SAS 8. Analysis of variance

3. ANOVA: Two Independent Variables

Suppose we ran the same experiment for comparing reading methods, but using 15 male and 15 female subjects. In addition to comparing reading-instruction methods, we could compare male versus female reading speeds. Finally, we might want to see if the effects of the reading methods are the same for males and females.

```
DATA TWOWAY;
      INPUT GROUP $ GENDER $ WORDS @@;
DATALINES;
X M 700 X M 850 X M 820 X M 640 X M 920
Y M 480 Y M 460 Y M 500 Y M 570 Y M 580
Z M 500 Z M 550 Z M 480 Z M 600 Z M 610
X F 900 X F 880 X F 899 X F 780 X F 899
Y F 590 Y F 540 Y F 560 Y F 570 Y F 555
Z F 520 Z F 660 Z F 525 Z F 610 Z F 645
;
PROC ANOVA DATA=TWOWAY;
      TITLE 'ANALYSIS OF READING DATA';
      CLASS GROUP GENDER;
      MODEL WORDS=GROUP | GENDER;
      MEANS GROUP | GENDER / DUNCAN;
RUN;
```

If we don’t include the vertical line, none of the interaction terms will be estimated.
In this case, the term GROUP | GENDER can be written as
GROUP GENDER GROUP*GENDER

The “|” symbol is especially useful when we have higher order factorial designs such as GROUP|GENDER|DOSE. Written the ong way, this would be GROUP GENDER DOSE GROUP*GENDER GROUP*DOSE GENDER*DOSE GROUP*GENDER*DOSE

Output:

Source	DF	Anova SS	Mean Square	F Value	Pr > F
group	2	503215.2667	251607.6333	56.62	<.0001
gender	1	25404.3000	25404.3000	5.72	0.0250
group*gender	2	2816.6000	1408.3000	0.32	0.7314

In a two-way analysis of variance, when we look at GROUP effects, we are comparing GROUP levels without regard to GENDER. That is, when the groups are compared we combine the data from both GENDERS. Conversely, when we compare males to females, we combine data from the three treatment groups.

The term GROUP*GENDER is called an interaction term. If group differences were not the same for males and females, we could have a significant interaction.

Means with the same letter
are not significantly different.

Duncan Grouping	Mean	N	GROUP

A	828.80	10	X
B	570.00	10	Z
B			
B	540.50	10	Y

The first comparison shows group X as being significantly different from Y and Z.

Means with the same letter are not significantly different.

Duncan Grouping	Mean	N	GENDER
A	675.53	15	F
B	617.33	15	M

the second table shows that females have significantly higher reading speeds than males.

5. Interpreting Significant Interactions

Now consider an example that has a significant interaction term. We have two groups of children. One group is considered normal; the other, hyperactive.

```
data ritalin;
      do group = 'normal' , 'hyper';
      do drug = 'placebo','ritalin';
                  do subj = 1 to 4;
                        input activity @@;
                        output;
                  end;
            end;
      end;
datalines;
50 45 55 52 67 60 58 65 70 72 68 75 51 57 48 55
;
proc anova data=ritalin;
      title 'activity study';
      class group drug;
      model activity=group | drug;
      means group | drug;
run;
```

Source	DF	Anova SS	Mean Square	F Value	Pr > F
group	1	121.0000000	121.0000000	8.00	0.0152
drug	1	42.2500000	42.2500000	2.79	0.1205
group*drug	1	930.2500000	930.2500000	61.50	<.0001

From the ANOVA table above, we notice that there is a strong GROUP*DRUG interaction term. When this occurs, we need to be very careful about interpreting any of the main effects (GROUP and DRUG in this example).

One way to read the data values is a straightforward INPUT DRGU $ GROUP $ ACTIVITY; statement. Instead we use nested Do loops, more as a demonstration of SAS programming rather than to shorten the program

Then we look closely at the means of the interaction groups. The best way to explain a two-way interaction is to take the cell means and plot them.

```
proc means data=ritalin nway noprint;
   class group drug;
   var activity;
   output out=means mean=M;
run;

symbol1 value=square color=black I=join;
symbol2 v=circle c=black i=join;
proc gplot data=means;
  plot M*drug=group;
run;
```

The graph shows that normal children increase their activity when given Ritalin, while hyperactive children are calmed by Ritalin. Before comparing the means of DRUG, we combined the data from normal and hyperactive children. Since the means of normal and hyperactive children tend to cancel the effect of DRUG of each other, the average activity with placebo and Ritalin is about the same. Then we can compare the means between placebo and Ritalin within the normal and hyperactive children. We can create a new variable named 'cond' with the four values: normal_placebo, normal_ritalin, hyper_placebo and hyper_ritalin.

```
data ritalin_new; set
ritalin; cond=group
|| drug; run;

proc anova data=ritalin_new;
      title 'one-way anova ritalin study';
      class cond;
      model activity = cond;
      means cond / duncan;
run;
```

Duncan Grouping	Mean	N	cond
A	71.250	4	hyper placebo
B	62.500	4	normal ritalin
C	52.750	4	hyper ritalin
C			
C	50.500	4	normal placebo

We can see that placebo is different from Ritalin within normal and hyper groups and while given Ritalin normal children act differently from hyperactive children.

6. N-Way Factorial Designs

With three independent variables, we have three main effects, three two-way interactions, and one three-way interaction. One usually hopes that the higher-order interactions are not significant since they complicate the interpretation of the main effects and the low-order interactions.

```
PROC ANOVA DATA=THREEWAY;
      TITLE 'THREE WAY ANALYSIS OF VARIANCE';
      CLASS GROUP GENDER DOSE;
      MODEL ACTIVITY = GROUP | GENDER | DOSE;
MEANS GROUP | GENDER | DOSE;
RUN;
```