ANOVA

1 One-Way ANOVA with Contrast Codes

Contrast coding is a useful tool to estimate and test differences, with respect to a continuous dependent variable, between different groups within the same categorical variable.

For example lets use the mammography data set with the SYMPT variable. This variable has four different levels; *Strongly Agree, Agree, Disagree, and Strongly Disagree.* We are interested in testing the differences, with respect to PB, between *Strongly Agree and Agree* versus *Disagree and Strongly Disagree.* There are three rules that we need to follow when using contrast codes (from pg 333 in Cohen and Cohen):

1. The sum of the weights for all groups must be zero.
2. The sum of the products for each pair must be zero.
3. The difference in the value of positive weights and negative weights should be one for each code variable

For example, the code variables for our question could be:

	Code variables		
SYMPT	C_1	C_2	C_3
Strongly Agree	1/2	1/2	0
Agree	1/2	-1/2	0
Disagree	-1/2	0	1/2
Strongly Disagree	-1/2	0	-1/2

Our model is:

$$Y_i = B_0 + B_1 C_{1i} + B_2 C_{2i} + B_2 C_{3i} + \epsilon_i \tag{1}$$

What does our B_0 estimate?

What does our B_1 estimate?

What does our B_2 estimate?

What does our B_3 estimate?

Now lets try this in SAS:

```
data new;
set tmp1.meexp2;
if sympt = 1 then C1 = 1/2;
if sympt = 2 then C1 = 1/2;
if sympt = 3 then C1 = -1/2;
if sympt = 4 then C1 = -1/2;
C2 = 0;
```

```
if sympt = 1 then C2 = 1/2;
if sympt = 2 then C2 = -1/2;
C3 = 0;
if sympt = 3 then C3 = 1/2;
if sympt = 4 then C3 = -1/2;
run;

proc print;
var sympt C1 C2 C3;
run;

proc glm;
model pb = c1 c2 c3/solution;
OUTPUT OUT=OUT1 R=RY P=PY;

PROC GPLOT DATA=OUT1;
PLOT RY*PY;
RUN;
QUIT;
```

2 Two-Way ANOVA (2 levels for each) with No Interaction

The main goal in two-way ANOVA is to compare the mean of a certain response variable across different levels of two factor effects.

Example One:
A treatment for a disease has severe side effects. Let Y be a numerical measure of the severity of these side effects. Let Factor A be the treatment for the disease itself (levels absent/present), and Factor B be a drug to minimize the severity of the side effects (levels absent present). Considering all combinations, there are 2 x 2 = 4 treatments under consideration.

		Taught BSE	
		No	Yes
HISTORY	No	$\bar{Y}_{11}$	$\bar{Y}_{12}$
	Yes	$\bar{Y}_{21}$	$\bar{Y}_{22}$

The Model that we would like to fit is:

$$Y_i \;=\; B_0 \;+\; B_1 History_i \;+\; B_2 BSE_i \;+\; \epsilon_i \tag{2}$$

Where our variables for the i^{th} individual:

Y_i = severity of the side effects

$$History_i = \begin{cases} 0, & No\ History \\ 1, & History \end{cases}$$

$$BSE_i = \begin{cases} 0, & Not\ Taught\ BSE \\ 1, & Taught\ BSE \end{cases}$$

What does our model look like when Disease = 0?

What does our model look like when Disease = 1?

What does our B_0 estimate?

What does our B_1 estimate?

What does our B_2 estimate?

What is the overall F-test testing in terms of our means?

What will the individual $H_0 : B_i = 0$ test in terms of our means?

Lets try fitting and interpreting this in SAS:

```
data new;
set tmp1.meexp2;
run;

proc sort;by HIST BSE;

proc means mean noprint;
by HIST BSE;
VAR PB;
OUTPUT OUT=YBAR MEAN=MeanPB;

PROC GPLOT;
symbol1 color=red interpol=join value=none;
symbol2 color=black interpol=join value=none;
PLOT MeanPB*HIST=BSE/ HAXIS=-1 TO 2 BY 1;
RUN;
QUIT;

PROC GLM DATA=MAMMO;
MODEL PB = HIST BSE/SOLUTION;
OUTPUT OUT=OUT1 R=RY P=PY;
```

```
PROC PLOT DATA=OUT1;
PLOT RY*PY;
RUN;
QUIT;
```

3 Two-Way ANOVA Full Model w/ Dummy Coding

Above we assumed that there was no interaction in the different effects. However as we saw in our first example that might not be a good assumption. Lets try refitting that model with an interaction term. Lets denote:

		Taught BSE	
		No	Yes
HISTORY	No	$\bar{Y}_{11}$	$\bar{Y}_{12}$
	Yes	$\bar{Y}_{21}$	$\bar{Y}_{22}$

Or lets fit the model:

$$PB_i \;=\; B_0 \;+\; B_1 HIST_i \;+\; B_2 BSE_i \;+\; B_3 HIST_i * BSE_i \;\epsilon_i \tag{3}$$

Using Dummy coding what will our:

What does our B_0 estimate?

What does our B_1 estimate?

What does our B_2 estimate?

What does our B_3 estimate?

What will the individual $H_0 : B_3 = 0$ test?

Lets try this in SAS:

```
proc means;
by HIST BSE;
VAR PB;

PROC GLM DATA=MAMMO;
MODEL PB = HIST BSE HIST*BSE/SOLUTION;
OUTPUT OUT=OUT1 R=RY P=PY;

PROC PLOT DATA=OUT1;
PLOT RY*PY;
RUN;
QUIT;
```

Since our interaction term is insignificant I would refit this model as an additive model (no interaction) and interpret the means separately.

4 Two-Way ANOVA w/ Unweighed Effect Coding

Now lets use the same model, but instead of having 0 as our control we will make -1 our control. First lets do this for just BSE, and we will leave the coding of History to be the same. BSE is now:

$$BSE_i = \begin{cases} -1, & Not\ Taught \\ 1, & Taught \end{cases}$$

We now have the same model.

$$PB_i \ = \ B_0 \ + \ B_1 HIST_i \ + \ B_2 BSE_i \ + \ \epsilon_i \tag{4}$$

But the interpretations have changed. So using effect coding:

What does our B_0 estimate?

What does our B_1 estimate?

What does our B_2 estimate?

Lets try this in SAS:

```
data mammo;
set TMP1.meexp2;
BSE2 = 1;
if BSE = 0 then BSE2 = -1;
run;

PROC GLM DATA=MAMMO;
MODEL PB = HIST BSE2 /SOLUTION;
OUTPUT OUT=OUT1 R=RY P=PY;
run;
quit;
```

5 Two-Way ANOVA Full Model w/ Unweighed Effect Coding

Now lets use the same model as above, but lets add an interaction term. The model becomes:

$$PB_i \ = \ B_0 \ + \ B_1 HIST_i \ + \ B_2 BSE_i \ + \ \ B_3 HIST_i * BSE_i \epsilon_i \tag{5}$$

Our B_0, B_1, B_2 stay the same.

What does our B_3 estimate?

Lets try this in SAS:

```
PROC GLM DATA=MAMMO;
MODEL PB = HIST BSE2 HIST*BSE2/SOLUTION;
OUTPUT OUT=OUT1 R=RY P=PY; run; quit;
```

What would happen if we changed our coding of BSE to: