass=.;
     manclass = class; if sex = 0 then manstat=. ; if age=0 then manstat=.;
     adclass = class; if age=0 then adclass=.;
     label manclass = 'Class: Adult males only'
           adclass = 'Class: Adults only'
           pasclass = 'Passenger class';
     format class manclass adclass clfmt.;
     format age agefmt.; format sex sexfmt.; format survived ynfmt.;

proc freq;
     title2 'Just Adults';
     tables sex*survived*adclass
            / norow nopercent chisq expected;
     tables adclass*survived*sex
            / norow nopercent chisq expected;

data kids;
     set iceberg;
     if age=0;

proc freq;
     title2 'Just Children';
     tables survived * (class sex)
            / norow nopercent chisq expected;
```


titanic2.lst

```
                   Titanic Data: STTA442 Final Check                           1
                              Just Adults    17:05 Sunday, December 4, 2005

                               The FREQ Procedure

                           Table 1 of survived by adclass
                             Controlling for sex=Female

             survived     adclass(Class: Adults only)

             Frequency|
             Expected |
             Col Pct  |Crew    |       1|       2|       3|  Total
             ---------+--------+--------+--------+--------+
             No       |      3 |      4 |     13 |     89 |    109
                      | 5.8988 | 36.932 | 23.852 | 42.318 |
                      |  13.04 |   2.78 |  13.98 |  53.94 |
             ---------+--------+--------+--------+--------+
             Yes      |     20 |    140 |     80 |     76 |    316
                      | 17.101 | 107.07 | 69.148 | 122.68 |
                      |  86.96 |  97.22 |  86.02 |  46.06 |
             ---------+--------+--------+--------+--------+
             Total          23      144       93      165      425

                             Frequency Missing = 45
```

```
                   Titanic Data: STTA442 Final Check                           2
                              Just Adults    17:05 Sunday, December 4, 2005

                               The FREQ Procedure

                      Statistics for Table 1 of survived by adclass
                             Controlling for sex=Female

            Statistic                     DF       Value      Prob
            ------------------------------------------------------
            Chi-Square                     3    117.3107    <.0001
            Likelihood Ratio Chi-Square    3    126.6013    <.0001
            Mantel-Haenszel Chi-Square     1     95.5274    <.0001
            Phi Coefficient                       0.5254
            Contingency Coefficient               0.4651
            Cramer's V                            0.5254

                           Effective Sample Size = 425
                             Frequency Missing = 45
```

Titanic Data: STTA442 Final Check 3
Just Adults 17:05 Sunday, December 4, 2005

The FREQ Procedure

Table 2 of survived by adclass
Controlling for sex=Male

survived adclass(Class: Adults only)

Frequency / Expected / Col Pct	Crew	1	2	3	Total
No	670	118	154	387	1329
	687.22	139.52	133.94	368.33	
	77.73	67.43	91.67	83.77	
Yes	192	57	14	75	338
	174.78	35.483	34.064	93.675	
	22.27	32.57	8.33	16.23	
Total	862	175	168	462	1667

Frequency Missing = 64

Titanic Data: STTA442 Final Check 4
Just Adults 17:05 Sunday, December 4, 2005

The FREQ Procedure

Statistics for Table 2 of survived by adclass
Controlling for sex=Male

Statistic	DF	Value	Prob
Chi-Square	3	37.9879	<.0001
Likelihood Ratio Chi-Square	3	39.6102	<.0001
Mantel-Haenszel Chi-Square	1	12.0997	0.0005
Phi Coefficient		0.1510	
Contingency Coefficient		0.1493	
Cramer's V		0.1510	

Effective Sample Size = 1667
Frequency Missing = 64

Titanic Data: STTA442 Final Check 5
Just Adults 17:05 Sunday, December 4, 2005

The FREQ Procedure

Table 1 of survived by sex
Controlling for adclass=Crew

survived sex

Frequency / Expected / Col Pct	Female	Male	Total
No	3	670	673
	17.49	655.51	
	13.04	77.73	
Yes	20	192	212
	5.5096	206.49	
	86.96	22.27	
Total	23	862	885

Titanic Data: STTA442 Final Check 6
Just Adults 17:05 Sunday, December 4, 2005

The FREQ Procedure

Statistics for Table 1 of survived by sex
Controlling for adclass=Crew

Statistic	DF	Value	Prob
Chi-Square	1	51.4522	<.0001
Likelihood Ratio Chi-Square	1	42.3508	<.0001
Continuity Adj. Chi-Square	1	47.9627	<.0001
Mantel-Haenszel Chi-Square	1	51.3941	<.0001
Phi Coefficient		-0.2411	
Contingency Coefficient		0.2344	
Cramer's V		-0.2411	

Fisher's Exact Test

Cell (1,1) Frequency (F)	3
Left-sided Pr <= F	1.655E-10
Right-sided Pr >= F	1.0000
Table Probability (P)	1.588E-10
Two-sided Pr <= P	1.655E-10

Sample Size = 885

```
                    Titanic Data: STTA442 Final Check                       7
                               Just Adults    17:05 Sunday, December 4, 2005

                       Table 2 of survived by sex
                        Controlling for adclass=1

                      survived     sex

                      Frequency|
                      Expected |
                      Col Pct  |Female  |Male    |  Total
                      ---------+--------+--------+
                      No       |      4 |    118 |    122
                               | 55.072 | 66.928 |
                               |   2.78 |  67.43 |
                      ---------+--------+--------+
                      Yes      |    140 |     57 |    197
                               | 88.928 | 108.07 |
                               |  97.22 |  32.57 |
                      ---------+--------+--------+
                      Total         144      175      319
```

```
                    Titanic Data: STTA442 Final Check                       8
                               Just Adults    17:05 Sunday, December 4, 2005

                             The FREQ Procedure

                  Statistics for Table 2 of survived by sex
                          Controlling for adclass=1

            Statistic                     DF       Value      Prob
            ------------------------------------------------------
            Chi-Square                     1    139.8018    <.0001
            Likelihood Ratio Chi-Square    1    166.9868    <.0001
            Continuity Adj. Chi-Square     1    137.0779    <.0001
            Mantel-Haenszel Chi-Square     1    139.3636    <.0001
            Phi Coefficient                      -0.6620
            Contingency Coefficient               0.5520
            Cramer's V                           -0.6620

                            Fisher's Exact Test
                      ----------------------------------
                      Cell (1,1) Frequency (F)         4
                      Left-sided Pr <= F       1.541E-37
                      Right-sided Pr >= F         1.0000

                      Table Probability (P)    1.521E-37
                      Two-sided Pr <= P        1.625E-37

                               Sample Size = 319
```



```
                    Titanic Data: STTA442 Final Check                       9
                               Just Adults    17:05 Sunday, December 4, 2005

                             The FREQ Procedure

                       Table 3 of survived by sex
                        Controlling for adclass=2

                      survived     sex

                      Frequency|
                      Expected |
                      Col Pct  |Female  |Male    |  Total
                      ---------+--------+--------+
                      No       |     13 |    154 |    167
                               | 59.506 | 107.49 |
                               |  13.98 |  91.67 |
                      ---------+--------+--------+
                      Yes      |     80 |     14 |     94
                               | 33.494 | 60.506 |
                               |  86.02 |   8.33 |
                      ---------+--------+--------+
                      Total          93      168      261
```

```
                    Titanic Data: STTA442 Final Check                      10
                               Just Adults    17:05 Sunday, December 4, 2005

                             The FREQ Procedure

                  Statistics for Table 3 of survived by sex
                          Controlling for adclass=2

            Statistic                     DF       Value      Prob
            ------------------------------------------------------
            Chi-Square                     1    156.7827    <.0001
            Likelihood Ratio Chi-Square    1    169.5028    <.0001
            Continuity Adj. Chi-Square     1    153.4296    <.0001
            Mantel-Haenszel Chi-Square     1    156.1820    <.0001
            Phi Coefficient                      -0.7750
            Contingency Coefficient               0.6126
            Cramer's V                           -0.7750


                            Fisher's Exact Test
                      ----------------------------------
                      Cell (1,1) Frequency (F)        13
                      Left-sided Pr <= F       4.040E-38
                      Right-sided Pr >= F         1.0000

                      Table Probability (P)    3.981E-38
                      Two-sided Pr <= P        4.040E-38

                               Sample Size = 261
```

The FREQ Procedure

Table 4 of survived by sex
Controlling for adclass=3

survived sex

Frequency / Expected / Col Pct	Female	Male	Total
No	89	387	476
	125.26	350.74	
	53.94	83.77	
Yes	76	75	151
	39.737	111.26	
	46.06	16.23	
Total	165	462	627

The FREQ Procedure

Statistics for Table 4 of survived by sex
Controlling for adclass=3

Statistic	DF	Value	Prob
Chi-Square	1	59.1594	<.0001
Likelihood Ratio Chi-Square	1	54.7194	<.0001
Continuity Adj. Chi-Square	1	57.5393	<.0001
Mantel-Haenszel Chi-Square	1	59.0651	<.0001
Phi Coefficient		-0.3072	
Contingency Coefficient		0.2936	
Cramer's V		-0.3072	

Fisher's Exact Test

Cell (1,1) Frequency (F)	89
Left-sided Pr <= F	1.414E-13
Right-sided Pr >= F	1.0000
Table Probability (P)	1.102E-13
Two-sided Pr <= P	2.336E-13

Sample Size = 627

The FREQ Procedure

Table of survived by class

survived class

Frequency / Expected / Col Pct	1	2	3	Total
No	0	0	52	52
	2.8624	11.45	37.688	
	0.00	0.00	65.82	
Yes	6	24	27	57
	3.1376	12.55	41.312	
	100.00	100.00	34.18	
Total	6	24	79	109

The FREQ Procedure

Statistics for Table of survived by class

Statistic	DF	Value	Prob
Chi-Square	2	37.7615	<.0001
Likelihood Ratio Chi-Square	2	49.4084	<.0001
Mantel-Haenszel Chi-Square	1	32.4416	<.0001
Phi Coefficient		0.5886	
Contingency Coefficient		0.5072	
Cramer's V		0.5886	

WARNING: 33% of the cells have expected counts less than 5. Chi-Square may not be a valid test.

Sample Size = 109

```
                    Titanic Data: STTA442 Final Check                      15
                              Just Children   17:05 Sunday, December 4, 2005

                           The FREQ Procedure

                        Table of survived by sex

                     survived     sex

                     Frequency|
                     Expected |
                     Col Pct  |Female  |Male    |  Total
                     ---------+--------+--------+
                     No       |     17 |     35 |     52
                              | 21.468 | 30.532 |
                              |  37.78 |  54.69 |
                     ---------+--------+--------+
                     Yes      |     28 |     29 |     57
                              | 23.532 | 33.468 |
                              |  62.22 |  45.31 |
                     ---------+--------+--------+
                     Total          45       64      109
```

```
                    Titanic Data: STTA442 Final Check                      16
                              Just Children   17:05 Sunday, December 4, 2005

                           The FREQ Procedure

                  Statistics for Table of survived by sex

           Statistic                     DF       Value      Prob
           ------------------------------------------------------
           Chi-Square                     1      3.0284    0.0818
           Likelihood Ratio Chi-Square    1      3.0502    0.0807
           Continuity Adj. Chi-Square     1      2.3885    0.1222
           Mantel-Haenszel Chi-Square     1      3.0006    0.0832
           Phi Coefficient                      -0.1667
           Contingency Coefficient               0.1644
           Cramer's V                           -0.1667


                          Fisher's Exact Test
                     ----------------------------------
                     Cell (1,1) Frequency (F)        17
                     Left-sided Pr <= F          0.0608
                     Right-sided Pr >= F         0.9739

                     Table Probability (P)       0.0347
                     Two-sided Pr <= P           0.1188

                          Sample Size = 109
```



```
/*  extoler.sas */
title 'Exercise Tolerence Data: STTA442 Final Exam';
options linesize=79 pagesize=200 noovp formdlim='_';

proc format;
     value sexfmt 1 = 'Male'  2 = 'Female';
     value fatfmt 1 = 'Low'   2 = 'High';
     value smfmt  1 = 'Light' 2 = 'Heavy';

data nopain;
     infile 'extoler.dat' firstobs=2; /* Skip the first line */
     input Toler Sex Fat Smoke;
     format sex sexfmt.; format fat fatfmt.; format smoke smfmt.;
     label toler = 'Exercise tolerence in minutes';

proc glm;
     class Sex Fat Smoke;
     model toler = Sex|Fat|Smoke;
     means Sex|Fat|Smoke;
```

extoler.lst

