The Stata Journal (*yyyy*) *vv*, Number *ii*, pp. 1–4

# Stata tip 87: Interpretation of interactions in non-linear models

Maarten L. Buis
Department of Sociology
Tübingen University
Tübingen, Germany
maarten.buis@uni-tuebingen.de

When estimating a non-linear model such as [R] **logit** or [R] **poisson**, we often have two options when it comes to interpreting the regression coefficients: compute some form of marginal effect; or exponentiate the coefficients, which will give us an odds ratio or incidence-rate ratio. The marginal effect is an approximation of how much the dependent variable is expected to increase or decrease for a unit change in an explanatory variable: that is, the effect is presented on an additive scale. The exponentiated coefficients give the ratio by which the dependent variable changes for a unit change in an explanatory variable: that is, the effect is presented on a multiplicative scale. An extensive overview is given by Long and Freese (2006). Sometimes we are also interested in how the effect of one variable changes when another variable changes, namely, the interaction effect. As there is more than one way in which we can define an effect in a non-linear model, there must also be more than one way in which we can define an interaction effect. This tip deals with how to interpret these interaction effects when we want to present effects as odds ratios or incidence-rate ratios. This can be an attractive alternative to interpreting interactions effects in terms of marginal effects.

The motivation for this tip is that there has been much discussion on how to interpret interaction effects when we want to interpret them in terms of marginal effects (Ai and Norton 2003; Norton et al. 2004; Cornelißen and Sonderhof 2009). (A separate concern about interaction effects in non-linear models whicht is often mentioned is the possible influence of unobserved heterogeneity on these estimates (e.g. Williams 2009), but I will not deal with that potential problem.) These authors point out a common mistake, interpreting the first derivative of the multiplicative term between two explanatory variables as the interaction effect. The problem with this is that we want the interaction effect between two variables ($x_1$ and $x_2$) to represent how much the effect of $x_1$ changes for a unit change in $x_2$. The effect of $x_1$, in the marginal effects metric, is the first derivative of the expected value of the dependent variable ($\mathrm{E}(y)$) with respect to $x_1$, which is an approximation of how much $\mathrm{E}(y)$ changes for a unit change in $x_1$. The interaction effect should thus be the cross partial derivative of $\mathrm{E}(y)$ with respect to $x_1$ and $x_2$, that is, an approximation of how much the derivative of $\mathrm{E}(y)$ with respect to $x_1$ changes for a unit change in $x_2$. In non-linear models this is typically different from the first derivative of $\mathrm{E}(y)$ with respect to the multiplicative term $x_1 \times x_2$. This is where programs like `inteff` by Norton et al. (2004) and `inteff3` by Cornelißen and Sonderhof (2009) come in.

Fortunately, we can interpret interactions without referring to any additional program by presenting effects as multiplicative effects (e.g. odds ratio, incidence-rate ratios,

 st0001

hazard ratios). However, the marginal effects and multiplicative effects answer subtly different questions, and thus it is a good idea to have both tools in your toolbox.

The interpretation of results is best explained using an example. Here we study whether the effect of having a college degree (`collgrad`) on the odds of obtaining an 'high' job (`high_occ`) differs between black and white women.

```
. sysuse nlsw88, clear
(NLSW, 1988 extract)
. gen byte high_occ = occupation < 3 if occupation < .
(9 missing values generated)
. gen byte black = race == 2 if race < .
. drop if race == 3
(26 observations deleted)
. gen byte baseline = 1
. logit high_occ black##collgrad baseline, or nocons nolog
Logistic regression                               Number of obs   =       2211
                                                  Wald chi2(4)    =     504.62
Log likelihood = -1199.4399                       Prob > chi2     =     0.0000
```

| high_occ | Odds Ratio | Std. Err. | z | P>\|z\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| 1.black | .4194072 | .0655069 | -5.56 | 0.000 | .3088072 | .5696188 |
| 1.collgrad | 2.465411 | .293568 | 7.58 | 0.000 | 1.952238 | 3.113478 |
| black# collgrad | | | | | | |
| 1 1 | 1.479715 | .4132536 | 1.40 | 0.161 | .8559637 | 2.558003 |
| baseline | .3220524 | .0215596 | -16.93 | 0.000 | .2824512 | .3672059 |

If we were to interpret these results in terms of marginal effects, we would typically look at the effect of the explanatory variables on the probability of attaining a high job. However, this example uses a `logit` model together with the `or` option, so the dependent variable is measured in the odds metric rather than the probability metric. Odds have a bad reputation for being hard to understand, but they are just the expected number of people with a high job for every person with a low job. For example, the baseline odds — the odds of having a high job for white women without a college degree — is .32, meaning that within this category we expect to find .32 women with a high job for every woman with a low job. Note that the trick I have used to display the baseline odds is discussed in an earlier tip (Newson 2003). The odds ratio for `collgrad` is 2.47, which means that the odds of having a high job is 2.47 times higher for women with a college degree. There is also an interaction effect between `collgrad` and `black`, so this effect of having a college degree refers to white women. The effect of college degree for black women is 1.48 times the effect of `collgrad` for white women. So the interaction effect tells by how much the effect of `collgrad` differs between black and white women, but does so in multiplicative terms. The results also show that this interaction is not significant.

