Alexandra Chang

Homework 6

1. The Viscosity Experiment

The viscosity experiment is described on page 695 where there are three treatment factors: A = sample, B = aliquot, C = subaliquot. Note that B is nested in A and C is nested in B. For this three-way nested structure, you can run `proc glm` as follows:

```
proc glm;
class A B C;
model y=A B(A) C(B A);
random A B(A) C(B A)/test;
run;
```

(a) [3 pts] Write down a statistical model for this experiment that involves three nested factors which all have random effects. Use appropriate notations.

(b) [3 pts] For testing $H_0 : \sigma_A^2 = 0$ vs $H_1 : \sigma_A^2 > 0$, what is the value of the test statistic? What is the p-value? Do you reject H_0 at the significance level $\alpha = 0.05$?

(c) [3 pts] For testing $H_0 : \sigma_{B(A)}^2 = 0$ vs $H_1 : \sigma_{B(A)}^2 > 0$, what is the value of the test statistic? What is the p-value? Do you reject H_0 at the significance level $\alpha = 0.05$?

2. The Cigarette Experiment

The cigarette experiment described on page 760 was run to determine the factors that affect the length of time that a cigarette will burn. There were three factors of interest:

- "Tar" (factor A) at two levels, "regular" and "ultra-light,"
- "Brand" (factor B) at two levels, "name brand" and "generic brand" (coded 1 and 2),
- "Age" (factor C) at three levels, "fresh," "24-hour air exposure," "48-hour air exposure."

The cigarettes were to be burned in whole plots of size six. This was to help with the difficulty of recording burning times and to help keep the amount of smoke in the room at a reasonable level. There were ten whole plots, and these were assigned at random to the tar levels so that each tar level was assigned five whole plots. The six split plots (time slots) in each whole plot were assigned at random to the six brand/age treatment combinations. Marks were made across the seam of each cigarette at a given distance apart. Each cigarette was lit at the beginning of its allotted time slot, and the time taken to burn between the two marks was recorded. The data can be downloaded at `https://www.stat.purdue.edu/~zhanghao/STAT514/Data/cigarettet.txt`. Note that factor D is created to denote the combination of factors B and C.

Answer the following questions by attaching the SAS code and relevant SAS output:

(a) [3 pts] Ignore Factors B and C and only consider factor A (assigned to whole plots) and factor D (assigned to the split-plots). Write SAS code for the analysis of this split-plot design and test the hypothesis that factor A has no effects on the burning time. State the p-value and your conclusion.

(b) [3 pts] Now the experimenters like to investigate the interaction effects of factors B and C. Run the analysis for a model that includes all two-way interaction terms and the three-way interaction term. State your conclusion about the interaction effects of factors B and C.

(c) [3 pts] Provide 95% simultaneous confidence intervals for all pairwise comparison of main effects of factor B using Tukey's method.

1. The Viscosity Experiment

The viscosity experiment is described on page 695 where there are three treatment factors: A = sample, B = aliquot, C = subaliquot. Note that B is nested in A and C is nested in B. For this three-way nested structure, you can run `proc glm` as follows:

```
proc glm;
class A B C;
model y=A B(A) C(B A);
random A B(A) C(B A)/test;
run;
```

Table 18.10 Viscosity determinations for the viscosity experiment

Sample A	Aliquot B	Subaliquot 1 C		Subaliquot 2	
		Part 1	Part 2	Part 1	Part 2
1	1	59.8	59.4	58.2	63.5
	2	66.6	63.9	61.8	62.0
	3	64.9	68.8	66.3	63.5
	4	62.7	62.2	62.9	62.8
	5	59.5	61.0	54.6	61.5
	6	69.0	69.0	60.6	61.8
	7	64.5	66.8	60.2	57.4
	8	61.6	56.6	64.5	62.3
	9	64.5	61.3	72.7	72.4
	10	65.2	63.9	60.8	61.2
2	1	59.8	61.2	60.0	65.0
	2	65.0	65.8	64.5	64.5
	3	65.0	65.2	65.5	63.5
	4	62.5	61.9	60.9	61.5
	5	59.8	60.9	56.0	57.2
	6	68.8	69.0	62.5	62.0
	7	65.2	65.6	61.0	59.3
	8	59.6	58.5	62.3	61.5
	9	61.0	64.0	73.0	71.7
	10	65.0	64.0	62.0	63.0

Source Johnson and Leone (1977). Copyright © 1977 Johnson and Leone. Reprinted with permission

```
data viscosity;
  do A = 1 to 2;                /* Sample */
    do B = 1 to 10;             /* Aliquot nested in A */
      do C = 1 to 2;            /* Subaliquot nested in B(A) */
        do part = 1 to 2;       /* Two replicate measurements */
          input y @@;          /* read into response y */
          output;
        end;
      end;
    end;
  end;
  drop part;
  lines;
  /* Sample A=1 */
  59.8 59.4 58.2 63.5
  66.6 63.9 61.8 62.0
  64.9 68.8 66.3 63.5
  62.7 62.2 62.9 62.8
  59.5 61.0 54.6 61.5
  69.0 69.0 60.6 61.8
  64.5 66.8 60.2 57.4
  61.6 56.6 64.5 62.3
  64.5 61.3 72.7 72.4
  65.2 63.9 60.8 61.2

  /* Sample A=2 */
  59.8 61.2 60.0 65.0
  65.0 65.8 64.5 64.5
  65.0 65.2 65.5 63.5
  62.5 61.9 60.9 61.5
  59.8 60.9 56.0 57.2
  68.8 69.0 62.5 62.0
  65.2 65.6 61.0 59.3
  59.6 58.5 62.3 61.5
  61.0 64.0 73.0 71.7
  65.0 64.0 62.0 63.0
;
run;
```

```
proc glm;
  class A B C;
  model y = A B(A) C(B A);
  random A B(A) C(B A)/test;
run;
```

The GLM Procedure

Class Level Information		
Class	**Levels**	**Values**
A	2	1 2
B	10	1 2 3 4 5 6 7 8 9 10
C	2	1 2

Number of Observations Read	80
Number of Observations Used	80

The GLM Procedure

Dependent Variable: y

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	39	958.279875	24.571279	9.01	<.0001
Error	40	109.115000	2.727875		
Corrected Total	79	1067.394875			

R-Square	Coeff Var	Root MSE	y Mean
0.897774	2.618047	1.651628	63.08625

Source	DF	Type I SS	Mean Square	F Value	Pr > F
A	1	0.0781250	0.0781250	0.03	0.8665
B(A)	18	503.4742500	27.9707917	10.25	<.0001
C(A*B)	20	454.7275000	22.7363750	8.33	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
A	1	0.0781250	0.0781250	0.03	0.8665
B(A)	18	503.4742500	27.9707917	10.25	<.0001
C(A*B)	20	454.7275000	22.7363750	8.33	<.0001

The GLM Procedure

Source	Type III Expected Mean Square
A	Var(Error) + 2 Var(C(A*B)) + 4 Var(B(A)) + 40 Var(A)
B(A)	Var(Error) + 2 Var(C(A*B)) + 4 Var(B(A))
C(A*B)	Var(Error) + 2 Var(C(A*B))

The GLM Procedure

Tests of Hypotheses for Random Model Analysis of Variance

Dependent Variable: y

Source	DF	Type III SS	Mean Square	F Value	Pr > F
A	1	0.078125	0.078125	0.00	0.9584
Error: MS(B(A))	18	503.474250	27.970792		

Source	DF	Type III SS	Mean Square	F Value	Pr > F
B(A)	18	503.474250	27.970792	1.23	0.3251
Error: MS(C(A*B))	20	454.727500	22.736375		

Source	DF	Type III SS	Mean Square	F Value	Pr > F
C(A*B)	20	454.727500	22.736375	8.33	<.0001
Error: MS(Error)	40	109.115000	2.727875		

(a) [3 pts] Write down a statistical model for this experiment that involves three nested factors which all have random effects. Use appropriate notations.

