****

Lab 11: <u>Randomized Block Design and Nested Design</u>

****

OBJECTIVES:

1. Perform ANOVA for factorial design and randomized block design using PROC MIXED.
2. Nested Design and simple CONTRAST statement.

The objective of an experimental design is to provide the maximum amount of reliable information at the minimum cost. In statistical terms, the reliability of information is measured by the standard error of estimates (that is directly related with the population variance, inversely related to sample size). Properly applied experimental design may effectively reduce the population variance, and/or could structure data collection to reduce the magnitude of the experimental error. Usually data resulting from the implementation of experimental designs are described by linear model and analyzed by the analysis of variance as was introduced last week.

Random block design is one of the simplest and probably the most popular experimental design. In this design the sample of experimental units is divided into groups or blocks and then treatments are randomly assigned to units in each block. In some experiments blocks may be of sufficient size to allow several units to be assigned to each treatment in a block. Such replication of treatments is referred to as randomized blocks with sampling. The linear model for data from such an experiment is

$$\mathbf{y_{ijk} = \mu + \tau_i + \beta_j + \tau\beta_{ij} + \varepsilon_{ijk}} \ (\boldsymbol{i} = 1, 2, \ldots, t;\ \boldsymbol{j} = 1, 2, \ldots, b;\ \boldsymbol{k} = 1, 2, \ldots, k)$$

Where $\mathbf{y_{ijk}}$ is the observed value of the response variable in the $\boldsymbol{k}^{th}$ replicate of the treatment ***i*** in block ***j***; $\boldsymbol{\mu}$ is the overall mean; $\boldsymbol{\tau_i}$ is the fixed effect of treatment ***i;*** $\boldsymbol{\beta_j}$ is the effect of block ***j***, a random variable with mean zero and variance σ^2_{β}; $\boldsymbol{\tau\beta_{ij}}$ is the experimental error, a random variable with mean zero and variance $\sigma^2_{\tau\beta}$; and $\boldsymbol{\varepsilon_{ijk}}$ is the sampling error, which is the measure of variation among units treated alike within a block, a random variable with mean zero and variance σ^2.

As you might be aware that in the above model, if the block effect is fixed, the interaction is also fixed so that it becomes a two-factor factorial design and F ratios for all the tests using the sampling error in the denominator.

Nested design refers to some experimental situations where experimental units may contain sampling units, which may, in turn, contain sample subunits. Since the design describes subsamplies nested within sample or experimental units, it is called nested or hierarchical design. The linear model for nest design is

$$\mathbf{y_{ijk} = \mu + \alpha_i + \beta_{j(i)} + \varepsilon_{k(ij)}} \ (\boldsymbol{i} = 1, 2, \ldots, a;\ \boldsymbol{j} = 1, 2, \ldots, b;\ \boldsymbol{k} = 1, 2, \ldots, n)$$

where $\mathbf{y_{ijk}}$ is the k^{th} observed value for level i of factor A and level j of factor B which is nested in the i^{th} level of factor A; $\boldsymbol{\mu}$ is the overall mean; α_i is the effect of the i^{th} level of factor A; $\boldsymbol{\beta_{j(i)}}$ is the effect of level j of factor B nested in the i^{th} level of factor A; and $\boldsymbol{\varepsilon_{k(ij)}}$ variation among sampled units and is the random error.

The subscrip $j(i)$ is used to denoted that different j subscripts occur within each value of i; that is, they are "nested" in i. Likewise, the k subscript is "nested" in groups identified by the combined ij subscript (Keep in mind that it is not interaction term).

In this week's lab, **PROC MIXED** will be used to analyze the data from Random Block Design with Sampling and Nested Design. Also, simple **CONTRAST** will be introduced as well.

LABORATORY INSTRUCTIONS

Housekeeping Statements

```
dm 'log; clear; output; clear';
options nodate nocenter pageno = 1 ls=78 ps=53;
title1 'EXST7014 lab 11, Name, Section#';
ods rtf  file = 'c:/temp/lab11.rtf';
ods html file = 'c:/temp/lab11.html';
```

Part I.

Nested Design with Subsampling:

The data set of nested design is from Statistical Methods II by Roger Cue of Department of Animal Science from McGill University. This experiment tests the effects of different treatments on the growth of apples. There are three types of treatments which are applied to 12 randomly selected apple trees (four trees per treatment). As you can see **the trees are nested in the treatment.** At the end of the experiment, six apples were randomly selected from each apple tree and the weight of each apple was recorded. **PROC MIXED** will be performed to test overall treatment effect, and simple **CONTRAST** will be used to test the difference between the treatments.

```
Data nested;
Title1 'ANOVA with nested design';
Input treatment tree apple weight;
Cards;
   1  1  1  313.063
   1  1  2  329.132
   1  1  3  334.278
   1  1  4  330.088
   1  1  5  334.987
   1  1  6  325.075
   1  2  1  333.936
   1  2  2  326.155
   1  2  3  352.854
   1  2  4  350.791
   1  2  5  318.560
   1  2  6  323.473
   1  3  1  345.494
   1  3  2  349.296
   1  3  3  339.190
   1  3  4  338.942
```

```
 1  3  5  331.370
 1  3  6  339.097
 1  4  1  340.840
 1  4  2  336.798
 1  4  3  313.810
 1  4  4  333.880
 1  4  5  343.068
 1  4  6  319.171
 2  1  1  349.271
 2  1  2  336.695
 2  1  3  352.797
 2  1  4  348.486
 2  1  5  352.077
 2  1  6  341.423
 2  2  1  356.880
 2  2  2  356.256
 2  2  3  364.950
 2  2  4  360.570
 2  2  5  362.104
 2  2  6  371.829
 2  3  1  324.161
 2  3  2  340.130
 2  3  3  334.580
 2  3  4  342.813
 2  3  5  327.415
 2  3  6  333.571
 2  4  1  338.742
 2  4  2  340.348
 2  4  3  362.837
 2  4  4  340.782
 2  4  5  348.730
 2  4  6  325.444
 3  1  1  387.868
 3  1  2  372.807
 3  1  3  380.505
 3  1  4  391.804
 3  1  5  388.935
 3  1  6  361.860
 3  2  1  377.948
 3  2  2  380.033
 3  2  3  361.913
 3  2  4  363.098
 3  2  5  365.375
 3  2  6  382.121
 3  3  1  363.583
 3  3  2  387.727
 3  3  3  373.021
 3  3  4  362.931
 3  3  5  378.928
 3  3  6  364.442
 3  4  1  374.851
 3  4  2  361.291
 3  4  3  377.389
 3  4  4  366.722
 3  4  5  374.187
 3  4  6  380.383
 ;
```

```
Proc mixed data=nested;
Class treatment tree;
Model weight = treatment;
Random tree(treatment);
Lsmeans treatment / adjust=tukey pdiff;
Contrast ' treatment 1 - (treatment 2+3)/2' treatment 1 -.5 -.5;
Estimate ' treatment 1 - (treatment 2+3)/2' treatment 1 -.5 -.5;
Run;
```

CONTRAST: specify the contrast between treatment 1 and the average of treatments 2 and 3. Note that the Sums of Squares, Mean Squares and F-ratio are not the same as those for the comparisons between treatments.
ESTIMATE: returns the estimate of the difference and its standard error. It is a standard t-test.

Part II.

Random Block Design with Sampling:

The data set of random block design comes from your textbook (Chapter10, Problem 8, Table 10.30 from Dr. Geaghan's website). An experiment is conducted to test the effectiveness of three types of gasoline additives for boosting gas mileage on a specific type of car. Three randomly selected cars are purchased for the experiment. Each additive is tested four (randomly ordered) times on each of the three cars. In this lab we will perform both a factorial (**Car as fixed effect**) and randomized block design ANOVA (**Car as a random block**) using **PROC MIXED**. The link of the data is http://www.stat.lsu.edu/exstweb/statlab/datasets/fwdata97/FW10P08.txt

The variables in the dataset are:

- **Additive**: Type of additive;
- **Car**: ID of cars;
- **Run**: ID of test on each car;
- **Mpg**: Mileage per gallon, the dependent variable.

```
Data gas;
Title2 'CHAPTER 10, PROBLEM 8, GASOLINE ADDITIVES';
Input OBS ADDITIVE RUN MPG CAR $;
Cards;
1 1 1 19.5 A
2 1 1 21.1 B
3 1 1 21.1 D
4 1 2 20.3 A
5 1 2 21.2 B
6 1 2 22.8 D
7 1 3 19.4 A
8 1 3 20.4 B
9 1 3 21.7 D
10 1 4 21.3 A
11 1 4 20.6 B
12 1 4 21.6 D
13 2 1 18.0 A
14 2 1 21.8 B
15 2 1 20.5 D
16 2 2 17.8 A
17 2 2 21.2 B
18 2 2 20.1 D
19 2 3 17.8 A
20 2 3 22.7 B
```

```
21 2 3 21.2 D
22 2 4 15.8 A
23 2 4 23.1 B
24 2 4 20.9 D
25 3 1 16.8 A
26 3 1 21.5 B
27 3 1 18.5 D
28 3 2 17.0 A
29 3 2 20.2 B
30 3 2 19.6 D
31 3 3 16.8 A
32 3 3 18.6 B
33 3 3 20.3 D
34 3 4 15.5 A
35 3 4 20.3 B
36 3 4 18.8 D
;
Proc boxplot data=gas;
Plot mpg*additive;
Run;
Proc mixed data=gas;
Title2 "Factorial design";
Class additive car;
Model mpg = additive car car*additive / ddfm=satterth outp=outdatacrd;
Run;
Proc univariate data=outdatacrd normal plot;
Var resid;
run;
Proc plot data=outdatacrd;
Plot resid*pred;
run;
Proc mixed data=gas;
Title2 "Randomized block design";
Class additive car;
Model mpg = additive / ddfm=satterth outp=outdatarbd;
Random car car*additive;
run;
Proc univariate data=outdatarbd normal plot;
Var resid;
Run;
Proc plot data=outdatarbd;
Plot resid*pred;
Run;
ods html close;
ods rtf close;
```

LAB ASSIGNMENT

Your assignment is to perform necessary analysis using SAS and answer the following questions (Please do not print all the output. Only print the graphs and tables that you think are relevant to your answers).

