STAT362 Homework Assignment 13

Sharon O'Boyle

Problem 1: Problem 1, 7.11 p. 235

SAS Program

```
* Sharon O'Boyle;
* Stat 362;
* Homework Assignment 13;

* Problem 1, 7.11 p. 235 ;
* Topic: Analysis of Covariance;

* Create dataset;

data math;
input GROUP $ M_SCORE AGE @@;
datalines;
A 90 16 B 92 18 C 97 18 A 88 15 B 88 13 C 92 17 A 72 12 B 76 12 C 88 16 A 82 14 B
78 14 C 92 17 A 65 12 B 90 17
C 99 17 A 74 13 B 68 12 C 82 14
;
run;

 proc print data=math;
 title 'math data';
 run;

 *** Part A - One-Way ANOVA ;

PROC ANOVA DATA = math;
 class group;
 MODEL m_score age = group;
 means group / snk;
 TITLE 'REGRESSION of m_score and age on group';
RUN;
QUIT;

 *** Part B - look for homogeneity of score and age among the 3 groups ;

PROC GLM DATA = math;
 class group;
 MODEL m_score = age group age*group;
 TITLE 'PROC GLM Homogeneity ';
RUN;
QUIT;

 *** Part C - analysis of covariance;

PROC GLM DATA = math;
 class group;
 MODEL m_score = age group /SS3 ;
```

```
 LSMEANS GROUP / PDIFF;
 TITLE 'Analysis of Covariance';
RUN;
QUIT;
```

SAS Log

```
NOTE: Copyright (c) 2002-2010 by SAS Institute Inc., Cary, NC, USA.
NOTE: SAS (r) Proprietary Software 9.3 (TS1M1)
      Licensed to GEORGE MASON UNIVERSITY-SFA T&R, Site 70008900.
NOTE: This session is executing on the W32_7PRO  platform.



NOTE: Updated analytical products:

SAS/STAT 9.3_M1, SAS/ETS 9.3_M1, SAS/OR 9.3_M1

NOTE: SAS initialization used:
      real time           26.53 seconds
      cpu time            2.74 seconds

1
2     * Sharon O'Boyle;
3     * Stat 362;
4     * Homework Assignment 13;
5
6     * Problem 1, 7.11 p. 235 ;
7     * Topic: Analysis of Covariance;
8
9     * Create dataset;
10
11    data math;
12    input GROUP $ M_SCORE AGE @@;
13    datalines;

NOTE: SAS went to a new line when INPUT statement reached past the end of a line.
NOTE: The data set WORK.MATH has 18 observations and 3 variables.
NOTE: DATA statement used (Total process time):
      real time           0.30 seconds
      cpu time            0.01 seconds


16    ;
17    run;
18
19     proc print data=math;
NOTE: Writing HTML Body file: sashtml.htm
20     title 'math data';
21     run;

NOTE: There were 18 observations read from the data set WORK.MATH.
NOTE: PROCEDURE PRINT used (Total process time):
      real time           2.21 seconds
      cpu time            0.54 seconds
```

```
22
23    *** Part A - One-Way ANOVA ;
24
25   PROC ANOVA DATA = math;
26    class group;
27    MODEL m_score age = group;
28    means group / snk;
29    TITLE 'REGRESSION of m_score and age on group';
30   RUN;

31   QUIT;

NOTE: PROCEDURE ANOVA used (Total process time):
      real time           5.91 seconds
      cpu time            0.78 seconds


32
33    *** Part B - look for homogeneity of score and age among the 3 groups ;
34
35   PROC GLM DATA = math;
36    class group;
37    MODEL m_score = age group age*group;
38    TITLE 'PROC GLM Homogeneity ';
39   RUN;

40   QUIT;

NOTE: PROCEDURE GLM used (Total process time):
      real time           0.67 seconds
      cpu time            0.12 seconds


41
42    *** Part C - analysis of covariance;
43
44   PROC GLM DATA = math;
45    class group;
46    MODEL m_score = age group /SS3 ;
47    LSMEANS GROUP / PDIFF;
48    TITLE 'Analysis of Covariance';
49   RUN;

50   QUIT;

NOTE: PROCEDURE GLM used (Total process time):
      real time           1.10 seconds
      cpu time            0.34 seconds
```

SAS Output

*** Part A - One-Way ANOVA ;

REGRESSION of m_score and age on group

The ANOVA Procedure

Class Level Information		
Class	Levels	Values
GROUP	3	A B C

Number of Observations Read	18
Number of Observations Used	18

REGRESSION of m_score and age on group

The ANOVA Procedure

Dependent Variable: M_SCORE

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	558.111111	279.055556	3.75	0.0479
Error	15	1116.833333	74.455556		
Corrected Total	17	1674.944444			

R-Square	Coeff Var	Root MSE	M_SCORE Mean
0.333212	10.26555	8.628763	84.05556

Source	DF	Anova SS	Mean Square	F Value	Pr > F
GROUP	2	558.1111111	279.0555556	3.75	0.0479

For M_SCORE (Math score) F= 3.75 and p=0.0479. So there is a significant result for math score.