```
                  Exercise Tolerence Data: STTA442 Final Exam                1
                                               23:36 Sunday, December 4, 2005

                               The GLM Procedure

                           Class Level Information

                     Class         Levels    Values

                     Sex                2    Female Male

                     Fat                2    High Low

                     Smoke              2    Heavy Light


                         Number of observations    24
```

```
Dependent Variable: Toler   Exercise tolerence in minutes

                                        Sum of
Source                      DF         Squares     Mean Square    F Value    Pr > F

Model                        7     588.5829167      84.0832738       9.01    0.0002

Error                       16     149.3666667       9.3354167

Corrected Total             23     737.9495833
```

R-Square	Coeff Var	Root MSE	Toler Mean
0.797592	18.77833	3.055391	16.27083

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Sex	1	176.5837500	176.5837500	18.92	0.0005
Fat	1	242.5704167	242.5704167	25.98	0.0001
Sex*Fat	1	13.6504167	13.6504167	1.46	0.2441
Smoke	1	70.3837500	70.3837500	7.54	0.0144
Sex*Smoke	1	11.0704167	11.0704167	1.19	0.2923
Fat*Smoke	1	72.4537500	72.4537500	7.76	0.0132
Sex*Fat*Smoke	1	1.8704167	1.8704167	0.20	0.6604

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Sex	1	176.5837500	176.5837500	18.92	0.0005
Fat	1	242.5704167	242.5704167	25.98	0.0001
Sex*Fat	1	13.6504167	13.6504167	1.46	0.2441
Smoke	1	70.3837500	70.3837500	7.54	0.0144
Sex*Smoke	1	11.0704167	11.0704167	1.19	0.2923
Fat*Smoke	1	72.4537500	72.4537500	7.76	0.0132
Sex*Fat*Smoke	1	1.8704167	1.8704167	0.20	0.6604

The GLM Procedure

Level of Sex	N	Toler Mean	Toler Std Dev
Female	12	13.5583333	4.72102807
Male	12	18.9833333	5.36145050

Level of Fat	N	Toler Mean	Toler Std Dev
High	12	13.0916667	3.72667493
Low	12	19.4500000	5.58089273

Level of Sex	Level of Fat	N	Toler Mean	Toler Std Dev
Female	High	6	11.1333333	3.61146323
Female	Low	6	15.9833333	4.67735680
Male	High	6	15.0500000	2.88218667
Male	Low	6	22.9166667	4.21920214

Level of Smoke	N	Toler Mean	Toler Std Dev
Heavy	12	14.5583333	4.85713315
Light	12	17.9833333	6.09065355

Level of Sex	Level of Smoke	N	Toler Mean	Toler Std Dev
Female	Heavy	6	11.1666667	3.20104150
Female	Light	6	15.9500000	5.00589652
Male	Heavy	6	17.9500000	3.74793276
Male	Light	6	20.0166667	6.82859185

Level of Fat	Level of Smoke	N	Toler Mean	Toler Std Dev
High	Heavy	6	13.1166667	4.82303501
High	Light	6	13.0666667	2.70012345
Low	Heavy	6	16.0000000	4.86333219
Low	Light	6	22.9000000	4.03782119

Level of Sex	Level of Fat	Level of Smoke	N	Toler Mean	Toler Std Dev
Female	High	Heavy	3	10.2000000	4.15090352
Female	High	Light	3	12.0666667	3.57258077
Female	Low	Heavy	3	12.1333333	2.36290781
Female	Low	Light	3	19.8333333	2.15483951
Male	High	Heavy	3	16.0333333	3.92470806
Male	High	Light	3	14.0666667	1.56950098
Male	Low	Heavy	3	19.8666667	2.94844592
Male	Low	Light	3	25.9666667	2.81128678

