EXST 7015
Fall 2014
Lab 14: Split-Plot Design with Repeated Measures

OBJECTIVES:

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.

For the last couple of weeks, several types of experimental designs (RBD, Nested Design and Latin Square Design) have been practice. You are getting familiar with how to recognize the experimental design based on the data structure and write the linear model and construct ANOVA source table based on different experimental design.

In this week's lab, another type of experimental design, **Split Plot Design**, will be introduced. Split-plot experiments began in the agricultural industry. Because one factor in the experiment is usually a fertilizer or irrigation method, it can only be applied to large sections of land called whole plots. The factor associated with this is therefore called a whole plot factor (Factor A). Within the whole plot, another factor, such as seed variety, is applied to smaller sections of the land, which are obtained by splitting the larger section of the land into subplots. This factor is therefore referred to as the subplot factor (Factor B). Therefore, the split-plot design has two levels, a whole plot with its own treatment (Factor A) and error term and a subplot with its own treatment (Factor b) and error term. The subplot error is expected to be smaller and have more degrees of freedom, which leads to more powerful subplot tests. The tests of interest could be assigned to the subplot. The linear model of the simplest split plot with balanced design is:

$$Y_{ijk} = \mu + \alpha_i + \gamma_{k(i)} + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk} \ (i = 1, 2, \ldots, a; \ j = 1, 2, \ldots, b; \ k = 1, 2, \ldots, n)$$

Where μ is the overall mean; a_i is the effect of ith level of **Facotr A**; $\gamma_{k(i)}$ is the whole plot error; β_j is the effect of j_{th} level of Factor B; $(\alpha\beta)_{ij}$ in the interaction effect between Factor A and Factor B. ε_{ijk} is the random error with mean zero and variance σ^2, which is the error term used to test β_j and $(\alpha\beta)_{ij}$.

The repeated measures design is similar to split plot. There is usually a "whole plot"(Factor A), which could be CRD, RBD or LSD. Then the repeated measures could be taken over time within the whole plots. If these "repeated measures" are independent, this TIME factor (Factor B) is just cross-classified with Factor A (Treatment).

PROC MIXED will be used to analyze a set of data with repeated measures of Split plot design. **Expected Mean Squares** will be required in PROC MIXED to further help you understand the ANOVA table.

LABORATORY INSTRUCTIONS

Housekeeping Statements
dm 'log; clear; output; clear';
options nodate nocenter pageno = **1** ls=**78** ps=**53**;
title1 'EXST7015 lab 13, Name, Section#';
ods rtf file = 'c:/temp/lab13.rtf';
ods html file = 'c:/temp/lab9.html';

Split Plot Design with Repeated Measures:

An experiment was carried out to study the differences in yield of four alfalfa cultivars. Five replications of these four varieties were organized according to a CRD, and four cuttings were made of each replication over time. The data represents the repeated measurements of yield (tons/acre) of the four cultivars and is analyzed as a split-plot CRD with repeated measures:

    4 levels of main plot A: **Cultivars** 1 – 4
    4 levels of subplot B: **Cut times** 1 – 4 (9/10/74, 6/25/75, 18/5/75, 9/16/75)

To analyze this data, we use PROC MIXED to carry out a standard split-plot analysis.

```
data one;
input rep A_var B_time yield @@;
cards;
1 1 1 2.80191 1 1 2 3.73092 1 1 3 3.09856 1 1 4 2.50965
1 2 1 2.76212 1 2 2 5.40530 1 2 3 3.82431 1 2 4 2.72992
1 3 1 2.29151 1 3 2 3.81140 1 3 3 2.92575 1 3 4 2.39863
1 4 1 2.56631 1 4 2 4.96070 1 4 3 2.81734 1 4 4 2.05752
2 1 1 2.96602 2 1 2 4.43545 2 1 3 3.10607 2 1 4 2.57299
2 2 1 3.09636 2 2 2 3.90683 2 2 3 3.26229 2 2 4 2.58614
2 3 1 2.54027 2 3 2 3.82716 2 3 3 2.86727 2 3 4 2.16287
2 4 1 2.31630 2 4 2 3.96629 2 4 3 2.91461 2 4 4 2.15764
3 1 1 2.43232 3 1 2 4.32311 3 1 3 2.81030 3 1 4 2.07966
3 2 1 3.09917 3 2 2 4.08859 3 2 3 3.13148 3 2 4 2.60316
3 3 1 2.41199 3 3 2 4.08317 3 3 3 3.03906 3 3 4 2.07076
3 4 1 2.65834 3 4 2 3.71856 3 4 3 2.92922 3 4 4 2.15684
4 1 1 2.93509 4 1 2 3.99711 4 1 3 2.77971 4 1 4 2.44033
4 2 1 2.65256 4 2 2 5.42879 4 2 3 2.70891 4 2 4 2.30163
4 3 1 2.30420 4 3 2 3.27852 4 3 3 2.72711 4 3 4 2.04933
4 4 1 2.47877 4 4 2 3.92048 4 4 3 3.06191 4 4 4 2.35822
5 1 1 2.42277 5 1 2 3.85657 5 1 3 3.24914 5 1 4 2.34131
5 2 1 2.63666 5 2 2 3.77458 5 2 3 3.09734 5 2 4 2.30082
5 3 1 2.36941 5 3 2 3.44835 5 3 3 2.50562 5 3 4 2.08980
5 4 1 2.23595 5 4 2 4.02985 5 4 3 2.85279 5 4 4 1.85736
;
```

```
proc mixed data=one method=type3;
  class rep A_var B_time;
  model yield=A_var B_time A_var*B_time/ddfm=satterth;
  random rep rep*A_var;
  LSMEANS A_var B_time/ adjust=tukey pdiff;
run;
```

METHOD = TYPE3: forces PROC MIXED procedure to produce the Expected Mean Squares.

LAB ASSIGNMENT:

1. Write the linear model the split plot design in symbolic notation; clearly explain each term in the model, and the range of the subscripts. Write the null hypotheses that you are testing.

2. Consider **Cultivars** and **Cut_ times** fixed effects; write the source ANOVA table, including degree of freedom. Then carefully exam the table of Expected Mean Squares produced by PROC MIXED to understand the error terms used to F- tests of different null hypothesis.

3. Use PROC MIXED to test the hypothesis of the model in Question 1. Report your results, including your F-values, and degree of freedom, p-value and conclusions