How can I get out-of-sample predicted values?

Authored by

stats writer

June 30, 2024

To obtain out-of-sample predicted values, a statistical or machine learning model can be trained using a set of existing data. The trained model can then be used to make predictions on new, unseen data points. These predictions are referred to as out-of-sample predicted values, as they are not part of the original data used to train the model. This method allows for the evaluation of the model's performance on new data and can be used to make informed decisions or forecasts based on the predicted values.

How can I get out-of-sample predicted values? | SPSS FAQ

**Sometimes it is useful to get predicted values for cases that were not used
in the regression analysis. There are two ways to do this in SPSS.
Let's use the hsb2 dataset and create some missing values in a variable.
Specifically, we will set the first nine values in the variable write to
be missing. Then we will use write as our outcome variable in an OLS regression analysis.
Of course, the cases with missing values will not be used in the analysis, but
we can still get the predicted values for those cases.**

```
get file ='d:https://stats.idre.ucla.edu/wp-content/uploads/2016/02/hsb2.sav'.
```

sort cases by id.
if id lt 10 write = $sysmis.list write read math
/cases=from 1 to 12.

write read math

. 34.00 40.00
. 39.00 33.00
. 63.00 48.00
. 44.00 41.00
. 47.00 43.00
. 47.00 46.00
. 57.00 59.00
. 39.00 52.00
. 48.00 52.00
54.00 47.00 49.00
46.00 34.00 45.00
44.00 37.00 45.00

Number of cases read: 12 Number of cases listed: 12

Method 1

When running the regression command, we can use the save subcommand to

save the predicted values to the current data file.
We have supplied the name for the new variable in parentheses after the SPSS
keyword pred. After
running the regression, we will list the first 12 cases in the data set for the
variables write and pred_1.

```
regression
/dependent write
/method = enter read math
/save pred(pred_1).

<output omitted>

list write pred_1
/cases from 1 to 12.

write pred_1

. 42.24554
. 40.81015
. 54.03857
. 45.58411
. 47.28941
```

. 48.53128

. 56.83733

. 48.67533

. 51.30748

54.00 49.77315

46.00 44.31532

44.00 45.19271

Number of cases read: 12 Number of cases listed: 12

Method 2

Another way to get out-of-sample predictions is to save the model information to an .xml file, use the model handle command to name the .xml file, and then use the ApplyModel function of the compute command to create the predicted values.

We will list the first 12 cases in the data file for the variables write

and yhat.

```
regression
/dependent write
/method = enter read math
/outfile=model('d:/data/working/hsb_m1.xml').
```

<output omitted>model handle name = m1 file='d:/data/working/hsb_m1.xml'.

compute yhat = ApplyModel(m1,'predict').

list write yhat
/cases from 1 to 12.
write yhat

. 42.25
. 40.81
. 54.04
. 45.58
. 47.29
. 48.53
. 56.84
. 48.68
. 51.31
54.00 49.77
46.00 44.32
44.00 45.19

Number of cases read: 12 Number of cases listed: 12

Now let's look at pred_1 and yhat side by side; as you

can see, they are the same.

formats pred_1 yhat (f8.5).
list write pred_1 yhat
/cases from 1 to 12.

write pred_1 yhat

. 42.24554 42.24554
. 40.81015 40.81015
. 54.03857 54.03857
. 45.58411 45.58411
. 47.28941 47.28941
. 48.53128 48.53128
. 56.83733 56.83733
. 48.67533 48.67533
. 51.30748 51.30748
54.00 49.77315 49.77315
46.00 44.31532 44.31532
44.00 45.19271 45.19271

Number of cases read: 12 Number of cases listed: 12