Let $i=1,2$ index Sample (A)
$j=1,..,10$ index Aliquot nested in sample ($B_{j(i)}$),
$k=1,2$ index Subaliquot nested in Aliquot ($C_{k(ij)}$),
$l=1,2$ index the two replicate parts.

A full random-effects model is: $\boxed{Y_{ijkl} = \mu + A_i + B_{j(i)} + C_{k(ij)} + \varepsilon_{l(ijk)}}$,
with
$A_i \sim N(0,\sigma_A^2)$, $B_{j(i)} \sim N(0,\sigma^2_{B(A)})$, $C_{k(ij)} \sim N(0,\sigma^2_{C(BA)})$, $\varepsilon_{l(ijk)} \sim N(0,\sigma^2)$,
all independent.

(b) [3 pts] For testing $H_0 : \sigma_A^2 = 0$ vs $H_1 : \sigma_A^2 > 0$, what is the value of the test statistic? What is the p-value? Do you reject H_0 at the significance level $\alpha = 0.05$?

Source	DF	Type III SS	Mean Square	F Value	Pr > F
A	1	0.078125	0.078125	0.00	0.9584
Error: MS(B(A))	18	503.474250	27.970792		

$\boxed{F = \frac{MSA}{\widehat{MS\ Error}}} = \frac{0.07813}{27.9708} \approx 0.00$

Since $p=0.9584 > 0.05$, we fail to reject Ho.
There is no evidence at the 5% level that σ_A^2 differs from zero

(c) [3 pts] For testing $H_0 : \sigma^2_{B(A)} = 0$ vs $H_1 : \sigma^2_{B(A)} > 0$, what is the value of the test statistic? What is the p-value? Do you reject H_0 at the significance level $\alpha = 0.05$?

Source	DF	Type III SS	Mean Square	F Value	Pr > F
B(A)	18	503.474250	27.970792	1.23	0.3251
Error: MS(C(A*B))	20	454.727500	22.736375		

$\boxed{F = \frac{MS_{B(A)}}{\widehat{MS\ Error}}} = \frac{27.9708}{22.7364} \approx 1.23$

Since $p=0.3251 > 0.05$, we fail to reject Ho.
There is no evidence at the 5% level that $\sigma^2_{B(A)}$ differs from zero.

2. The Cigarette Experiment

The cigarette experiment described on page 760 was run to determine the factors that affect the length of time that a cigarette will burn. There were three factors of interest:

- "Tar" (factor A) at two levels, "regular" and "ultra-light,"
- "Brand" (factor B) at two levels, "name brand" and "generic brand" (coded 1 and 2),
- "Age" (factor C) at three levels, "fresh," "24-hour air exposure," "48-hour air exposure."

The cigarettes were to be burned in whole plots of size six. This was to help with the difficulty of recording burning times and to help keep the amount of smoke in the room at a reasonable level. There were ten whole plots, and these were assigned at random to the tar levels so that each tar level was assigned five whole plots. The six split plots (time slots) in each whole plot were assigned at random to the six brand/age treatment combinations. Marks were made across the seam of each cigarette at a given distance apart. Each cigarette was lit at the beginning of its allotted time slot, and the time taken to burn between the two marks was recorded. The data can be downloaded at `https://www.stat.purdue.edu/~zhanghao/STAT514/Data/cigarettet.txt`. Note that factor D is created to denote the combination of factors B and C.

