EXST 7005

Fall 2010

Lab #10: Factorial Experiment Design

A factorial experiment is one in which responses are observed for every combination of factor levels. It allows looking at not only the effect of each main factor but also the effects of interactions between factors on the response variable. Suppose we have two variables A and B and each have two levels, we will get four combination effects of A and B (a1b1, a1b2, a2b1, a2b2). And then we assign these four treatments randomly to experimental units, as is done for the one-way or CRD experiment. It is a 2 x 2 factorial design. For a factorial experiment, we often assume that: 1) there are two or more independently sampled experimental units for each combination of factor levels and also 2) the numbers of experimental units for each combination are equal. The first assumption is to make sure there are replications in consideration of the validity of the experiment; the second assumption is to get a balanced design. A factorial experiment may be thought of as an extension of one-way ANOVA since the same theory and methodology is used.

Example: Suppose that an experimenter is interested in evaluating the relative effectiveness of three drugs in bringing about behavioral changes in two categories of patients: schizophrenics and depressives. What is considered to be a random sample of nine patients belonging to the first category (schizophrenics) is divided at random into three subgroups, with three patients in each subgroup. Each subgroup is then assigned to one of the drug conditions. An analogous procedure is followed for a random sample of nine patients belonging to the second category (depressives). Criterion ratings are made on each patient before and after the administration of the drugs. The numerical entries in the table below represent the differences between the two ratings on each of the patients.

	Drug 1			**Drug 2**			**Drug 3**		
Schizophrenics	8	4	0	10	8	6	8	6	4
Depressives	14	10	6	4	2	0	15	12	6

```
data RATINGS;
input CATEGORY DRUG DIFF @@;
datalines;
1 1 8 1 1 4 1 1 0 1 2 10 1 2 8 1 2 6 1 3 8 1 3 6 1 3 4
2 1 14 2 1 10 2 1 6 2 2 4 2 2 2 2 2 0 2 3 15 2 3 12 2 3 9
;
proc glm;
class CATEGORY DRUG;
model DIFF=CATEGORY DRUG CATEGORY*DRUG;
output out=OUT1 p=YHAT residual=RESID;
run;
```

The SAS code is almost the same as for one-way ANOVA. The response variable is difference between the two ratings (DIFF). Since we have two variables representing main factors (CATEGORY & DRUG), we should put both of them into the CLASS statement and include these two variables as well as their interaction in the MODEL statement.

Source	DF	Type III SS	Mean Square	F Value	Pr > F
category	1	18.0000000	18.0000000	2.04	0.1789
drug	2	48.0000000	24.0000000	2.72	0.1063
category*drug	2	144.0000000	72.0000000	8.15	0.0058

P-value=0.1789 for **category**: there is no significant difference between the two categories of patients.
P-value=0.1063 for **drug**: there is no significant difference among the three drugs used.
P-value=0.0058 for **category*drug**: category and drug do interact. It means that the differences among the effectiveness of the three drugs are not the same for the two categories of patients.

Since there is no significant difference between the two categories of patients and among the three drugs used, post hoc tests are not used for either of the main effects (Post hoc tests are only used once you have determined that differences exist among the means of a main effect.) Since the interaction term has a significant influence, we add LSMEANS statement into the PROC GLM procedure to find which levels are significantly different:

```
lsmeans CATEGORY*DRUG/stderr pdiff adjust=tukey;
```

CATEGORY	DRUG	DIFF LSMEAN	Standard Error	Pr > |t|	LSMEAN Number
1	1	4.0000000	1.7159384	0.0380	1
1	2	8.0000000	1.7159384	0.0005	2
1	3	6.0000000	1.7159384	0.0044	3
2	1	10.0000000	1.7159384	<.0001	4
2	2	2.0000000	1.7159384	0.2664	5
2	3	12.0000000	1.7159384	<.0001	6

Least Squares Means for effect CATEGORY*DRUG Pr > |t| for H0: LSMean(i)=LSMean(j) Dependent Variable: DIFF						
i/j	1	2	3	4	5	6
1		0.5857	0.9572	0.2068	0.9572	0.0555
2	0.5857		0.9572	0.9572	0.2068	0.5857
3	0.9572	0.9572		0.5857	0.5857	0.2068
4	0.2068	0.9572	0.5857		0.0555	0.9572
5	0.9572	0.2068	0.5857	0.0555		0.0138
6	0.0555	0.5857	0.2068	0.9572	0.0138	

To check the assumption of homogeneous variance, a new method will be used for today.

```
proc plot data=out1;
plot resid*yhat;
run;
```

Plot of resid*yhat. Legend: A = 1 obs, B = 2 obs, etc.