Paper 350-2011

Visualizing Marginal Effects from Interactions in Generalized Linear Models

Xiao Chen, UCLA Academic Technology Services, Los Angeles, CA
Brigid Wilson, UCLA Department of Statistics, Los Angeles, CA

ABSTRACT

Interactions are often included in generalized linear models (GLM). Interpreting these interactions in the transformed scale of the linear equation is like interpreting an interaction in an OLS regression. However, this is rarely the scale in which results are discussed, and interpreting interactions in the non-transformed scale is not straightforward.

Focusing on a continuous by categorical interaction in a logistic regression, we present code for visualizing a marginal effect in the probability scale. This visualization is a useful tool for understanding a model effect that is difficult to intuit from the standard model output. By presenting the results in the researcher's scale of interest, it allows the researcher to better communicate the model results.

INTRODUCTION

A researcher running a generalized linear model might be thrilled to see that the interaction he included in the model has a low p-value. His suspicions have been confirmed! The findings are significant! But what exactly has he found? While a researcher may believe an interaction effect exists, he may still be unsure of the precise nature of the interaction and unable to understand this aspect of his model.

This paper aims to give researchers a general approach and some SAS® code for understanding such an interaction using visualizations of the marginal effects of the interaction variables. We will not be discussing model methodology or inference made from models. Our aim is to present an informal tool that allows researchers to see the relationships defined by interaction effects in GLM.

We will focus on a logistic model that includes the interaction of a categorical variable and a continuous variable. For simplicity, we will assume that the model consists of these two interacted variables. The categorical variable in our example has two levels.

A macro version of the code presented in this paper can be found at http://www.ats.ucla.edu/stat/sas/macros/sugi2011.sas.txt. The dataset used in the examples shown here can be found at http://www.ats.ucla.edu/stat/sas/macros/logitcatcon.sas7bdat.

THE EASY WAY OUT

Generalized linear models are, by definition, linear in some scale. In that scale, we can interpret an interaction as we would an interaction in an OLS regression and, technically, we are not wrong in doing so.

Since we are looking at a logistic model in SAS, we will use the OUTMODEL option from PROC LOGISTIC to save our estimated model. Figure 1 shows the Maximum Likelihood Estimates output from this model.

```
proc logistic data = logitcatcon descending outmodel = sasuser.l1;
  model y = f|s;
run;
```

Figure 1.
Parameter estimates from PROC LOGISTIC

The LOGISTIC Procedure

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-9.2536	1.9419	22.7085	<.0001
f	1	5.7866	2.3025	6.3162	0.0120
s	1	0.1773	0.0364	23.6885	<.0001
f*s	1	-0.0895	0.0439	4.1580	0.0414

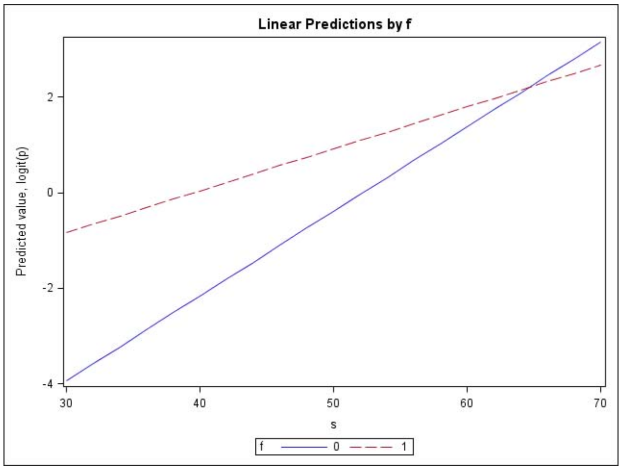
Then we can create a dataset containing all the interaction values of interest. Using this new dataset, we can again use the INMODEL option of PROC LOGISTIC to score this dataset, generating predicted probabilities for both 0 and 1 from which we can calculate the predicted logit values.

```
data test;
  do s = 30 to 70 by 2;
    do f = 0 to 1;
      output;
    end;
  end;
run;

proc logistic inmodel = sasuser.l1;
  score data = test out = predp;
run;

data xbeta; set predp;
  linpred = log( p_1 / p_0 );
run;
```

We can plot these linear predictions and see the distinct lines for f=0 and f=1.

```
proc sgplot data = xbeta;
  series x = s y = linpred / group = f;
  yaxis label="Predicted value, logit(p)";
  xaxis label="s";
  title "Linear Predictions by f";
run;
```

Figure 2.
Predicted values by categorical variable levels, logit scale



PROBLEMS WITH THE EASY WAY OUT

GLM models are not usually fit with the link-defined scale in mind. When predicting a dichotomous outcome, one generally aims to discuss the odds or the probability of an outcome level, not the logit or probit. Therein lies the problem with the easy way out: it is easy, but not very useful for explaining results.