Answer the following questions by attaching the SAS code and relevant SAS output:

Table 19.34 Burning times for the cigarette experiment

Whole plot (Time)	*A* (Tar)	Levels of *BC* (Burning times in seconds)					
1	1	22 (301)	11 (326)	23 (260)	13 (290)	12 (312)	21 (292)
2	2	11 (329)	12 (331)	13 (285)	21 (306)	22 (258)	23 (276)
3	2	22 (290)	11 (380)	12 (335)	13 (309)	23 (243)	21 (334)
4	2	11 (321)	21 (337)	23 (275)	12 (316)	13 (307)	22 (250)
5	2	22 (308)	11 (345)	21 (307)	23 (288)	13 (321)	12 (330)
6	1	11 (344)	23 (283)	21 (281)	22 (261)	13 (307)	12 (292)
7	1	21 (274)	13 (310)	12 (304)	22 (279)	23 (277)	11 (330)
8	1	13 (302)	12 (325)	22 (301)	11 (338)	23 (270)	21 (297)
9	2	12 (323)	13 (334)	23 (265)	11 (326)	22 (269)	21 (297)
10	1	23 (309)	13 (314)	22 (259)	11 (344)	21 (310)	12 (322)

```
data cigarette;
  input WP A @@;                /* Whole-plot ID and Tar level */
  do SP = 1 to 6;               /* Six Brand×Age observations per whole plot */
    input B C time @@;          /* Brand, Age, and burn time */
    D = 10*B + C;               /* Combined split-plot factor */
    output;
  end;
  drop SP;
  lines;
1 1  2 2 301  1 1 326  2 3 260  1 3 290  1 2 312  2 1 292
2 2  1 1 329  1 2 331  1 3 285  2 1 306  2 2 258  2 3 276
3 2  2 2 290  1 1 380  1 2 335  1 3 309  2 3 243  2 1 334
4 2  1 1 321  2 1 337  2 3 275  1 2 316  1 3 307  2 2 250
5 2  2 2 308  1 1 345  2 1 307  2 3 288  1 3 321  1 2 330
6 1  1 1 344  2 3 283  2 1 281  2 2 261  1 3 307  1 2 292
7 1  2 1 274  1 3 310  1 2 304  2 2 279  2 3 277  1 1 330
8 1  1 3 302  1 2 325  2 2 301  1 1 338  2 3 270  2 1 297
9 2  1 2 323  1 3 334  2 3 265  1 1 326  2 2 269  2 1 297
10 1  2 3 309  1 3 314  2 2 259  1 1 344  2 1 310  1 2 322
;
run;
```

(a) [3 pts] Ignore Factors B and C and only consider factor A (assigned to whole plots) and factor D (assigned to the split-plots). Write SAS code for the analysis of this split-plot design and test the hypothesis that factor A has no effects on the burning time. State the p-value and your conclusion.

```
/* 2(a) Ignore B and C, test H0: no Tar (A) effect using WP(A) as the whole-plot error */
proc mixed data=cigarette;
  class WP A D;
  model time = A D;
  random WP(A);
run;
```

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
A	1	8	1.51	0.2547
D	5	45	20.58	<.0001

Since $p = 0.2547 > 0.05$, we fail to reject H_0.
There is no evidence at the 5% level that Tar (Factor A) affects burning time.

(next page for Q2 (b) (c))

(b) [3 pts] Now the experimenters like to investigate the interaction effects of factors B and C. Run the analysis for a model that includes all two-way interaction terms and the three-way interaction term. State your conclusion about the interaction effects of factors B and C.

```
/* 2(b) Full split-plot model with all two- and three-way interactions */
proc mixed data=cigarette;
  class WP A B C;
  model time = A
               B
               C
               A*B
               A*C
               B*C
               A*B*C;
  random WP(A);
run;
```

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
A	1	8	1.51	0.2547
B	1	40	73.82	<.0001
C	2	40	17.73	<.0001
A*B	1	40	0.42	0.5186
A*C	2	40	1.26	0.2939
B*C	2	40	0.34	0.7117
A*B*C	2	40	2.57	0.0891

Since $p=0.7117 > 0.05$, we fail to reject H_0.
There is no evidence at the 5% level of a Brand × Age interaction.
In other words, the effect of Brand does not depend on cigarette Age, and vice versa.