```
/* finalfarm.sas */
options linesize=79 pagesize=100 noovp formdlim='_';
title 'STA442F05 Final Exam: Farm data from NWK Prob 29.20, p. 1204';

data wheat;
     infile 'farm2.dat'  firstobs=2; /* Skip the first line */
     input Field IgMethod Yield1 Yield2;
     label igmethod = 'Irrigation Method'
           yield1 = 'Crop Yield with Fertilizer 1'
           yield2 = 'Crop Yield with Fertilizer 2';

proc means mean stddev n;
     class igmethod;
     var Yield1 Yield2;

proc glm;
     class igmethod;
     model yield1 yield2 = igmethod;
     repeated Fertlizr / short summary;
```

finalfarm.lst

```
_______________________________________________________________________________

         STA442F05 Final Exam: Farm data from NWK Prob 29.20, p. 1204         1
                                           20:38 Monday, December 5, 2005

                                The MEANS Procedure

      Irrigation    N
          Method  Obs  Variable  Label                                   Mean
    ----------------------------------------------------------------------
               1    5  Yield1    Crop Yield with Fertilizer 1    35.4000000
                       Yield2    Crop Yield with Fertilizer 2    39.2000000

               2    5  Yield1    Crop Yield with Fertilizer 1    52.2000000
                       Yield2    Crop Yield with Fertilizer 2    55.8000000
    ----------------------------------------------------------------------

    Irrigation    N
        Method  Obs  Variable  Label                                Std Dev   N
  ---------------------------------------------------------------------------
             1    5  Yield1    Crop Yield with Fertilizer 1     6.5038450   5
                     Yield2    Crop Yield with Fertilizer 2     6.8337398   5

             2    5  Yield1    Crop Yield with Fertilizer 1     7.0498227   5
                     Yield2    Crop Yield with Fertilizer 2     8.5848704   5
  ---------------------------------------------------------------------------
```



```
_______________________________________________________________________________

         STA442F05 Final Exam: Farm data from NWK Prob 29.20, p. 1204         2
                                           20:38 Monday, December 5, 2005

                                The GLM Procedure

                            Class Level Information

                        Class         Levels    Values

                        IgMethod           2    1 2


                         Number of observations    10

_______________________________________________________________________________

         STA442F05 Final Exam: Farm data from NWK Prob 29.20, p. 1204         3
                                           20:38 Monday, December 5, 2005

                                The GLM Procedure

Dependent Variable: Yield1   Crop Yield with Fertilizer 1

                                        Sum of
Source                      DF         Squares     Mean Square    F Value    Pr > F

Model                        1      705.600000      705.600000      15.34    0.0044

Error                        8      368.000000       46.000000

Corrected Total              9     1073.600000


              R-Square     Coeff Var      Root MSE    Yield1 Mean

              0.657228      15.48477      6.782330       43.80000


Source                      DF       Type I SS     Mean Square    F Value    Pr > F

IgMethod                     1     705.6000000     705.6000000      15.34    0.0044


Source                      DF     Type III SS     Mean Square    F Value    Pr > F

IgMethod                     1     705.6000000     705.6000000      15.34    0.0044
```

The GLM Procedure

Dependent Variable: Yield2 Crop Yield with Fertilizer 2

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	688.900000	688.900000	11.44	0.0096
Error	8	481.600000	60.200000		
Corrected Total	9	1170.500000			

R-Square	Coeff Var	Root MSE	Yield2 Mean
0.588552	16.33445	7.758866	47.50000

Source	DF	Type I SS	Mean Square	F Value	Pr > F
IgMethod	1	688.9000000	688.9000000	11.44	0.0096

Source	DF	Type III SS	Mean Square	F Value	Pr > F
IgMethod	1	688.9000000	688.9000000	11.44	0.0096

The GLM Procedure
Repeated Measures Analysis of Variance

Repeated Measures Level Information

Dependent Variable	Yield1	Yield2
Level of Fertlizr	1	2

Manova Test Criteria and Exact F Statistics
for the Hypothesis of no Fertlizr Effect
H = Type III SSCP Matrix for Fertlizr
E = Error SSCP Matrix

S=1 M=-0.5 N=3

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.14916097	45.63	1	8	0.0001
Pillai's Trace	0.85083903	45.63	1	8	0.0001
Hotelling-Lawley Trace	5.70416667	45.63	1	8	0.0001
Roy's Greatest Root	5.70416667	45.63	1	8	0.0001

Manova Test Criteria and Exact F Statistics for
the Hypothesis of no Fertlizr*IgMethod Effect
H = Type III SSCP Matrix for Fertlizr*IgMethod
E = Error SSCP Matrix

S=1 M=-0.5 N=3

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.99585062	0.03	1	8	0.8597
Pillai's Trace	0.00414938	0.03	1	8	0.8597
Hotelling-Lawley Trace	0.00416667	0.03	1	8	0.8597
Roy's Greatest Root	0.00416667	0.03	1	8	0.8597

The GLM Procedure
Repeated Measures Analysis of Variance
Tests of Hypotheses for Between Subjects Effects

Source	DF	Type III SS	Mean Square	F Value	Pr > F
IgMethod	1	1394.450000	1394.450000	13.32	0.0065
Error	8	837.600000	104.700000		

The GLM Procedure
Repeated Measures Analysis of Variance
Univariate Tests of Hypotheses for Within Subject Effects

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Fertlizr	1	68.45000000	68.45000000	45.63	0.0001
Fertlizr*IgMethod	1	0.05000000	0.05000000	0.03	0.8597
Error(Fertlizr)	8	12.00000000	1.50000000		