From the regression output in Figure 1, the coefficient estimate of the interaction is the difference in the slopes of the two plotted lines: the f=1 line has a slope that is 0.0895 less than the f = 0 line. The p-value in the output suggests that the difference in slopes is statistically significant, assuming alpha = 0.05.

From Figure 2, we can see this difference in slopes. Without bothering with the algebra, we can see that the predicted values are closest when s is between 60 and 70; f = 0 has lower predicted values than f = 1 for lower s values and higher predicted values than f = 1 for higher s values. The difference in predicted logit values is also linear in s.

But what can we say about the probability of the outcome? Or the difference in probabilities between f = 1 and f = 0 for a given value of s?

THE TRANSFORMED SCALE

In SAS, GLM and OLS regression are similarly easy to implement and produce some similar output tables. Perhaps as a result, the GLM link function often goes unexamined and the interpretation GLM too often ends with the p-values of the parameter estimates.

The relationship between the transformed outcome and the independent variables can be unintuitive and quite nuanced. Graphing this relationship can be eye-opening.

As s increases or decreases, the resulting changes in the linear predictions can easily be described, but not so easily the changes in the predicted probabilities. By the nature of the logit transformation, the predicted probabilities from a logistic model are *not* linear in s.

Figure 2 illustrated our model's interaction in the logit scale. The interaction defined two linear relationships in s: one where f = 0 and one where f = 1. To examine this interaction in the probability scale, we will first plot the predicted probability curves for each level of f. To better understand the marginal effect of f in the probability scale, we can calculate the difference these predicted probability curves over a range of s values. We can plot these probability differences and generate confidence intervals about them.

In earlier code, we used the INMODEL and OUTMODEL options of PROC LOGISTIC to generate predicted values. SAS 9.2 offers some more flexible options that allow for the easy generation of model predictions.

- The STORE statement (available in PROC LOGISTIC and PROC GLM) saves model results that, among other things, can be accessed for out-of-sample prediction.
- PROC PLM generates out-of-sample predictions in both un-transformed and transformed scales.

We start by running the same model, storing results, and then creating a dataset containing every combination of our interaction variables of interest. Using PROC PLM and our stored results, we generate predicted probabilities and calculate differences in these values.

```
proc logistic data = logitcatcon noprint descending;
  model y = f|s;
  store sasuser.l2;
run;

proc plm source=sasuser.l2;
  score data=test out=_temp1_ predicted / ilink;
run;

proc sgplot data = _temp1_;
  series x = s y = predicted / group = f;
  yaxis label = "Predicted probability";
  xaxis label = "s";
  title "Predicted probability by f";
run;

data _temp1_; set _temp1_;
  by s;
  retain diff;
  if f = 0 then diff = predicted;
  if f = 1 then do;
    diff = predicted - diff;
    output;
  end;
run;

proc sgplot data = _temp1_;
  series x = s y = diff;
  yaxis label = "Probability difference";
  xaxis label = "s";
  refline 0 / axis = y;
  title "Difference in predicted probabilities";
run;
```

We can plot these probability differences against s and see the non-linear nature of this relationship. From this plot, we can see that when s is less than 65, f = 1 has a higher predicted probability than f = 0 and the reverse is true for higher values. We could have intuited this from the intersection point of the lines in Figure 2. We can also see that the greatest difference in predicted probabilities between s = 40 and s = 45. This may be very interesting to a researcher, but he will never arrive at this conclusion by glancing at the output table and the logit scale predictions. By generating this plot, we have made obvious a result that was previously obscured.

Figure 3.
Predicted values by categorical variable levels, probability scale

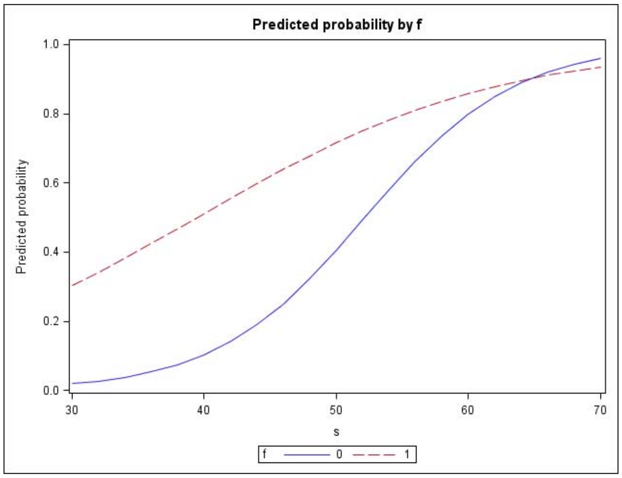
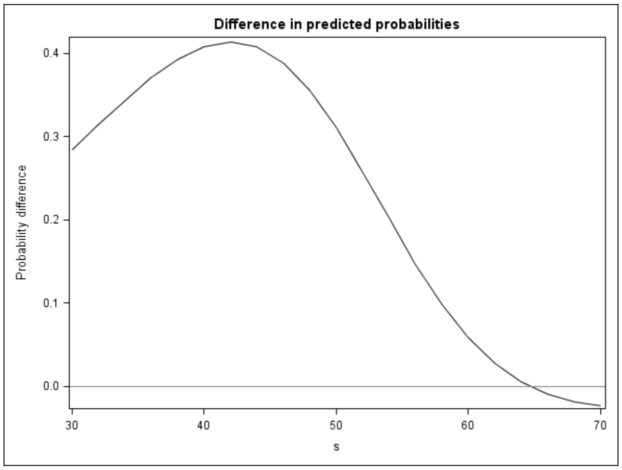


Figure 4.
Difference in predicted probabilities



BOOTSTRAP CONFIDENCE INTERVALS

To generate bootstrap confidence intervals about the differences plotted above, we can use PROC SURVEYSELECT to sample with replacement from our original dataset. We can then model BY REPLICATE and repeat the prediction step used earlier. Using the BY statement, we save logistic regression results for each sampled dataset and PROC PLM allows the user to apply each replicate's model estimates to a specified dataset.

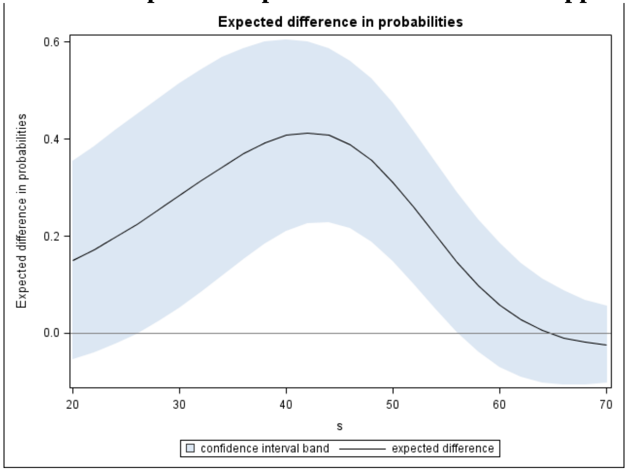
```
proc surveyselect data= logitcatcon out=_bsample_
  seed = 1234 method = urs
  samprate = 1 outhits rep = 100;
run;

proc logistic data = _bsample_ descending;
  by replicate;
  model y = f|s;
  store sasuser.l2;
run;

proc plm source=sasuser.l2;
  score data=test out=_temp2_ predicted / ilink;
```

```
run;

data _temp2_; set _temp2_;
  by replicate s;
  retain diff;
  if f = 0 then diff = predicted;
  if f = 1 then do;
    diff = predicted - diff;
    output;
  end;
run;

proc sql;
  create table _temp3_ as
  select  distinct
    mean(diff) as diff,
    std(diff) as std_diff,
    s
  from _temp2_
  group by s;
quit;

proc sql;
  create table outdata as
  select a.s as con_at, a.diff as est_diff, a.diff - b.diff as bias,
    a.diff - 1.96*std_diff as lb,
    a.diff + 1.96*std_diff as ub
  from _temp1_ as a, _temp3_ as b
  where a.s = b.s
  order by con_at;
quit;
```

In generating the point-wise confidence intervals, we are assuming normally distributed errors about the estimated difference. We also calculate the bias of the original model estimates as their distance from the mean of the estimated differences from the bootstrap samples.

```
proc sgplot data = outdata;
  band x = con_at upper = ub lower = lb / transparency = .5
  legendlabel = "Confidence interval band";
  series x = con_at y = est_diff /
  legendlabel = "Expected difference";
  refline 0 / axis = y;
  yaxis label = "Expected difference in probabilities";
  xaxis label = "s";
  title "Expected difference in probabilities";
run;
```

Figure 5.
Difference in predicted probabilities with bootstrapped confidence intervals



ANOTHER EASY WAY OUT

If the interaction is going to be difficult to explain in the scale of interest, why not simply leave it out of the model? The model with the interaction omitted is, effectively, the same model with the interaction coefficient constrained to zero. In our example, the logit predictions for f = 1 and f=0 would be parallel lines. However, in the probability scale, the difference between the prediction curves changes with values of s.

