Topic 22 - Interaction in Two Factor ANOVA

STAT 525 - Fall 2013

Outline

- Strategies for Analysis
  - when interaction not present
  - when interaction present
  - when $n_{ij} = 1$
  - when factor(s) quantitative

General Plan

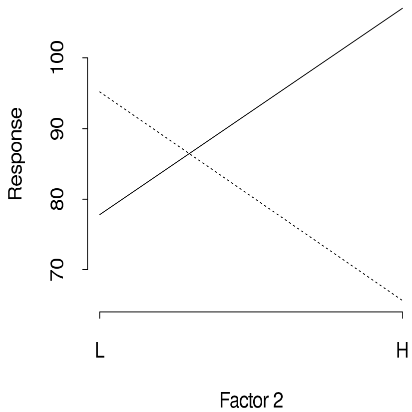
- Construct scatterplot / interaction plot
- Run <u>full</u> model
- Check assumptions
  - Residual plots
  - Histogram / QQplot
  - Ordered residuals plot
- Check significance of interaction

Interaction Not Significant

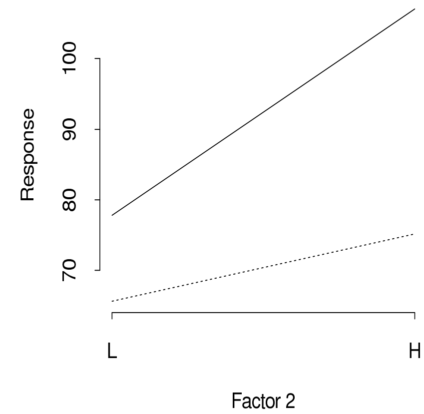
- Determine whether pooling is beneficial
- If yes, rerun analysis without interaction
- Check significance of main effects
- If factor insignificant, determine whether pooling is beneficial
- If yes, rerun analysis as one-way ANOVA
- If statistically significant factor has more than two levels, use multiple comparison procedure to assess differences
- Contrasts and linear combinations can also be used

Interactions

- Opposite behavior (no Factor 2 effect?)



- Not similar increase (still Factor 2 effect?)



If Interaction Significant

- Determine if interaction "important"
  - May not be of practical importance
  - May be like plot #2
  - Often due to one cell mean
- If no, use previously described methods making sure to leave interaction in the model (no pooling). Carefully interpret the marginal means as averages over the levels of the other factor and not a main effect
- If yes, take approach of one-way ANOVA with ab levels. Use linear combinations to compare various means (e.g., levels of factor A for each level of factor B). Use the interaction plots for discussion purposes.

Using Estimate Statement

- Must formulate in terms of factor effects model
- Order of factors determined by order in class statement not the model statement
- Example from Castle Bread Company
  - $H_0 : \mu_{2.} = \mu_{1.} + \mu_{3.}$
  - Rewriting in factor effect terms

$$\begin{aligned}
\mu_{2.} &= \mu_{21} + \mu_{22} \\
&= \mu + \alpha_2 + \beta_1 + (\alpha\beta)_{21} + \\
&\quad\ \mu + \alpha_2 + \beta_2 + (\alpha\beta)_{22} \\
&= 2\mu + 2\alpha_2 + \beta_. + (\alpha\beta)_{2.} \\
\mu_{1.} + \mu_{3.} &= \mu_{11} + \mu_{12} + \mu_{31} + \mu_{32} \\
&= 4\mu + 2\alpha_1 + 2\alpha_3 + 2\beta_. + (\alpha\beta)_{1.} + (\alpha\beta)_{3.}
\end{aligned}$$

Using Slice Statement

- Slice option performs one-way ANOVA for fixed level of other factor
- Can also express that as contrast statement
- Following output presents results from two contrasts
  - $H_0 : 2\mu_{2.} = \mu_{1.} + \mu_{3.}$
  - $H_0 : \mu_{11} = \mu_{21} = \mu_{31}$
- See if you can come up with the same contrast statements

SAS Commands

```
proc glm data=a1;
   class height width;
   model sales=height width height*width;
   estimate 'middle is sum of other two heights'
      intercept -2    height -2 2 -2    width -1 -1
           height*width -1 -1 1 1 -1 -1;
   contrast 'middle two vs all others'
      height -.5 1 -.5    height*width -.25 -.25 .5 .5 -.25 -.25;
   estimate 'middle two vs all others'
      height -.5 1 -.5    height*width -.25 -.25 .5 .5 -.25 -.25;
   contrast 'height same for normal width'
      height 1 -1 0 height*width 1 0 -1 0 0 0,
      height 0 1 -1 height*width 0 0 1 0 -1 0;
   means height*width;

proc glm data=a1;
   class height width;
   model sales=height width height*width;
   lsmeans height*width / slice=width;
```


Output

```
Contrast           DF  Contrast SS  Mean Square  F Value  Pr > F
middle vs others    1  1536.000000  1536.000000   148.65  <.0001
height for normal   2   700.000000   350.000000    33.87  0.0005
Parameter                 Estimate   Std  Error t Value Pr > |t|
middle is sum of others -38.0000000  5.56776436   -6.83  0.0005
middle vs others         24.0000000  1.96850197   12.19  <.0001

Level of     Level of           ------------sales------------
height       width        N            Mean            Std Dev
1            1            2       45.0000000        2.82842712
1            2            2       43.0000000        4.24264069
2            1            2       65.0000000        4.24264069
2            2            2       69.0000000        2.82842712
3            1            2       40.0000000        1.41421356
3            2            2       44.0000000        2.82842712
           height*width Effect Sliced by width for sales
width   DF   Sum of Squares    Mean Square    F Value    Pr > F
1        2      700.000000      350.000000      33.87    0.0005
2        2      868.000000      434.000000      42.00    0.0003
```