Part I. Answer the following questions for data set of nested design:

1. According to the model, do different treatments have significant effect on the apple weight?
2. Using the contrast to test the hypothesis that the difference between treatment 1 and the average of treatment 2 and treatment 3 is equal to zero. What is you conclusion?

<u>Part II.</u> Answer the following questions for data set of random block design:

1. Take the variable Car as a random block. Do different additives have significantly different effect on boosting gas mileage?

2. Take the variable Car as a fixed effect. Do different additives have significantly different effect on boosting gas mileage?

3. Are your answers for question 1 and 2 consistent? If not, why (that is, why the test statistic and p-value are different)?

4. Which model is more appropriate? Why?

5. What is the additional assumption we must have for the RBD model? Can we test it?

Remember to attach your SAS log with your lab report.

*

The linear model and symbolic notation of source ANOVA table for the nested design with subsampling.

The linear model:

$\mathbf{y_{ijk} = \mu + \alpha_i + \beta_{j(i)} + \varepsilon_{k(ij)}}$ **(*i* = 1, 2, …, a; *j* = 1, 2, …, b; *k* =1, 2, …, n)**

Source	DF	EMS
Treatment	a-1	$\sigma^2 + n\,\sigma^2_{\beta} + nb(\Sigma\,\alpha_i^2/(a-1))$
Sampling (Treatment)	a(b-1)	$\sigma^2 + n\,\sigma^2_{\beta}$
Subsampling	ab(n-1)	σ^2

Effect	F test
Treatment	$F = \dfrac{\sigma 2 + n\,\sigma 2\boldsymbol{\beta} + nb(\Sigma\,\alpha i2/(a-1))}{\sigma 2 + n\,\sigma 2\boldsymbol{\beta}}$

The linear model and the symbolic notation of source ANOVA table for the two- factor factorial design.

The linear model:

$\mathbf{y_{ijk} = \mu + \tau_i + \tau_j + \tau\tau_{ij} + \varepsilon_{ijk}}$ **(*i* = 1, 2, …,t; *j* = 1, 2, …, b; *k* =1, 2, …, k)**

Source	DF	EMS
Treatment 1	$t_1 - 1$	$\sigma^2 + n\,t_2(\Sigma\,\tau_i^2/(t_1 - 1))$
Treatment 2	$t_2 - 1$	$\sigma^2 + n\,t_1(\Sigma\,\tau_j^2/(t_2 - 1))$
Treatment 1 *Treatment 2	$(t_1 - 1)(t_2 - 1)$	$\sigma^2 + n\,(\Sigma\,\tau_i\tau_j^2/(t_1 - 1)(t_2 - 1))$
Error	$t_1 * t_2 * (n - 1)$	σ^2
Total	$t_1\,t_2 n - 1$	

Effect	F test
Treatment 1	$F = \dfrac{\sigma 2 + n\,t2(\Sigma\,\tau i2/(t1-1))}{\sigma 2}$
Treatment 2	$F = \dfrac{\sigma 2 + n\,t1(\Sigma\,\tau j2/(t2-1))}{\sigma 2}$
Treatment 1 *Treatment 2	$F = \dfrac{\sigma 2 + n\,(\Sigma\,\tau i\tau j2/(t1-1)(t2-1))}{\sigma 2}$

The linear model and the symbolic notation of source ANOVA table for the randomized block design with sampling.

The linear model:

$$\mathbf{y_{ijk} = \mu + \tau_i + \beta_j + \tau\beta i_j + \varepsilon_{ijk}} \quad (i = 1, 2, \ldots, t;\ j = 1, 2, \ldots, b;\ k = 1, 2, \ldots, k)$$

Source	**DF**	**EMS**
Treatment	t – 1	$\sigma^2 + n\sigma_{\tau\beta}^2 + nb(\,\Sigma\, \tau_i^2/(\,t-1))$
Block	b – 1	$\sigma^2 + n\sigma_{\tau\beta}^2 + nt\sigma_{\beta}^2$
Treatment*Block (Exp. Error)	(t – 1)(b – 1)	$\sigma^2 + n\sigma_{\tau\beta}^2$
Sampling Error	t *b*(n -1)	σ^2
Total	tbn – 1=35	

Effect	**F test**
Treatment	$F = \dfrac{\sigma 2\ +\ n\sigma\tau\beta 2\ +\ nb(\,\Sigma\,\tau i2/(\,t-1))}{\sigma 2\ +\ n\sigma\tau\beta 2}$
Block	$F = \dfrac{\sigma 2\ +\ n\sigma\tau\beta 2\ +\ nt\sigma\beta 2}{\sigma 2\ +\ n\sigma\tau\beta 2}$
Treatment*Block	$\boldsymbol{F = \dfrac{\sigma 2\ +\ n\sigma\tau\beta 2}{\sigma 2}}$