

predict and adjust with logistic regression

Maarten L. Buis
 Department of Social Research Methodology
 Vrije Universiteit Amsterdam
 Amsterdam, the Netherlands
 m.buis@fsw.vu.nl

Abstract. Within Stata there are two ways of getting average predicted values for different groups after an estimation command: `adjust` and `predict`. After OLS regression (`regress`) these two ways give the same answer. However, after logistic regression the average predicted probabilities differ. This article discusses where that difference comes from and the consequent subtle difference in interpretation.

Keywords: st0001, adjust, predict, logistic regression

1 Introduction

A useful way of interpreting the results from a regression model is to compare predicted values from different groups. Within Stata both `adjust` and `predict` can be used after an estimation command to set up values at which predictions are desired and then display those predictions in a tabular form. In a Stata FAQ Poi (2002) showed the following example:

```
. sysuse auto, clear
(1978 Automobile Data)
. regress mpg weight length foreign
(output omitted)
. adjust, by(rep78)
```

Dependent variable: mpg Command: regress
 Variables left as is: weight, length, foreign

Repair Record 1978	xb
1	21.3651
2	19.3989
3	19.9118
4	21.86
5	24.9181

Key: xb = Linear Prediction

In this example `adjust` shows the average predicted millage for different states of repair. To show that that is exactly what `adjust` does, Poi actually computes the

predicted millage for each observation with `predict` and then shows that the averages for each state of repair corresponds exactly with the output from `adjust`.

```
. predict yhat, xb
. tabstat yhat, statistics(mean) by(rep78)
Summary for variables: yhat
      by categories of: rep78 (Repair Record 1978)
```

rep78	mean
1	21.36511
2	19.39887
3	19.91184
4	21.86001
5	24.91809
Total	21.20081

However, when the same procedure is applied to predicted probabilities from logistic regression, than the average predicted probabilities no longer match the output from `adjust`. The aim of this article is to explain where that difference comes from, and to discuss the resulting difference in interpretation of the results from `adjust` and `predict`.

```
. use http://www.stata-press.com/data/r9/lbw,clear
(Hosmer & Lemeshow data)
. gen black = race==2
. gen other = race==3
. logit low age lwt black other smoke,nolog
(output omitted)
. predict p
(option p assumed; Pr(low))
. tabstat p, statistics(mean) by(ht)
Summary for variables: p
      by categories of: ht (has history of hypertension)
```

ht	mean
0	.3154036
1	.2644634
Total	.3121693

(Continued on next page)

```
. adjust, pr by(ht)
```

```
Dependent variable: low      Command: logit
Variables left as is: age, lwt, smoke, black, other
```

has history of hypertens ion	pr
0	.291936
1	.251055

Key: pr = Probability

2 Computing predicted probabilities involve a nonlinear transformation

The key in understanding this difference is noticing that getting predicted probabilities from logistic regression requires a nonlinear transformation. In the example `logit` modeled the probability of getting a child with low birthweight according to equation 1 below.

$$\Pr(\text{low} = 1) = \frac{e^{xb}}{1 + e^{xb}} \quad (1)$$

whereby xb is usually called the linear predictor and is given by:

$$xb = \beta_0 + \beta_1 \text{age} + \beta_2 \text{lwt} + \beta_3 \text{black} + \beta_4 \text{other} + \beta_5 \text{smoke}$$

Once the model is estimated, computing the predicted probabilities involves two steps: First the predicted values for the linear predictor are calculated. Next the linear predictor is transformed to the probability metric using equation 2. Predicted values are identified by a $\hat{}$ on top of their symbol.

$$\widehat{\Pr} = \frac{e^{\widehat{xb}}}{1 + e^{\widehat{xb}}} \quad (2)$$

The difference between `predict` and `adjust` is that `predict` first applies the transformation to the linear predictor and then computes the mean, while `adjust` first computes the mean of the linear predictor and then applies the transformation (see: [R] **adjust**). In order to see why this matters it is instructive to first look at a special case where it does not matter. This is the case if xb is distributed symmetrically around

0. This is shown in figure 1. It shows that the transformation ‘squeezes’ the values of xb on the unit interval. Furthermore, it squeezes values further away from zero harder than values closer to zero. So in the transformed metric the smallest value became less extreme because it got squeezed a lot. Remember that extreme values influence the mean more than less extreme values. So, the lowest value exerts less influence on the mean in the transformed probability metric than in the original linear predictor metric. However, the change in mean due to the loss of influence of the lowest value was exactly balanced by the change in mean due to the loss of influence from the largest value, since the linear predictor was symmetrically distributed around zero.

