EXST 7005

Fall 2010

Lab #11: Randomized Complete Block Design & Simple Linear Regression

---

Randomized Complete Block Design

A Randomized Complete Block Design (RCBD) is a restricted randomization design in which the experimental units are first sorted into homogeneous groups, called blocks, and the treatments are then assigned at random within the blocks. In a Randomized Complete Block Design, each level of a "treatment" appears once in each block, and each block contains all the treatments. The experimental units are assumed to be homogeneous within each block. Variability coming from the block variable is not of interest to the researcher, so the addition to SSE from variation between the blocks is removed from consideration in the ANOVA table. By doing so, we hope to increase the power of the test. This would allow us to detect smaller differences between treatments.

Example: The researchers want to test the effects of three insecticides which are used to protect seedlings. To get a larger sample, they get 4 plots of land with the same seedlings and apply all the three insecticide on each of the plot. After a period, the researcher measures the number of healthy seedlings (per 100 seedlings) on each plot. The researchers are not interested in the difference of plots. They are only interested in the effects of insecticides.

	Plot 1	Plot 2	Plot 3	Plot 4
Insecticide 1	56	48	66	62
Insecticide 2	83	78	94	93
Insecticide 3	80	72	83	85

```
title1 'Randomized Complete Block Design';

data SEEDS;
title2 'Analysis of the effects of three insecticides';
input SEEDLINGS INSECTICIDE PLOT;
cards;
56 1 1
48 1 2
66 1 3
62 1 4
83 2 1
78 2 2
94 2 3
93 2 4
80 3 1
72 3 2
83 3 3
85 3 4
;
```

We will use PROC MIXED procedure in which the random effect variable (PLOT) is not included in the MODEL statement. Instead, it goes into the RANDOM statement. Since the AVONA results show that the three insecticides have significantly different effects, we use LSMEANS to find how they differ. The CONTRAST statement is also used to see whether the mean of 1 & 2 significantly differs from 3.

```
proc mixed;
title3 'Analysis of variance with PROC MIXED';
class INSECTICIDE PLOT;
model SEEDLINGS=INSECTICIDE/ddfm=satterth outp=OUT1;
random PLOT;
lsmeans INSECTICIDE/adjust=tukey pdiff;
contrast "insecticide 1 & 2 vs 3" INSECTICIDE 1 1 -2;
run;
```

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
Insecticide	2	6	211.38	<.0001

The following table suggests insecticide 1 and 2 are significantly different with each other. So are 2 & 3, 1 & 3.

Differences of Least Squares Means									
Effect	INSECTICIDE	_INSECTICIDE	Estimate	Standard Error	DF	t Value	Pr > |t|	Adjustment	Adj P
INSECTICIDE	1	2	-29.0000	1.4720	6	-19.70	<.0001	Tukey-Kramer	<.0001
INSECTICIDE	1	3	-22.0000	1.4720	6	-14.95	<.0001	Tukey-Kramer	<.0001
INSECTICIDE	2	3	7.0000	1.4720	6	4.76	0.0031	Tukey-Kramer	0.0075

To test the assumptions of normality and homogeneous variances, the following codes are used.

```
proc univariate data=OUT1 normal plot;
var RESID;
run;

proc plot data=OUT1;
plot RESID*PRED;
run;
```

Simple Linear Regression

The purpose of a simple linear regression analysis is to examine the linear relationship between a regressor (or independent variable) and a response (or dependent variable) and then set up a regression model. We can use this model to predict Y according to the value of X for a new observation. Fitting this model with the REG procedure requires only the following MODEL statement, where y is the response variable and x is the regressor variable. For example, you might use regression analysis to find out how well you can predict a child's weight if you know that child's height. The following data are from a study of nineteen children. Height and weight are measured for each child. The SAS code will be:

```
data Class;
input Name $ Height Weight @@;
datalines;
Alfred 69.0 112.5 Alice 56.5 84.0 Barbara 65.3 98.0 Carol 62.8 102.5 Henry 63.5
102.5 James  57.3 83.0 Jane 59.8 84.5 Janet 62.5 112.5 Jeffrey 62.5 84.0 John
59.0 99.5 Joyce 51.3 50.5 Judy 64.3 90.0 Louise 56.3 77.0 Mary  66.5 112.0 Philip
72.0 150.0 Robert 64.8 128.0 Ronald 67.0 133.0 Thomas 57.5 85.0 William 66.5 112.0
;
proc reg data=Class;
model Weight=Height;
output out=OUT1 p=YHAT r=RESID;
run;
```

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	1	7193.24912	7193.24912	57.08	<.0001
Error	17	2142.48772	126.02869		
Corrected Total	18	9335.73684			

The F statistic for the overall model is highly significant (F=57.08, p<0.0001), indicating that the model explains a significant portion of the variation in the data or height has significant influence on weight.

Root MSE	11.22625	**R-Square**	0.7705
Dependent Mean	100.02632	**Adj R-Sq**	0.7570
Coeff Var	11.22330		

The **R-Square** and **Adj R-Square** are two statistics used in assessing the fit of the model. Values close to 1 indicate a better fit. The R-Square of 0.77 indicates that Height accounts for 77% of the variation in Weight.

Parameter Estimates					
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	1	-143.02692	32.27459	-4.43	0.0004
Height	1	3.89903	0.51609	7.55	<.0001

The above table contains the estimates of β_0 and β_1 as well as the t tests results associated with them. The p-values (t=-4.43, p=0.0004 and t=7.55, p<0.0001) indicate that both the intercept and Height parameter estimates are significantly different from zero. From the parameter estimates, the fitted model is

- Weight = -143.0 + 3.9Height

The following command can now be issued to test the homogeneity of variance: