How can I obtain Somers' D after performing a logistic regression in Stata?

Authored by

stats writer

July 1, 2024

To obtain Somers' D after performing a logistic regression in Stata, one can use the "somersd" command. This command takes the predicted probability and the outcome variable from the logistic regression as input and calculates Somers' D, which is a measure of rank correlation between the two variables. It also provides a confidence interval and a p-value for the calculated D value. This allows for an assessment of the strength and significance of the relationship between the predicted probability and the outcome variable in the logistic regression model. The "somersd" command is a useful tool for evaluating the predictive ability of logistic regression models in Stata.

How can I get a Somers' D after logistic regression in Stata? | Stata FAQ

Recently we received the following question: The output from SAS proc logistic gives a values for Somers' D. How can I get this using Stata?

The Somers' D, in logistic regression, provides an estimate of the rank correlation of the observed binary response variable and the predicted probabilities. Thus, it can be used as an indicator of model fit.

Its not difficult to get a Somers' D in Stata once you download the user contributed program somersd written by Roger Newson. To get the program just type, ssc install somersd, in Stata's command window and follow the prompts to download the program. Alternatively, you can type search somersd and follow the prompts (see How can I use the search command to

search for programs and get additional help? for more information about using search).

Once you have downloaded somersd, run your logistic regression, compute the predicted probability and then run somersd with your binary response variable and the predicted probability.

Example

We will use the hsb2 dataset for this example. It doesn't have a good binary response variable so we will create one by dichotomizing write at the value 60 and calling it honors (for honors English). Then we will use it in a logistic regression with read, female and prog (the type of program each student is in) as predictors. After the logit command we will use predict to get the predicted probabilities.

```
use https://stats.idre.ucla.edu/stat/data/hsbdemo, clear

logit honors read female i.prog

Iteration 0: log likelihood = -115.64441
Iteration 1: log likelihood = -86.845312
Iteration 2: log likelihood = -84.560995
```

```
Iteration 3: log likelihood = -84.542357
Iteration 4: log likelihood = -84.542348
Iteration 5: log likelihood = -84.542348

Logistic regression Number of obs = 200
LR chi2(4) = 62.20
Prob > chi2 = 0.0000
Log likelihood = -84.542348 Pseudo R2 = 0.2689

------------------------------------------------------------------------------
honors | Coef. Std. Err. z P>|z|
-------------+----------------------------------------------------------------
read | .1352861 .0242218 5.59 0.000 .0878123 .18276
female | 1.08343 .4094357 2.65 0.008 .2809511 1.885909
|
prog |
2 | .5559416 .5053125 1.10 0.271 -.4344527 1.546336
3 | .0016408 .6611702 0.00 0.998 -1.294229 1.29751
|
_cons | -9.41691 1.481922 -6.35 0.000 -12.32142
-6.512397
------------------------------------------------------------------------------

predict pprob
```

Now that we have the predicted probabilities, pprob, we can run somersd.

somersd honors pprob

Somers' D with variable: honors
Transformation: Untransformed
Valid observations: 200

Symmetric 95% CI

--
| Jackknife
honors | Coef. Std. Err. z P>|z|
-------------+--
pprob | .6761648 .05778 11.70 0.000 .5629182 .7894115
--

The value of Somers' D is 0.676. We can compare this value of Somers' D to one from a model that uses only prog as a predictor.

logit honors i.prog

Iteration 0: log likelihood = -115.64441
Iteration 1: log likelihood = -107.7993

```
Iteration 2: log likelihood = -107.57279
Iteration 3: log likelihood = -107.5719
Iteration 4: log likelihood = -107.5719

Logistic regression Number of obs = 200
LR chi2(2) = 16.15
Prob > chi2 = 0.0003
Log likelihood = -107.5719 Pseudo R2 = 0.0698

------------------------------------------------------------------------------
honors | Coef. Std. Err. z P>|z|
-------------+----------------------------------------------------------------
prog |
2 | 1.206168 .4577746 2.63 0.008 .3089465 2.10339
3 | -.3007541 .5988045 -0.50 0.615 -1.474389 .8728812
|
_cons | -1.691676 .4113064 -4.11 0.000 -2.497822 -
.8855303
------------------------------------------------------------------------------

predict pprob2

(option pr assumed; Pr(honors))

somersd honors pprob2
```

Somers' D with variable: honors

Transformation: Untransformed

Valid observations: 200

Symmetric 95% CI

--

| Jackknife

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

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

pprob2 | .3228084 .0737763 4.38 0.000 .1782095 .4674073

--

As you can see the Somers' D for this model is much smaller than for the previous one.