This example points up the difference between marginal effects and multiplicative effects. Now we can compute the marginal effect as the difference between the expected odds of women with and without a college degree, rather than as the derivative of the expected odds with respect to `collgrad`. The reason for computing the marginal effect as a difference is that `collgrad` is a categorical variable, so this discrete difference corresponds more closely with what would actually be observed. Although it is a slight abuse of terminology, I will continue to call it the marginal effect. The `margins` command below shows the odds of attaining a high job for every combination of `black` and `collgrad`. The odds of attaining a high job for white women without a college degree is .32, while the odds for white women with a college degree is .79. The marginal effect of `collgrad` for white women is thus .47. The marginal effect of `collgrad` for black women is only .36. The marginal effect of `collgrad` is thus larger for white women than for black women, while the multiplicative effect of `collgrad` is larger for black women.

```
. margins , over(black collgrad) expression(exp(xb())) post
Predictive margins                              Number of obs   =       2211
Model VCE    : OIM

Expression   : exp(xb())
over         : black collgrad

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      black#
    collgrad |
        0 0  |   .3220524   .0215596    14.94   0.000     .2797964    .3643084
        0 1  |   .7939914    .078188    10.15   0.000     .6407457    .9472371
        1 0  |   .1350711   .0190606     7.09   0.000      .097713    .1724292
        1 1  |   .4927536   .1032487     4.77   0.000       .29039    .6951173
------------------------------------------------------------------------------

. lincom 0.black#1.collgrad - 0.black#0.collgrad
 ( 1)  - 0bn.black#0bn.collgrad + 0bn.black#1.collgrad = 0

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |    .471939    .081106     5.82   0.000     .3129742    .6309038
------------------------------------------------------------------------------

. lincom 1.black#1.collgrad - 1.black#0.collgrad
 ( 1)  - 1.black#0bn.collgrad + 1.black#1.collgrad = 0

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   .3576825   .1049933     3.41   0.001     .1518994    .5634656
------------------------------------------------------------------------------
```

The reason for this difference is that the multiplicative effects are relative to the baseline odds in their own category. In this example these baseline odds differ substantially between black and white women: for white women without a college degree we expect to find .32 women with a high job for every woman with a low job, while for black women without a college degree we expect to find only .14 women with a high job for every woman with a low job. So, even though the increase in odds as a result of getting a college degree is higher for the white women than for black women, this increase as a percentage of the baseline value is less for white women than for black women. The multiplicative effects control in this way for differences between the groups in baseline odds. However, notice that marginal and multiplicative effects are both accurate representations of the effect of a college degree. Which effect one wants to report depends on the substantive question, whether or not one wants to control for differences in the baseline odds.

The example here is relatively simple with only binary variables and no controlling variables. However, the basic argument still holds when using continuous variables and when controlling variables are added. Moreover, the argument is not limited to results obtained from [R] **logit**. It applies to all forms of multiplicative effects, and so, for example, to odds ratios from other models such as [R] **ologit** and [R] **glogit**; relative risk ratios ([R] **mlogit**); incidence-rate ratios (for example [R] **poisson**, [R] **nbreg**, and [R] **zip**); or hazard ratios (for example [ST] **streg** and [R] **cloglog**).

## Acknowledgment

I thank Richard Williams for helpful comments.

## References

Ai, C., and E. Norton. 2003. Interaction terms in logit and probit models. *Economics Letters* 80: 123–129.

Cornelißen, T., and K. Sonderhof. 2009. Partial effects in probit and logit models with a triple dummy-variable interaction term. *The Stata Journal* 9: 571–583.

Long, J., and J. Freese. 2006. *Regression Models for Categorical Dependent Variables Using Stata*. 2nd ed. College Station, TX: Stata Press.

Newson, R. 2003. Stata tip 1: The eform() option of regress. *The Stata Journal* 3: 445.

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

Williams, R. 2009. Using heterogenous choice models to compare logit and probit coefficients across groups. *Sociological Methods & Research* 37: 531–559.