How can I manually generate the predicted counts from a ZIP or ZINB model based on the parameter estimates?

Authored by
stats writer

July 1, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I manually generate the predicted counts from a ZIP or ZINB model based on the parameter estimates?*. PSYCHOLOGICAL SCALES. Retrieved from https://scales.arabpsychology.com/?p=164893

The process of manually generating predicted counts from a ZIP or ZINB model involves using the parameter estimates obtained from the model to calculate the expected number of counts for a given dataset. This can be done by first determining the probability of a count being equal to zero, and then using a combination of the remaining probabilities and the estimated dispersion parameter to calculate the expected count for each observation. This method allows for a deeper understanding of the model's predictions and can be useful in validating the effectiveness of the model.

How can I manually generate the predicted counts from a ZIP or ZINB model based on the parameter estimates? | Stata FAQ

This page shows some examples on how to generate the predicted count from a zero-inflated Poisson or a zero-inflated negative binomial model based on the parameter estimates. Zero-inflated models allow us to model two processes simultaneously. Let's take ZIP as an example. Basically, zero outcome arises from two different processes. In one process, the outcome is always zero and in the other process, zero outcome, as well as other outcomes obey the Poisson process. With the two parts of the model, how do we generate the predicted count after running the

model? The examples demonstrate the steps to this end.

Example 1. Zero-inflated Poisson model with logit inflation model

webuse fish, clear
zip count persons livebait, inf(child camper) nolog

Zero-inflated Poisson regression Number of obs = 250
Nonzero obs = 108
Zero obs = 142

Inflation model = logit LR chi2(2) = 506.48
Log likelihood = -850.7014 Prob > chi2 = 0.0000

--

| Coef. Std. Err. z P>|z|
-------------+--
count |
persons | .8068853 .0453288 17.80 0.000 .7180424 .8957281
livebait | 1.757289 .2446082 7.18 0.000 1.277866 2.236713
_cons | -2.178472 .2860289 -7.62 0.000 -2.739078 -1.617865

-------------+--

inflate |

child | 1.602571 .2797719 5.73 0.000 1.054228 2.150913

camper | -1.015698 .365259 -2.78 0.005 -1.731593 -.2998038

_cons | -.4922872 .3114562 -1.58 0.114 -1.10273 .1181558

----

predict p

The variable p created above is the predicted count based on this model.

Now we show the steps to create the same p using the parameter

estimates. Basically, it has two parts, the model for the usual Poisson

process and the model for the process of zeros. Variable a1 below is

the linear prediction based on the first model and variable a2 is the

linear prediction for the second model which is a logit model by default.

Variable pzero is the predicted probability for being in the first

process which only produces zero count. Variable pcount is then the
predicted count based on the two processes.

```
gen a1 = -2.178472 + .8068853*persons + 1.757289*livebait
gen a2 = -.4922872 + 1.602571*child -1.015698*camper
gen pzero = exp(a2)/(1+exp(a2))
gen pcount = exp(a1)*(1-pzero) /*for logit model*/
sum p pcount

Variable | Obs Mean Std. Dev. Min Max
-------------+--------------------------------------------------------
p | 250 2.770999 3.269588 .079269 13.55015
pcount | 250 2.770997 3.269585 .0792689 13.55014
```

Example 2. Zero-inflated Poisson model with probit inflation model

The only difference between this example and the previous one is that the
inflation part in this one is modeled by probit model instead of logit
model.

```
webuse fish, clear
```

zip count persons livebait, inf(child camper) probit nolog

Zero-inflated Poisson regression Number of obs = 250
Nonzero obs = 108
Zero obs = 142

Inflation model = probit LR chi2(2) = 506.29
Log likelihood = -850.3968 Prob > chi2 = 0.0000

--

| Coef. Std. Err. z P>|z|

-------------+--

count |

persons | .8062521 .0453179 17.79 0.000 .7174306 .8950736

livebait | 1.755824 .2444357 7.18 0.000 1.276739 2.234909

_cons | -2.174616 .2858538 -7.61 0.000 -2.734879 -1.614353

-------------+--

inflate |

child | .9658273 .1576773 6.13 0.000 .6567855 1.274869

camper | -.6112131 .2146819 -2.85 0.004 -1.031982 -.1904442

_cons | -.295569 .1869964 -1.58 0.114 -.6620753 .0709372

----

predict p
gen a1 = -2.174616 + .8062521*persons + 1.755824 *livebait
gen a2 = -.295569 + .9658273*child -.6112131*camper
gen pzero = normal(a2) /*for probit model*/
gen pcount = exp(a1)*(1-pzero)
sum p pcount