Student-Newman-Keuls Test for M_SCORE

Means with the same letter are not significantly different.

SNK Grouping	Mean	N	GROUP
A	91.667	6	C
A			
B A	82.000	6	B
B			
B	78.500	6	A

SNK shows no differerence in Math Score between Group A and Group B or between Group B and Group C. However there is a difference between Group A and Group C.

Dependent Variable: AGE

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	26.33333333	13.16666667	3.52	0.0559
Error	15	56.16666667	3.74444444		
Corrected Total	17	82.50000000			

R-Square	Coeff Var	Root MSE	AGE Mean
0.319192	13.04533	1.935057	14.83333

Source	DF	Anova SS	Mean Square	F Value	Pr > F
GROUP	2	26.33333333	13.16666667	3.52	0.0559

For AGE F= 3.52 and p=0.0559. So there is not a significant result for age at the .05 level (although it would be significant at the 0.1 level).

Student-Newman-Keuls Test for AGE

Means with the same letter are not significantly different.

SNK Grouping	Mean	N	GROUP
A	16.500	6	C
A			
A	14.333	6	B
A			
A	13.667	6	A

SNK shows no differerence in Age between between any of the Groups at .05 level.

*** Part B - Homogeneity of relationship among 3 groups ;

Source	DF	Type III SS	Mean Square	F Value	Pr > F
AGE	1	768.8994431	768.8994431	39.76	<.0001
GROUP	2	79.2386314	39.6193157	2.05	0.1716
AGE*GROUP	2	77.0383234	38.5191617	1.99	0.1790

In the test for significant interaction between AGE and GROUP, the F-value=1.99 and p-value=0.1790. So there is no significant interaction and we can do the analysis of covariance.

*** Part C - Analysis of Covariance ;

Dependent Variable: M_SCORE

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	1365.867788	455.289263	20.62	<.0001
Error	14	309.076657	22.076904		
Corrected Total	17	1674.944444			

R-Square	Coeff Var	Root MSE	M_SCORE Mean
0.815471	5.589882	4.698607	84.05556

Source	DF	Type III SS	Mean Square	F Value	Pr > F
AGE	1	807.7566766	807.7566766	36.59	<.0001
GROUP	2	12.3240166	6.1620083	0.28	0.7606

For Age, F=36.59 and P <.0001, so the result for Age is very significant. ie Math Score is affected by age.

For Group, F=0.28 and P =0.7606, so the result for Group is not significant. ie Math Score is not affected by group.

Problem 2: Problem 2, 9.2 p. 316

SAS Program

```
* Problem 1 from Lesson 12, 9.1 p. 315 ;
* Topic: Multiple Regression;

* Create dataset;

data tomato;
input yield light water @@;
datalines;
 12 1  1  20 2  2
 9  1  1  16 2  2
 8  1  1  16 2  2
 13 1  2  18 3  1
 15 1  2  25 3  1
 14 1  2  20 3  1
 16 2  1  25 3  2
 14 2  1  27 3  2
 12 2  1  29 3  2
 ;
 run;

 proc print data=tomato;
 title 'tomato data';
 run;

 data tomato2;
 set tomato;
 light_int = light * 5;
 run;

 proc print data=tomato2;
 title 'tomato2 data';
 run;

PROC REG DATA = tomato2;
 MODEL yield = light_int water;
 TITLE 'REGRESSION ON TOMATO DATA';
RUN;
QUIT;


* Problem 2 , 9.2 p. 316 ;
* Topic: Re-do 9.1 Multiple Regression using Dummy Variables;

 data tomato3;
 set tomato;

 if light = 1 then L5=1; else L5 = 0; /* Reference- will not be used in model */
 if light = 2 then L10=1; else L10 = 0;
 if light = 3 then L15=1; else L15 = 0;

 if water = 1 then W1=1; else W1 = 0; /* Reference- will not be used in model */
```

```
 if water = 2 then W2=1; else W2 = 0;
 run;

 proc print data=tomato3;
 title 'tomato3 data';
 run;

PROC REG DATA = tomato3;
 MODEL yield = L10 L15 W2;
 TITLE 'REGRESSION ON TOMATO DATA using dummy variables';
RUN;
QUIT;
```