```
/* finalshoe.sas */
options linesize=79 pagesize=100 noovp formdlim='_';
title 'STA442F05 Final Exam: Shoe data from NWK Table 29.10';

data uvfoot;
     infile 'shoes.dat' firstobs=2; /* Skip the first line */
     input row  ident  period  campaign  sales;

proc mixed cl;
     class period campaign;
     model sales = period|campaign;
     repeated / type = un  subject = ident;
     lsmeans period / adjust=bon;
```

finalshoe.lst

```
             STA442F05 Final Exam: Shoe data from NWK Table 29.10            1
                                           20:15 Monday, December 5, 2005

                              The Mixed Procedure

                               Model Information

            Data Set                     WORK.UVFOOT
            Dependent Variable           sales
            Covariance Structure         Unstructured
            Subject Effect               ident
            Estimation Method            REML
            Residual Variance Method     None
            Fixed Effects SE Method      Model-Based
            Degrees of Freedom Method    Between-Within


                           Class Level Information

             Class       Levels    Values

             period           3    1 2 3
             campaign         2    1 2


                                  Dimensions

                      Covariance Parameters             6
                      Columns in X                     12
                      Columns in Z                      0
                      Subjects                         10
                      Max Obs Per Subject               3
                      Observations Used                30
                      Observations Not Used             0
                      Total Observations               30
```



```
                              Iteration History

         Iteration    Evaluations    -2 Res Log Like       Criterion

                 0              1       347.69129449
                 1              1       256.26815630      0.00000000


                          Convergence criteria met.


                      Covariance Parameter Estimates

      Cov Parm    Subject    Estimate     Alpha       Lower       Upper

      UN(1,1)     ident         75290      0.05       34350      276326
      UN(2,1)     ident         79440      0.05     1502.45      157377
      UN(2,2)     ident         84197      0.05       38414      309017
      UN(3,1)     ident         72731      0.05     1410.04      144051
      UN(3,2)     ident         76682      0.05     1373.22      151990
      UN(3,3)     ident         70440      0.05       32138      258527


                               Fit Statistics

                    -2 Res Log Likelihood          256.3
                    AIC (smaller is better)        268.3
                    AICC (smaller is better)       273.2
                    BIC (smaller is better)        270.1


                      Null Model Likelihood Ratio Test

                        DF    Chi-Square      Pr > ChiSq

                         5         91.42          <.0001


                       Type 3 Tests of Fixed Effects

                              Num     Den
            Effect             DF      DF    F Value    Pr > F

            period              2       8      54.03    <.0001
            campaign            1       8       0.73    0.4166
            period*campaign     2       8       2.94    0.1101

                            Least Squares Means

                                          Standard
   Effect    period    Estimate       Error      DF    t Value    Pr > |t|

   period    1           648.40     86.7695       8       7.47      <.0001
   period    2           728.80     91.7588       8       7.94      <.0001
   period    3           616.40     83.9284       8       7.34      <.0001
```

```
STA442F05 Final Exam: Shoe data from NWK Table 29.10            2
                                   20:15 Monday, December 5, 2005