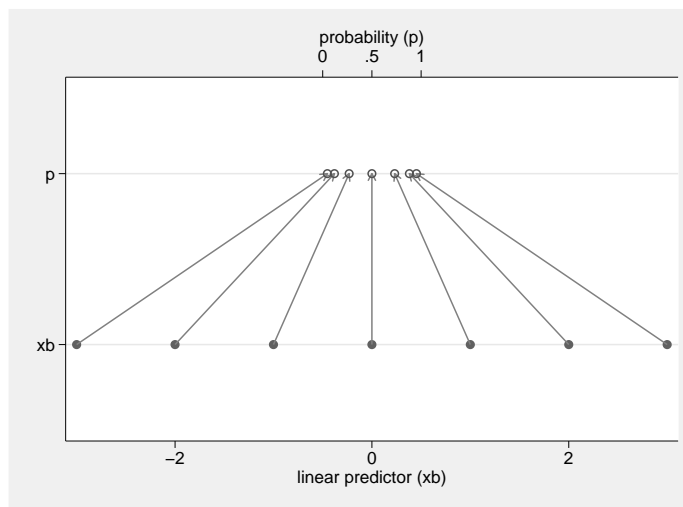
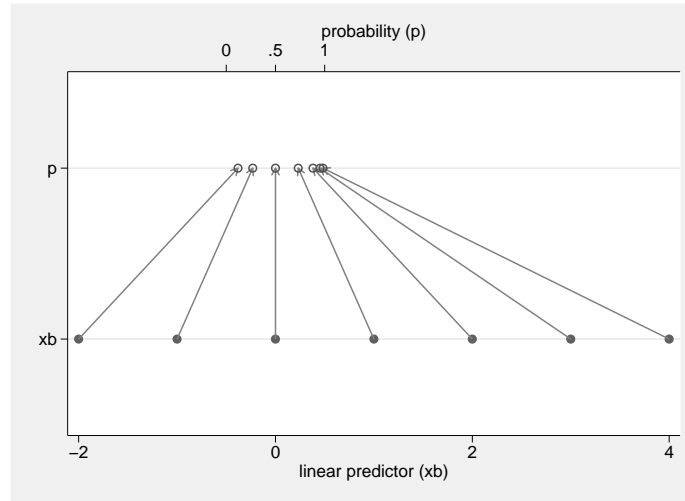


Figure 1: Logit transformation if xb is symmetric around 0

The likelihood that a real model on real data will yield a distribution of linear predictors that are symmetric around zero is extremely small. Figure 2 shows what happens if the distribution is asymmetric around 0. The loss in influence for the largest values is not balanced by the loss of influence for the smallest values. As a consequence the largest values exert more influence on the mean in the original linear predictor metric than in the transformed probability metric. So, in the case of figure 2 those who first compute the mean and then transform (use `adjust`) will find a larger probability than those who first transform and then compute the mean (use `predict`).

Figure 2: Logit transformation if xb is asymmetric around 0

3 What does this difference mean?

In order to make sense of this difference it is helpful to see that the average linear predictor is the linear predictor for someone with average values on its explanatory variables. Equations 3 till 6 show why.

$$\overline{xb}_k = \frac{\sum_k \beta_0 + \beta_1 x_1 + \beta_2 x_2}{N_k} \quad (3)$$

$$= \frac{\sum_k \beta_0}{N_k} + \frac{\sum_k \beta_1 x_1}{N_k} + \frac{\sum_k \beta_2 x_2}{N_k} \quad (4)$$

$$= \frac{N_k \beta_0}{N_k} + \beta_1 \frac{\sum_k x_1}{N_k} + \beta_2 \frac{\sum_k x_2}{N_k} \quad (5)$$

$$= \beta_0 + \beta_1 \overline{x}_{1k} + \beta_2 \overline{x}_{2k} \quad (6)$$

Say we have two explanatory variables, imaginatively called x_1 and x_2 , and we want to compute the mean linear predictor for group k . Equation 3 is just the definition of that mean. Equation 4 shows that that fraction can be broken up. Equation 5 is based on the fact that the β s are constant, so they can be moved outside the summation sign. And finally equation 6 is again based on the definition of the mean. Note that \overline{x}_{1k} and \overline{x}_{2k} are the means for group k only, not the overall means. `adjust` does have facilities to fix (some of) the explanatory variables at their grand mean, or other values, but that is not being discussed here.

The consequence of this is that there is a subtle difference in interpretation between the results of `predict` and `adjust`. If we get back to our logistic regression example

and look at someone with hypertension, than `predict` will give us the average predicted probability for someone with hypertension while `adjust` will give us the predicted probability for someone with average values on age, lwt, black, other, and smoke for someone with hypertension. It is the difference between a typical predicted probability for someone within a group and the predicted probability for someone with typical values on the explanatory variables for someone within that group. Furthermore, this is by no means unique to logistic regression. The same difference in interpretation between `predict` and `adjust` also applies to for instance probit regression, multinomial logistic regression, ordered logistic regression, and any generalized linear model (glm) with a link function other than the identity function.

4 Acknowledgement

I thank Fred Wolfe whose question on Statalist provided the initial spark that resulted in this article.

5 References

Poi, B. P. 2002. Predict and adjust. <http://www.stata.com/support/faqs/stat/adjust.html>.