SAS Log

```
51
52
53   * Problem 1 from Lesson 12, 9.1 p. 315 ;
54   * Topic: Multiple Regression;
55
56   * Create dataset;
57
58   data tomato;
59   input yield light water @@;
60   datalines;

NOTE: SAS went to a new line when INPUT statement reached past the end of a line.
NOTE: The data set WORK.TOMATO has 18 observations and 3 variables.
NOTE: DATA statement used (Total process time):
      real time           0.00 seconds
      cpu time            0.01 seconds


70    ;
71    run;
72
73    proc print data=tomato;
74    title 'tomato data';
75    run;

NOTE: There were 18 observations read from the data set WORK.TOMATO.
NOTE: PROCEDURE PRINT used (Total process time):
      real time           0.02 seconds
      cpu time            0.03 seconds


76
77    data tomato2;
78    set tomato;
79    light_int = light * 5;
80    run;

NOTE: There were 18 observations read from the data set WORK.TOMATO.
NOTE: The data set WORK.TOMATO2 has 18 observations and 4 variables.
NOTE: DATA statement used (Total process time):
      real time           0.02 seconds
      cpu time            0.01 seconds


81
82    proc print data=tomato2;
83    title 'tomato2 data';
84    run;

NOTE: There were 18 observations read from the data set WORK.TOMATO2.
NOTE: PROCEDURE PRINT used (Total process time):
      real time           0.02 seconds
      cpu time            0.03 seconds
```

```
85
86   PROC REG DATA = tomato2;
87    MODEL yield = light_int water;
88    TITLE 'REGRESSION ON TOMATO DATA';
89   RUN;

90   QUIT;

NOTE: PROCEDURE REG used (Total process time):
      real time           1.85 seconds
      cpu time            0.40 seconds

91
92
93   * Problem 2 , 9.2 p. 316 ;
94   * Topic: Re-do 9.1 Multiple Regression using Dummy Variables;
95
96    data tomato3;
97    set tomato;
98
99    if light = 1 then L5=1; else L5 = 0; /* Reference- will not be used in model
*/
100   if light = 2 then L10=1; else L10 = 0;
101   if light = 3 then L15=1; else L15 = 0;
102
103   if water = 1 then W1=1; else W1 = 0; /* Reference- will not be used in model
*/
104   if water = 2 then W2=1; else W2 = 0;
105   run;

NOTE: There were 18 observations read from the data set WORK.TOMATO.
NOTE: The data set WORK.TOMATO3 has 18 observations and 8 variables.
NOTE: DATA statement used (Total process time):
      real time           0.01 seconds
      cpu time            0.00 seconds


106
107   proc print data=tomato3;
108   title 'tomato3 data';
109   run;

NOTE: There were 18 observations read from the data set WORK.TOMATO3.
NOTE: PROCEDURE PRINT used (Total process time):
      real time           0.04 seconds
      cpu time            0.03 seconds


110
111  PROC REG DATA = tomato3;
112   MODEL yield = L10 L15 W2;
113   TITLE 'REGRESSION ON TOMATO DATA using dummy variables';
114  RUN;

115  QUIT;

NOTE: PROCEDURE REG used (Total process time):
      real time           1.39 seconds
      cpu time            0.40 seconds
```

SAS Output

[Note: Here is my original result from Problem 9.1

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-1.83333	2.29689	-0.80	0.4372
light_int	1	1.21667	0.14066	8.65	<.0001
water	1	4.55556	1.14845	3.97	0.0012

The original equation from Problem 9.1 is:

yield = -1.83333 + (1.21667 * light_int) + (4.55556 * water)]

REGRESSION ON TOMATO DATA using dummy variables

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	9.55556	1.04485	9.15	<.0001
L10	1	3.83333	1.27967	3.00	0.0096
L15	1	12.16667	1.27967	9.51	<.0001
W2	1	4.55556	1.04485	4.36	0.0007

New equation:

```
yield = 9.55556 + (3.83333 * L10) + (12.16667 * L15)
                + (4.55556 * water)
```

Note that the parameter estimate for the water variable stays the same using both methods. But the intercept and light parameters changed.