

Demo: Generating Predictions Using PROC PLM

Filename: **st107d07.sas**

Let's run the LOGISTIC procedure again, with the same effects as before, except let's add in the STORE statement to save the model information and score new data.



```
PROC LOGISTIC DATA=SAS-data-set <options>;
CLASS variable <(options)> ... </ options>;
MODEL variable <(variable_options)> = <effects> </ options>;
UNITS <independent1=list1> ... </ options>;
STORE <OUT=> item-store-name </ LABEL='label'>;
RUN;
```

```
PROC PLM RESTORE=item-store-specification <options>;
SCORE DATA=SAS-data-set <OUT=SAS-data-set> <keyword<=name>> ... </ options>;
RUN;
```

1. Open program st107d07.sas.



```
/*st107d07.sas*/
```

```
ods select none;
proc logistic data=STAT1.ameshousing3;
    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;
    store out=isbonus;
run;
ods select all;
```

```
data newhouses;
    length Lot_Shape_2 $9;
    input Fireplaces Lot_Shape_2 $ Basement_Area;
    datalines;
    0 Regular 1060
    2 Regular 775
    2 Irregular 1100
    1 Irregular 975
    1 Regular 800
    ;
run;

proc plm restore=isbonus;
    score data=newhouses out=scored_houses / ILINK;
    title 'Predictions using PROC PLM';
run;
```

```
proc print data=scored_houses;  
run;
```

Before the PROC LOGISTIC step, we add the statement ODS SELECT NONE, which suppresses the output, and we add ODS SELECT ALL at the end of the step to make sure that we get the output from the next step we run. In the STORE statement, we specify the name of the item store we want to save, isbonus. Next, we'll create a data set named newhouses that contains the new data we want to score.

Finally, we'll use PROC PLM to generate predictions for the newhouses data set. The RESTORE option specifies that the predictions will be based on the analysis results saved in the item store isbonus. The SCORE statement specifies that SAS will score the data set named newhouses and will write the results into a new data set named scored_houses. The ILINK option requests predictions on the probability scale as opposed to the logit scale. This makes the predictions easier to interpret.

We'll close our program with a PRINT procedure so that we can view the scored data.

2. Submit the code.

3. [Review the output.](#)

As expected, the Predictions table produced by PROC PLM shows that the house with the highest predicted probability of being bonus eligible (0.306) has an irregular lot shape, 1 fireplace, and a basement area of 975 square feet. The house with the lowest predicted probability (0.0004) has a regular lot shape, 2 fireplaces, and a basement area of 775. Again, the predicted values in the last column are probabilities because we used the ILINK option. Otherwise, the last column would be the predicted logit values.

Be sure that you generate predictions only for new data records that fall within the range of the training data. If not, predictions could be invalid due to extrapolation. We assume that the modeled relationship between predictors and the response holds across the span of the observed data. We should not assume that this relationship holds everywhere.