Figure 6.
Predicted values by categorical variable levels, logit scale, no interaction in model

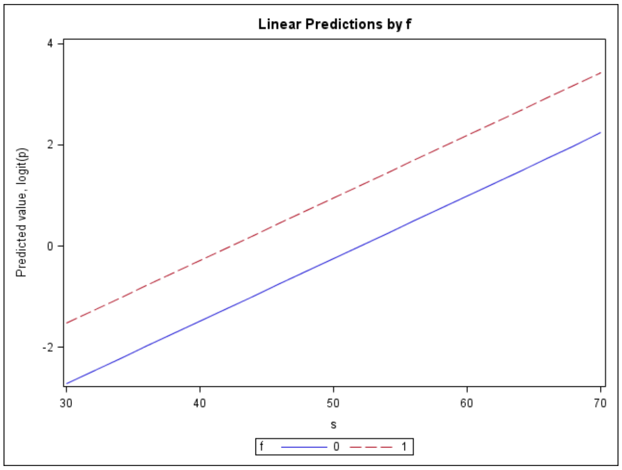
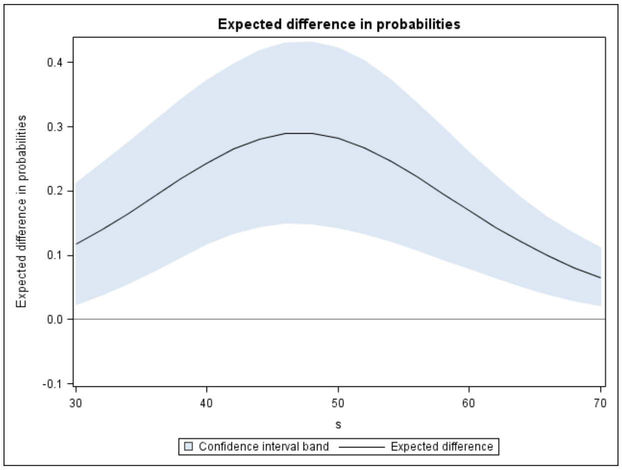


Figure 7.
Difference in predicted probabilities with bootstrapped confidence intervals, no interaction in model



Thus, the model defines different relationships between s and the outcome for the two levels of f. So even though we have no interaction in the logit scale, we have an interaction in the probability scale! It is possible for an interaction to appear significant in the logit scale, but not in a non-linearly transformed scale. And similarly, it is possible for an interaction to appear non-significant or be constrained to zero in the logit scale, but appear significant in the transformed scale. A researcher must consider this when both conceptualizing and interpreting a model. Simply leaving the interaction out of the GLM model does not mean you will avoid seeing one in the transformed scale.

STATISTICAL INFERENCE AND INTERACTIONS IN GLM

While we have created several plots and presented a fair amount of code, we have not performed any formal tests or statistical inference in this paper. Statistical inference based on interaction effects in GLM is a topic of ongoing discussion that we will leave to others to tackle. Regardless of the formal interpretation of the model, we believe there is value in understanding the relationships the model defines. We presented the bootstrap confidence intervals in Figure 5 to further illustrate how unintuitive the interaction effect is when viewed in the probability scale, not necessarily as a way of showing where the effect is or is not significant.

If we were to view the SAS parameter estimates from our initial logistic regression, we would see that the interaction coefficient is positive and is statistically different from zero at alpha = .05. Too often, model analysis ends here. The steps we describe allow a researcher to see the more complex picture of the interaction. Visualizing interaction effects requires little effort but allows the researcher to more fully understand the model and better evaluate their hypotheses.

CONCLUSION

Regression output from GLM models with interaction effects are difficult to mentally translate into the scale of interest. Generating transformed model predictions and producing graphics to illustrate interaction effects in the scale of interest allow researchers to more fully understand their models.

REFERENCES

Ai, C.R. and Norton E.C. 2003. Interaction terms in logit and probit models. *Economics Letters* 80(1): 123-129.

Cassell, David L (2007), "Don't be Loopy: Re-Sampling and Simulating the SAS Way". *Proceedings of the SAS Global Forum 2007 Conference,* Cary, NC: SAS Institute Inc.

Greenland, S. and Rothman, K.J. 1998. *Modern Epidemiology, 2nd Ed.* Phildephlia: Lippincott Williams and Wilkins.

Norton, E.C., Wang, H., and Ai, C. 2004 Computing interaction effects and standard errors in logit and probit models. *Stata Journal* 4(2): 154-167.

Tobias, R. and Cai, W. (2010), "Introducing PROC PLM and Postfitting Analysis for Very General Linear Models in SAS/STAT 9.22". *Proceedings of the SAS Global Forum 2010 Conference,* Cary, NC: SAS Institute Inc.

ACKNOWLEDGMENTS

We thank our colleagues Phil Ender, Rose Medeiros, and Christine Wells, as well as the many UCLA researchers who have challenged us with their non-linear interactions.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Name: Xiao Chen
Enterprise: UCLA Academic Technology Services
E-mail: atsstat@ucla.edu
Web: http://www.ats.ucla.edu/stat/

Name: Brigid Wilson
Enterprise: UCLA Academic Technology Services/UCLA Department of Statistics
E-mail: brigid@stat.ucla.edu
Web: http://www.stat.ucla.edu/~brigid/