One Observation Per Cell

- Do not have enough information to estimate **both** the interaction effect and error variance
- With interaction, error degrees of freedom is $ab(n-1) = 0$
- Common to assume there is no interaction (i.e., pooling)
  - $\text{SSE}^* = \text{SSAB} + 0$
  - $\text{df}^*_E = \text{df}_{AB} + 0$
- Can also test for less general type of interaction that requires fewer degrees of freedom

Tukey's Test for Additivity

- Consider special type of interaction
- Assume following model

$$Y_{ij} = \mu + \alpha_i + \beta_j + \theta\alpha_i\beta_j + \varepsilon_{ij}$$

- Uses up only one degree of freedom
- Other variations possible (e.g., $\theta_i\beta_j$)
- Want to test $H_0 : \theta = 0$
- Will use regression after estimating factor effects to test θ

Example Page 882

- Y is the premium for auto insurance
- Factor A is the size of the city
  - $a = 3$: small, medium, large
- Factor B is the region
  - $b = 2$: east, west
- Only one city per cell was observed

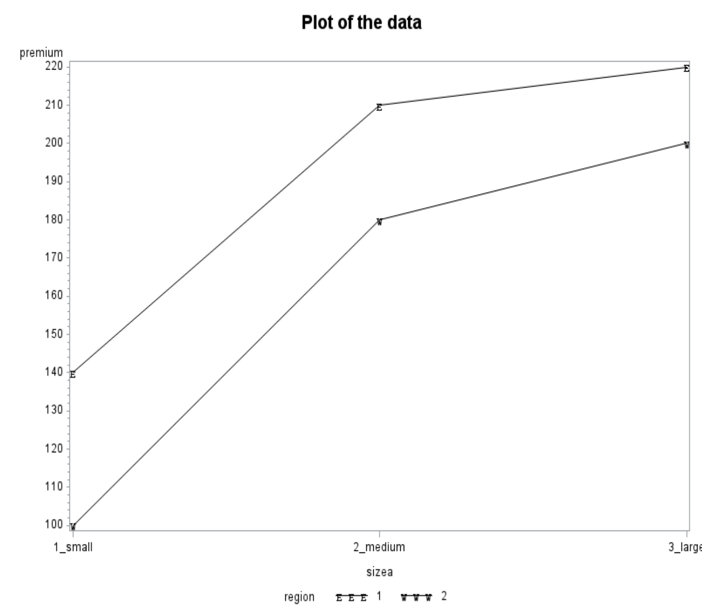
SAS Commands

```
data a1; infile 'u:\.www\datasets525\CH20TA02.txt';
   input premium size region;

if size=1 then sizea='1_small ';
if size=2 then sizea='2_medium';
if size=3 then sizea='3_large ';

proc glm data=a1;
   class sizea region;
   model premium=sizea region / solution;
   means sizea region / tukey;

symbol1 v='E' i=join c=black; symbol2 v='W' i=join c=black;
title1 'Plot of the data';
proc gplot data=a2;
   plot premium*sizea=region/frame;
```

SAS Commands

```
proc glm data=a1;
   model premium=;  output out=aall p=muhat;

proc glm data=a1; class size;
   model premium=size;  output out=aA p=muhatA;

proc glm data=a1; class region;
   model premium=region;  output out=aB p=muhatB;

data a2; merge aall aA aB;
   alpha=muhatA-muhat; beta=muhatB-muhat;  atimesb=alpha*beta;

proc print data=a2;
   var size region atimesb;

proc glm data=a2;
   class size region;
   model premium=size region atimesb/solution;
run;
```


Output

Obs	size	region	atimesb
1	1	1	-825
2	1	2	825
3	2	1	300
4	2	2	-300
5	3	1	525
6	3	2	-525

Note: These estimates are based on the factor effects model where $\sum \alpha = 0$ and $\sum \beta = 0$. While not shown, the following were used to compute atimesb: $\hat{\mu} = 175$, $\hat{\mu}_{1.} = 120$, $\hat{\mu}_{2.} = 195$, $\hat{\mu}_{3.} = 210$, $\hat{\mu}_{.1} = 190$, and $\hat{\mu}_{.2} = 160$.

Output

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	10737.09677	2684.27419	208.03	0.0519
Error	1	12.90323	12.90323		
Corrected Total	5	10750.00000			

Source	DF	Type I SS	Mean Square	F Value	Pr > F
size	2	9300.000000	4650.000000	360.37	0.0372
region	1	1350.000000	1350.000000	104.62	0.0620
atimesb	1	87.096774	87.096774	6.75	0.2339

Parameter		Estimate	Std Error	t Value	Pr > |t|
Intercept		195.0000000 B	2.93294230	66.49	0.0096
size	1	-90.0000000 B	3.59210604	-25.05	0.0254
size	2	-15.0000000 B	3.59210604	-4.18	0.1496
size	3	0.0000000 B	.	.	.
region	1	30.0000000 B	2.93294230	10.23	0.0620
region	2	0.0000000 B	.	.	.
atimesb		-0.0064516	0.00248323	-2.60	0.2339

Note: These are the same parameter estimates as the original model without the interaction term.

One Quantitative Factor

- Similar to regression with one indicator or categorical variable
- Plot the means vs the quantitative factor for each level of the categorical factor
- Based on this plot,
  - Consider linear/quadratic relationships for the quantitative factor
  - Consider different slopes for the different levels of the categorical factor
  - Can perform lack of fit analysis
- If two quantitative variables, can consider linear and quadratic terms. Interactions modeled as the direct product. Lack of fit test very useful. Again very similar to linear regression models.

Background Reading

- KNNL Chapters 19, 20
- knnl849.sas, knnl883.sas
- KNNL Chapter 23