Demo: Fitting a Multiple Logistic Regression Model with Interactions Using PROC LOGISTIC

Filename: **st107d06.sas**

In this demonstration, we fit a multiple logistic regression model using the backward elimination method. The full model should include all the main effects and two-way interactions.

```
PROC LOGISTIC DATA=SAS-data-set <options>;
    CLASS variable <(options)> ... < / options>;
    MODEL variable <(variable_options)> = <effects> < / options>;
    UNITS <independent1=list1> ... < / options>;
RUN;
```

1. Open program st107d06.sas.

```
/*st107d06.sas*/  /*Part A*/
proc logistic data=STAT1.ameshousing3 plots(only)=(effect oddsratio);
    class Fireplaces(ref='0') Lot_Shape_2(ref='Regular') / param=ref;
    model Bonus(event='1')=Basement_Area|Fireplaces|Lot_Shape_2 @2 /
            selection=backward clodds=pl slstay=0.10;
    units Basement_Area=100;
    title 'LOGISTIC MODEL (3): Backward Elimination '
            'Bonus=Basement_Area|Fireplaces|Lot_Shape_2';
run;

/*st107d06.sas*/  /*Part B*/
proc logistic data=STAT1.ameshousing3
              plots(only)=oddsratio(range=clip);
    class Fireplaces(ref='0') Lot_Shape_2(ref='Regular') / param=ref;
    model Bonus(event='1')=Basement_Area|Lot_Shape_2 Fireplaces;
    units Basement_Area=100;
    oddsratio Basement_Area / at (Lot_Shape_2=ALL) cl=pl;
    oddsratio Lot_Shape_2 / at (Basement_Area=1000 1500) cl=pl;
    title 'LOGISTIC MODEL (3.1): Bonus=Basement_Area|Lot_Shape_2 Fireplaces';
run;
```

The code in Part A is the same as the previous demonstration, except for changes to the MODEL statement. The MODEL statement now lists the three main effects, Basement_Area, Fireplaces, and Lot_Shape_2, as well as the vertical bar operator between each pair to specify their interactions. Following the last effect, the @2 indicates to include only the two-factor interactions along with the main effects. If you increase it to @3, then you construct a model with all the main effects, the two-factor interactions, and the three-factor interaction.

The SELECTION= option specifies the backward elimination method for variable selection, and SLSTAY changes the significance level to 0.10. Using an iterative process, PROC LOGISTIC eliminates terms one at a time, starting with the least significant interaction, the one with the largest p-value. By default, PROC LOGISTIC preserves model hierarchy, meaning, if an interaction is in the model, the main effects contributing to it must remain in the model, whether they are significant or not. Therefore, the procedure only considers main effects for possible removal if the corresponding interaction has first been removed.

2. Submit the code in Part A.

3. <u>Review the output.</u>

The Model Information table is equivalent to previous demos since we are using the same data, response, model, and optimization technique. However, a note in the Response Profile indicates that one observation was deleted due to missing value information.

The Class Level Information table is the same, and this model did converge, so we can trust our results.

Let's jump to Summary of Backward Elimination table (in the Model Building Summary) to view the stepwise process. At Step 1, the interaction between Lot_Shape_2 and Fireplaces was removed, because it had the least significant p-value. Then, at step 2, the interaction between Basement_Area and Fireplaces was removed for the same reason. The procedure stops after the two interactions involving Fireplaces are removed.

The next table, the Joint Tests, tests that all the parameters associated with that effect are 0 for the final model. Notice the main effects are included, but only the interaction for Basement_Area by Lot_Shape_2 is included. Also, notice that individually, each effect is significant at the alpha=.10 level.

The Parameter Estimates table, Analysis of Maximum Likelihood Estimates, displays the estimates and significance for each parameter in the final model. Notice that Fireplaces was significant at the 0.10 level when tested simultaneously in the Joint Tests table, but here, both parameters have a p-value greater than 0.10 and are still included in the final model.

In the Association Statistics table, Association of Predicted Probabilities and Observed Responses, we see that the c value is a slight improvement over the previous model that only included the main effects. Here, it's almost 0.94 compared to 0.93 previously.

Odds ratios are not calculated for effects involved in interactions. Any single odds ratio for Basement_Area or for Lot_Shape_2 would be misleading, because the effects vary for each at different levels of the other variable. Again, the odds of being bonus eligible are more than two times the odds for homes with one fireplace compared to the odds for homes with 0 fireplaces. But the odds decrease when comparing 2 versus 0 fireplaces.

In the effect plot, Predicted Probabilities for Bonus=1, notice the sigmoidal probability curves overlap and are not all equally shaped. This visually displays the Basement_Area by Lot_Shape_2 interaction we discovered with backward elimination. The effect of Basement_Area depends on the lot shape of the home, causing an interaction effect.

Statistics 1: Introduction to ANOVA, Regression, and Logistic Regression

Close