The Stata Journal (2022)
22, Number 2, pp. 460–464 DOI: 10.1177/1536867X221106437

Stata tip 146: Using margins after a Poisson regression model to estimate the number of events prevented by an intervention

Milena Falcaro
King's College London
London, U.K.
milena.falcaro@kcl.ac.uk

Roger B. Newson
King's College London
London, U.K.
roger.newson@kcl.ac.uk

Peter Sasieni
King's College London
London, U.K.
peter.sasieni@kcl.ac.uk

After fitting a Poisson regression model to evaluate the effect of an intervention in a cohort study, one might be interested in estimating the number of events prevented by the intervention (assuming the observed associations are causal). This can be derived as the difference in the intervention group between the predicted number of events under the counterfactual (no intervention) and the factual (intervention) scenarios. One could use the `predict` command to obtain the predicted number of events under the two scenarios and then sum up the differences, but this approach would not be convenient for several reasons. One would need to change the intervention variable to get the counterfactual predicted values, and the confidence intervals would not be readily available (`bootstrap` or `jackknife` could be used, but this could be particularly time consuming if the dataset is large).

We here suggest using the `margins` command. Its use, however, is not straightforward for our specific problem because `margins` computes predictions for each observation (like `predict`) and then takes the average of these predicted values. For example, if our data are aggregated in years, `margins` will provide an average of the year-specific predictions. When `margins` is applied over N records and $\widehat{P}_i$ is the predicted value for the ith observation ($i = 1, \ldots, N$), the result is simply the average of these predicted values, that is, $\left(\sum_{i=1}^{N} \widehat{P}_i\right)/N$. If we want `margins` to calculate the sum of the predictions instead of the mean, we can multiply each observation-specific prediction by the number of observations (that is, N), and the result of `margins` will be $\left(\sum_{i=1}^{N} N\widehat{P}_i\right)/N = \sum_{i=1}^{N} \widehat{P}_i$.

Let's consider a simple example using simulated data. Specifically, we use a Poisson distribution to generate a variable, `cases`, containing the number of events of interest (for example, the number of cancer cases) as a function of an intervention indicator (`trt` = `1` if treated, `0` otherwise); two covariates (`x1` and `x2`); and an offset (`pyar` = person-years at risk).

```
. clear
. set seed 12345
. set obs 1000
. generate x1=runiform(50,100)
. generate x2=rbinomial(1,0.3)
. generate trt=rbinomial(1,0.5)
. generate pyar=runiformint(200,400)
. generate m=exp(0.01-0.2*trt-0.05*x1+0.8*x2 + ln(pyar))
. generate cases=rpoisson(m)
```


We then fit a Poisson regression model:

```
. poisson cases i.trt x1 i.x2, exp(pyar)
Iteration 0:   log likelihood = -2452.9776
Iteration 1:   log likelihood = -2452.9125
Iteration 2:   log likelihood = -2452.9125
Poisson regression                                      Number of obs =   1,000
                                                        LR chi2(3)    = 6654.25
                                                        Prob > chi2   =  0.0000
Log likelihood = -2452.9125                             Pseudo R2     =  0.5756

------------------------------------------------------------------------------
       cases | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
       1.trt |  -.1925239   .0188441   -10.22   0.000    -.2294576   -.1555902
          x1 |  -.0497789    .000752   -66.19   0.000    -.0512529    -.048305
        1.x2 |   .7863367   .0188159    41.79   0.000     .7494582    .8232151
       _cons |   .0105769   .0514323     0.21   0.837    -.0902284    .1113823
    ln(pyar) |          1  (exposure)
------------------------------------------------------------------------------
```

To obtain an estimate of the number of events prevented by the intervention and its 95% confidence interval, the `margins` command will need to include the following (see [R] **margins** for more details):

- an `if` qualifier (that is, `if trt==1`) or the corresponding `subpop()` option (the latter must be used if the `vce(unconditional)` option is specified too);
- two `at()` options: one for the factual scenario (that is, `at((asobserved) _all)`) and one for the counterfactual scenario (that is, `at(trt=0)`);
- `expression(predict(n)*r)`, where `r` is the size of the group of observations over which the `margins` command averages the predictions (it is here retrieved from the two command lines `count if trt==1 & e(sample)==1` and `scalar r=r(N)`); and
- the `pwcompare` option.