                    The Mixed Procedure

           Differences of Least Squares Means

                                        Standard
 Effect   period   _period   Estimate      Error     DF   t Value   Pr > |t|

 period   1        2         -80.4000     7.7910      8    -10.32     <.0001
 period   1        3          32.0000     5.1764      8      6.18     0.0003
 period   2        3           112.40    11.2821      8      9.96     <.0001

           Differences of Least Squares Means

        Effect   period   _period   Adjustment       Adj P

        period   1        2         Bonferroni      <.0001
        period   1        3         Bonferroni      0.0008
        period   2        3         Bonferroni      <.0001
```

```
/* finalwine.sas */
options linesize=79 pagesize=100 noovp formdlim='_';
title 'STA442F05 Final Exam: Wine tasting data from NWK p. 1169';

data wine;
     infile 'wine.dat'  firstobs=2; /* Skip the first line (header) */
     input Judge  Wine  Rating;

proc means mean stddev;
     class wine;
     var rating;

proc mixed;
     class wine;
     model rating = wine;
     repeated / type = cs  subject = judge;
     lsmeans wine / adjust=bon;
```


finalwine.lst

```
STA442F05 Final Exam: Wine tasting data from NWK p. 1169          1
                                   21:58 Monday, December 5, 2005

                    The MEANS Procedure

                Analysis Variable : Rating

                     N
            Wine   Obs            Mean         Std Dev
    --------------------------------------------------
               1     6      20.0000000       3.7416574

               2     6      22.0000000       3.1622777

               3     6      26.6666667       2.6583203

               4     6      26.0000000       2.6076810
    --------------------------------------------------
```

```
STA442F05 Final Exam: Wine tasting data from NWK p. 1169          2
                                   21:58 Monday, December 5, 2005

                    The Mixed Procedure

                    Model Information

  Data Set                     WORK.WINE
  Dependent Variable           Rating
  Covariance Structure         Compound Symmetry
  Subject Effect               Judge
  Estimation Method            REML
  Residual Variance Method     Profile
  Fixed Effects SE Method      Model-Based
  Degrees of Freedom Method    Between-Within


                 Class Level Information

     Class    Levels    Values

     Wine          4    1 2 3 4


                    Iteration History

 Iteration    Evaluations    -2 Res Log Like       Criterion

         0              1       108.88011634
         1              1        82.62155007      0.00000000
```

Convergence criteria met.

Covariance Parameter Estimates

Cov Parm	Subject	Estimate
CS	Judge	8.4000
Residual		1.0667

Fit Statistics

-2 Res Log Likelihood	82.6
AIC (smaller is better)	86.6
AICC (smaller is better)	87.3
BIC (smaller is better)	86.2

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
1	26.26	<.0001

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
Wine	3	15	57.50	<.0001

Least Squares Means

Effect	Wine	Estimate	Standard Error	DF	t Value	Pr > |t|
Wine	1	20.0000	1.2561	15	15.92	<.0001
Wine	2	22.0000	1.2561	15	17.51	<.0001
Wine	3	26.6667	1.2561	15	21.23	<.0001
Wine	4	26.0000	1.2561	15	20.70	<.0001

Differences of Least Squares Means

Effect	Wine	_Wine	Estimate	Standard Error	DF	t Value	Pr > |t|	Adjustment
Wine	1	2	-2.0000	0.5963	15	-3.35	0.0043	Bonferroni
Wine	1	3	-6.6667	0.5963	15	-11.18	<.0001	Bonferroni
Wine	1	4	-6.0000	0.5963	15	-10.06	<.0001	Bonferroni
Wine	2	3	-4.6667	0.5963	15	-7.83	<.0001	Bonferroni
Wine	2	4	-4.0000	0.5963	15	-6.71	<.0001	Bonferroni
Wine	3	4	0.6667	0.5963	15	1.12	0.2811	Bonferroni

The Mixed Procedure

Differences of Least Squares Means

Effect	Wine	_Wine	Adj P
Wine	1	2	0.0261
Wine	1	3	<.0001
Wine	1	4	<.0001
Wine	2	3	<.0001
Wine	2	4	<.0001
Wine	3	4	1.0000