(c) [3 pts] Provide 95% simultaneous confidence intervals for all pairwise comparison of main effects of factor B using Tukey's method.

```
/* 2(c) 95% simultaneous Tukey confidence intervals for main effects of Brand (B) */
proc mixed data=cigarette;
  class WP A B C;
  model time = A
               B
               C
               A*B
               A*C
               B*C
               A*B*C;
  random WP(A);
  lsmeans B / cl adjust=Tukey;
run;
```

Differences of Least Squares Means														
Effect	B	_B	Estimate	Standard Error	DF	t Value	Pr > |t|	Adjustment	Adj P	Alpha	Lower	Upper	Adj Lower	Adj Upper
B	1	2	36.5000	4.2483	40	8.59	<.0001	Tukey-Kramer	<.0001	0.05	27.9138	45.0862	27.9140	45.0860

- Brand (Factor B) only has 2 levels, so only one pairwise comparison.

Adjusted 95% Tukey's CI:
[27.9140, 45.0860]

Homework 6

1. The Viscosity Experiment

The viscosity experiment is described on page 695 where there are three treatment factors: A = sample, B = aliquot, C = subaliquot. Note that B is nested in A and C is nested in B. For this three-way nested structure, you can run `proc glm` as follows:

```
proc glm;
class A B C;
model y=A B(A) C(B A);
random A B(A) C(B A)/test;
run;
```

(a) [3 pts] Write down a statistical model for this experiment that involves three nested factors which all have random effects. Use appropriate notations.

The statistical model is:

$$y_{ijkl} = \mu + A_i + B_{j(i)} + C_{k(ij)} + \epsilon_{ijkl}$$

where $A_i \sim N(0, \sigma_A^2)$, $B_{j(i)} \sim N(0, \sigma^2_{B(A)})$, and $C_{k(ij)} \sim N(0, \sigma^2_{C(BA)})$.

(b) [3 pts] For testing $H_0 : \sigma_A^2 = 0$ vs $H_1 : \sigma_A^2 > 0$, what is the value of the test statistic? What is the p-value? Do you reject H_0 at the significance level $\alpha = 0.05$?

The F-value is 0.0028 (=0.078125/27.970792) and the p-value is 0.9584. We fail to reject the null hypothesis.

(c) [3 pts] For testing $H_0 : \sigma^2_{B(A)} = 0$ vs $H_1 : \sigma^2_{B(A)} > 0$, what is the value of the test statistic? What is the p-value? Do you reject H_0 at the significance level $\alpha = 0.05$?

The test statistic is 1.23, and the p-value is 0.3251. We fail to reject the null hypothesis.

SAS Code for Solutions:

```
data viscosity;
input sample aliquot @@;
do subali=1 to 2;
    do part=1 to 2;
    input viscosity @@;
    output;
end; end;
lines;
1 1 59.8 59.4 58.2 63.5
1 2 66.6 63.9 61.8 62.0
1 3 64.9 68.8 66.3 63.5
1 4 62.7 62.2 62.9 62.8
1 5 59.5 61.0 54.6 61.5
1 6 69.0 69.0 60.6 61.8
1 7 64.5 66.8 60.2 57.4
1 8 61.6 56.6 64.5 62.3
```

```
1 9 64.5 61.3 72.7 72.4
1 10 65.2 63.9 60.8 61.2
2 1 59.8 61.2 60.0 65.0
2 2 65.0 65.8 64.5 64.5
2 3 65.0 65.2 65.5 63.5
2 4 62.5 61.9 60.9 61.5
2 5 59.8 60.9 56.0 57.2
2 6 68.8 69.0 62.5 62.0
2 7 65.2 65.6 61.0 59.3
2 8 59.6 58.5 62.3 61.5
2 9 61.0 64.0 73.0 71.7
2 10 65.0 64.0 62.0 63.0
;
proc print;
run;

proc glm data=viscosity;
class sample aliquot subali part;
model viscosity=sample aliquot(sample) subali(sample aliquot);
random sample aliquot(sample) subali(sample aliquot)/test;
run;
```

2. The Cigarette Experiment

The cigarette experiment described on page 760 was run to determine the factors that affect the length of time that a cigarette will burn. There were three factors of interest:

- "Tar" (factor A) at two levels, "regular" and "ultra-light,"
- "Brand" (factor B) at two levels, "name brand" and "generic brand" (coded 1 and 2),
- "Age" (factor C) at three levels, "fresh," "24-hour air exposure," "48-hour air exposure."