Hence, the commands and results are as follows:

```
. count if trt==1 & e(sample)==1
  504

. scalar r=r(N)

. margins, at((asobs) _all) at(trt=0) exp(predict(n)*r) subpop(if trt==1)
> pwcompare

Pairwise comparisons of predictive margins            Number of obs   = 1,000
Model VCE: OIM                                        Subpop. no. obs =   504

Expression: predict(n)*r
1._at: (asobserved)
2._at: trt   = 0

──────────────────────────────────────────────────────────────
             │                Delta-method         Unadjusted
             │   Contrast   std. err.     [95% conf. interval]
─────────────┼────────────────────────────────────────────────
         _at │
     2 vs 1  │   1082.121    105.661      875.0296    1289.213
──────────────────────────────────────────────────────────────
```

The above results show that the intervention is estimated to have prevented 1,082 (95% CI: 875 to 1,289) cancer cases in our sample. Had we used the above `margins` command without the `expression()` option, we would have obtained the average of the observation-specific predicted number of events:

```
. margins, at((asobs) _all) at(trt=0) subpop(if trt==1) pwcompare

Pairwise comparisons of predictive margins            Number of obs   = 1,000
Model VCE: OIM                                        Subpop. no. obs =   504

Expression: Predicted number of events, predict()
1._at: (asobserved)
2._at: trt   = 0

──────────────────────────────────────────────────────────────
             │                Delta-method         Unadjusted
             │   Contrast   std. err.     [95% conf. interval]
─────────────┼────────────────────────────────────────────────
         _at │
     2 vs 1  │   2.147066   .2096448       1.73617    2.557962
──────────────────────────────────────────────────────────────
```

To better understand the above output, one can generate the variables (here called `pred1` and `pred2`) containing the observation-specific predictions for the two scenarios and then look at their means. The `pwcompare` option will be omitted because it is not allowed when the `generate()` option is specified too.

```
. margins, at((asobs) _all) at(trt=0) subpop(if trt==1) generate(pred)

Predictive margins                                 Number of obs   = 1,000
Model VCE: OIM                                     Subpop. no. obs =   504

Expression: Predicted number of events, predict()
1._at: (asobserved)
2._at: trt   = 0

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
         _at |
          1  |    10.1131   .1416533    71.39   0.000      9.83546    10.39073
          2  |   12.26016   .1545487    79.33   0.000     11.95725    12.56307
------------------------------------------------------------------------------

. sum pred1 pred2
    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
       pred1 |        504     10.1131    8.375062   1.288584   45.00473
       pred2 |        504    12.26016    10.15313   1.562158   54.55948
```

If we calculate the difference between the means of `pred2` (counterfactual scenario) and `pred1` (factual scenario), we obtain the value reported in the above `margins` command, where we omitted both the `expression()` and `pwcompare` options ($12.26016 - 10.1131 = 2.14706$). If we now generate the difference between `pred2` and `pred1` (that is, `generate diff=pred2-pred1`) and use the `total` command, we will obtain the point estimate reported by `margins` with the `expression()` and `pwcompare` options.

```
. generate diff=pred2-pred1
(496 missing values generated)

. total pred1 pred2 diff

Total estimation                          Number of obs = 504

--------------------------------------------------------------
             |      Total   Std. err.     [95% conf. interval]
-------------+------------------------------------------------
       pred1 |       5097   188.0197      4727.599    5466.401
       pred2 |   6179.121   227.9373      5731.295    6626.948
        diff |   1082.121   39.91762      1003.696    1160.547
--------------------------------------------------------------
```

Extensions to interventions with two or more levels (for example, `0` = no treatment, `1` = low-dosage treatment, `2` = high-dosage treatment) or other counterfactual scenarios would be straightforward. For example, if we want to estimate how many fewer cases we would have observed in the nonintervention group (that is, `trt` = `0`) if everybody had received the treatment, then we would specify the following:

```
. quietly poisson cases i.trt x1 i.x2, exp(pyar)

. count if trt==0 & e(sample)==1
  496

. scalar s=r(N)

. margins, at((asobs) _all) at(trt=1) exp(predict(n)*s) subpop(if trt==0)
> pwcompare

Pairwise comparisons of predictive margins            Number of obs   = 1,000
Model VCE: OIM                                        Subpop. no. obs =   496

Expression: predict(n)*s
1._at: (asobserved)
2._at: trt   = 1

─────────────────────────────────────────────────────────────
             │              Delta-method         Unadjusted
             │   Contrast   std. err.     [95% conf. interval]
─────────────┼───────────────────────────────────────────────
         _at │
     2 vs 1  │  -1106.267   107.9831      -1317.91   -894.6243
─────────────────────────────────────────────────────────────
```

Thus, our model estimates that if everyone in the nonintervention group had been administered the treatment, there would have been 1,106 (95% CI: 895 to 1318) fewer cancer cases. Note that the contrast is negative, corresponding to fewer cases had everyone been treated. This is because we are comparing the counterfactual scenario represented by `at(trt=1)` (that is, scenario 2 = untreated patients are treated) versus the factual scenario specified by `at((asobs) _all)` (that is, scenario 1 = untreated patients are untreated).

What is discussed in this Stata tip could also be extended to case–control studies by using inverse-probability-of-sampling weights to estimate absolute rates.

Acknowledgment

This work was supported by Cancer Research U.K. (grant number: C8162/A27047).