EXST 7005

Fall 2010

Lab #8: Single Factor CRD & One-way ANOVA

Single Factor CRD

In a single factor Completely Randomized Design (CRD), there is only one treatment factor and the levels of the factor are randomly assigned to the experimental units. Each experimental unit receives one and only one experimental treatment. For example, an experiment could be conducted to compare the mean yields of several varieties of wheat. The field plots (experimental units) are randomly assigned one of the varieties of wheat.

One-way ANOVA

The ANOVA (short for "the analysis of variance") is one of the most powerful and general techniques for data analysis. It is applicable to a variety of experimental situations. However, before applying the appropriate ANOVA technique, it is extremely important that you understand the experimental design. The ANOVA applied to a single factor Completely Randomized Design is often called the "one-way ANOVA."

The ANOVA could be viewed as an extension and generalization of the two-sample t-test to the case of two or more groups. When you are comparing three or more means, doing multiple two-sample t-tests would result in a largely increased chance of committing a type I error. The ANOVA is a solution to this problem.

The ANOVA examines the data for evidence of differences in the corresponding population means by looking at the ratio of **among group means variation** to the **within group observation variatio**n. If this ratio is large, there is evidence against the null hypothesis of equal group means. In other words, the ANOVA is actually a test of the equality of population means through the comparison of sample variations.

There are many ways to approach the ANOVA in SAS. We will focus on only two of these, **PROC GLM** and **PROC MIXED**. For simple one-way ANOVA, we can start with only **PROC GLM** (the generalized linear models procedure).

Example

Suppose you are comparing the time to relief of three headache medicines -- brands 1, 2, and 3. The time to relief data is reported in minutes. For this experiment, 15 subjects were randomly placed on one of the three medicines. Which medicine (if any) is the most effective? The data for this example are as follows:

Brand 1	Brand 2	Brand 3
24.5	28.4	26.1
23.5	34.2	28.3
26.4	29.5	24.3
27.1	32.2	26.2
29.9	30.1	27.8

```
data ACHE;
input BRAND RELIEF;
datalines;
1 24.5
1 23.5
1 26.4
1 27.1
1 29.9
2 28.4
2 34.2
2 29.5
2 32.2
2 30.1
3 26.1
3 28.3
3 24.3
3 26.2
3 27.8
;
proc glm data=ACHE;
class BRAND;
model RELIEF=BRAND;
output out=OUT1 p=YHAT R=RESID;
run;
```

The PROC GLM statement will produce the following ANOVA table. This tests the overall model to determine if there is a difference in means between BRANDS. In this case, since the p-value is smaller than 0.05, you can conclude that there is evidence that there is a statistically significant difference in brands.

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	66.7720000	33.3860000	7.14	0.0091
Error	12	56.1280000	4.6773333		
Corrected Total	14	122.9000000			

In the OUTPUT statement, out= OUT1 produces a new dataset OUT1 consisting of two new variables YHAT (predicted Y) and RESID (residual ε =Y-YHAT). You can use the PROC PRINT statement to show the OUT1 dataset:

```
proc print data=OUT1;
run;
```

Obs	BRAND	RELIEF	YHAT	RESID
1	1	24.5	26.28	-1.78
2	1	23.5	26.28	-2.78
3	1	26.4	26.28	0.12
4	1	27.1	26.28	0.82
5	1	29.9	26.28	3.62
6	2	28.4	30.88	-2.48
7	2	34.2	30.88	3.32
8	2	29.5	30.88	-1.38
9	2	32.2	30.88	1.32
10	2	30.1	30.88	-0.78
11	3	26.1	26.54	-0.44
12	3	28.3	26.54	1.76
13	3	24.3	26.54	-2.24
14	3	26.2	26.54	-0.34
15	3	27.8	26.54	1.26

Normality of residuals is an important assumption of ANOVA. You can use **PROC UNIVARIATE** to check normality of residuals:

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

Assignment

A firm wishes to compare four programs for training workers to perform a manual task. Twenty employees are randomly assigned to the training programs (5 in each program). At the end of the training period, a test is conducted to see how quickly trainees can perform the task. The number of times the task is performed per minute is recorded for each trainee, with the following results:
Program 1: 9, 12, 14, 11, 13
Program 2: 10, 6, 9, 9, 10
Program 3: 12, 14, 11, 13, 11
Program 4: 9, 8, 11, 7, 8

1. Identify the treatment and experimental unit in this design.
2. Construct the ANOVA table. Report the F-value, degrees of freedom, p-value.
3. Using $\alpha = .01$, discuss the testing result and make your conclusion.
4. Show your residual values.
5. Check whether the assumption of normality residuals is satisfied.