The cigarettes were to be burned in whole plots of size six. This was to help with the difficulty of recording burning times and to help keep the amount of smoke in the room at a reasonable level. There were ten whole plots, and these were assigned at random to the tar levels so that each tar level was assigned five whole plots. The six split plots (time slots) in each whole plot were assigned at random to the six brand/age treatment combinations. Marks were made across the seam of each cigarette at a given distance apart. Each cigarette was lit at the beginning of its allotted time slot, and the time taken to burn between the two marks was recorded. The data can be downloaded at `https://www.stat.purdue.edu/~zhanghao/STAT514/Data/cigarettet.txt`. Note that factor D is created to denote the combination of factors B and C.

Answer the following questions by attaching the SAS code and relevant SAS output:

(a) [3 pts] Ignore Factors B and C and only consider factor A (assigned to whole plots) and factor D (assigned to the split-plots). Write SAS code for the analysis of this split-plot design and test the hypothesis that factor A has no effects on the burning time. State the p-value and your conclusion.

SAS Code for Solution:

```
data cigarett;
input WP A @@;
do SP=1 to 6;

input B C time @@;
D=10*B+C;
```

```
output;

end;

lines;
1 1 2 2 301 1 1 326 2 3 260 1 3 290 1 2 312 2 1 292
2 2 1 1 329 1 2 331 1 3 285 2 1 306 2 2 258 2 3 276
3 2 2 2 290 1 1 380 1 2 335 1 3 309 2 3 243 2 1 334
4 2 1 1 321 2 1 337 2 3 275 1 2 316 1 3 307 2 2 250
5 2 2 2 308 1 1 345 2 1 307 2 3 288 1 3 321 1 2 330
6 1 1 1 344 2 3 283 2 1 281 2 2 261 1 3 307 1 2 292
7 1 2 1 274 1 3 310 1 2 304 2 2 279 2 3 277 1 1 330
8 1 1 3 302 1 2 325 2 2 301 1 1 338 2 3 270 2 1 297
9 2 1 2 323 1 3 334 2 3 265 1 1 326 2 2 269 2 1 297
10 1 2 3 309 1 3 314 2 2 259 1 1 344 2 1 310 1 2 322
;
proc print;
run;

proc mixed data=cigarett;
class WP A D;
model time=A D A*D;
random WP(A);
run;
```

The p-value is 0.2547. We fail to reject the null hypothesis. No evidence to indicate that the main effects of A are not all equal.

(b) [3 pts] Now the experimenters like to investigate the interaction effects of factors B and C. Run the analysis for a model that includes all two-way interaction terms and the three-way interaction term. State your conclusion about the interaction effects of factors B and C.

SAS Code for Solution:

```
proc mixed data=cigarett;
class WP A B C;
model time=A B C B*C A*B A*C A*B*C;
random WP(A);
lsmeans C/cl pdiff adjust=Tukey;
lsmeans B/cl pdiff adjust=Tukey;
run;
```

The p-value for the BC interaction is 0.7117. We fail to reject the null hypothesis. No significant evidence to suggest BC interactions.

(c) [3 pts] Provide 95% simultaneous confidence intervals for all pairwise comparison of main effects of factor B using Tukey's method.

The simultaneous confidence intervals for all pairwise comparisons for factor B are given below. Since B has only two levels, there is no need to use Tukey's method to adjust.

Effect	Estimate	DF	Standard Error	95% CI
$B_1 - B_2$	36.5000	40	4.2483	$(27.9138, 45.0862)$