Variable | Obs Mean Std. Dev. Min Max
-------------+--
p | 250 2.754194 3.272803 .0649889 13.53128
pcount | 250 2.754194 3.272803 .0649889 13.53128

Example 3. Zero-inflated negative binomial model with logit inflation model

Now we switch to zero-inflated negative binomial model. The way to
calculate the predicted values is exactly the same as for zero-inflated
Poisson models.

webuse fish, clear

zinb count persons livebait, inf(child camper) nolog

Zero-inflated negative binomial regression Number of obs = 250

Nonzero obs = 108

Zero obs = 142

Inflation model = logit LR chi2(2) = 82.23

Log likelihood = -401.5478 Prob > chi2 = 0.0000

| Coef. Std. Err. z P>|z|

-------------+--

count |

persons | .9742984 .1034938 9.41 0.000 .7714543 1.177142

livebait | 1.557523 .4124424 3.78 0.000 .7491503 2.365895

_cons | -2.730064 .476953 -5.72 0.000 -3.664874 -1.795253

-------------+--

inflate |

child | 3.185999 .7468551 4.27 0.000 1.72219 4.649808

camper | -2.020951 .872054 -2.32 0.020 -3.730146 -

.3117567

_cons | -2.695385 .8929071 -3.02 0.003 -4.44545 -.9453189

-------------+--

/lnalpha | .5110429 .1816816 2.81 0.005 .1549535 .8671323

-------------+--

alpha | 1.667029 .3028685 1.167604 2.380076

----

predict p

gen a1 = -2.730064 + .9742984*persons + 1.557523*livebait

gen a2 = -2.695385 + 3.185999*child -2.020951*camper

gen pzero = exp(a2)/(1+exp(a2))

gen pcount = exp(a1)*(1-pzero)

sum p pcount

Variable | Obs Mean Std. Dev. Min Max

-------------+--

p | 250 3.131795 4.189243 .0159387 15.11586

pcount | 250 3.131795 4.189243 .0159391 15.11586

Example 4. Zero-inflated Poisson model with logit inflation model again:
general setup

In previous examples, we have manually generated these variables using the parameter estimates. In this example, we make use of the Stata's stored matrix for parameter coefficients. This is the general and more useful approach in practice.

webuse fish, clear
zip count persons livebait, inf(child camper) nolog

Zero-inflated Poisson regression Number of obs = 250
Nonzero obs = 108
Zero obs = 142

Inflation model = logit LR chi2(2) = 506.48
Log likelihood = -850.7014 Prob > chi2 = 0.0000

--

| Coef. Std. Err. z P>|z|

-------------+--

count |

persons | .8068853 .0453288 17.80 0.000 .7180424 .8957281

livebait | 1.757289 .2446082 7.18 0.000 1.277866 2.236713

_cons | -2.178472 .2860289 -7.62 0.000 -2.739078 -1.617865

-------------+--

inflate |

child | 1.602571 .2797719 5.73 0.000 1.054228 2.150913

camper | -1.015698 .365259 -2.78 0.005 -1.731593 -.2998038

_cons | -.4922872 .3114562 -1.58 0.114 -1.10273 .1181558

--

predict p

matrix list e(b)

e(b)

count: count: count: inflate: inflate: inflate:

persons livebait _cons child camper _cons

y1 .80688527 1.7572894 -2.1784716 1.6025705 -1.0156983 -.49228716

gen a1 = _b + _b*persons + _b*livebait

gen a2 = _b + _b*child +_b*camper

gen pzero = exp(a2)/(1+exp(a2))

gen pcount = exp(a1)*(1-pzero) /*for logit model*/

sum p pcount

Variable | Obs Mean Std. Dev. Min Max

-------------+--

p | 250 2.770999 3.269588 .079269 13.55015

pcount | 250 2.770999 3.269588 .079269 13.55015