Subject: How to arrive at final wealth index score using PCA
Posted by annabellebronson on Mon, 19 Feb 2024 10:10:10 GMT
View Forum Message <> Reply to Message

smash karts
I need to arrive at a final wealth index score for my analysis using PCA in stata. I am working with the Demographic and Health Survey dataset which provides only an asset-based wealth index for measuring socio-economic welfare. I am unsure about the variable normalisation process to arrive at the final wealth index score for each household. I have followed the variable preparation process using the Filmer and Pritchett /Demographic and Health Survey and coded each variable as a dummy. I have always assumed that if I predict the scores using the command predict score1 for example and apply the survey weights to generate break the scores up into quintiles, I arrive at quintiles representing different categories of wealth from the poorest to wealthiest. I, however, have been unable to duplicate the survey provided wealth index and quintile distribution using this approach. They provide the code in SPSS, and I am working in stata. I need to be sure what I am doing is right and not necessarily arrive at the very same index.

I did some more reading used the following
document;http://www.psi.org/equity-wealth-quintileguide from the Population Service International (PSI). They show how to construct a wealth index based on a survey. In the document which is a guide they use the PCA command and then generate a factor score using the predict command, it appears they do not use this as the final wealth index score. They, however, go on to standardise each variable using the mean and standard deviation. They then proceed to create a final wealth index score for each household by multiplying the standardised variable by the factor scores obtained from the PCA and summing these up for each household. I guess my questions are;
1. Is predicting the relevant factor score using the predict command in stata alone adequate to arrive at a wealth Index score based on which to rank households.
2. Is ranking the households using quintiles and survey weight command sufficient to reflect the distribution of wealth based on this index.
3. is the process of generating the index using the standardised variable as described above same or similar to the stata predict command, ie. another way toobtain the score?

In the interest of brevity my commands are as follows;
pca n1, n2, n3 ... ni [aweight = weight], means
predict wealthscore1 wealthscore2
sort wealthscore1
xtile quintile= wealthscore1 [pweight=weight], nq(5)
Some clarity as soon as possible will be greatly appreciated. regards,

Subject: Re: How to arrive at final wealth index score using PCA
Posted by Janet-DHS on Wed, 28 Feb 2024 19:01:53 GMT
View Forum Message <> Reply to Message

Following is a response from DHS staff member, Tom Pullum:

Yes, you can match the SPSS procedure with Stata. Have you looked at the DHS website

(https://www.dhsprogram.com/topics/wealth-index/index.cfm)? MANY variables go into that PCA. Various people have tried to simplify the construction, and I support that, but if you simplify it you will not get a match.

The conversion to quintiles is a little more complex. The following lines, applied the HR file, show how to do it ("mem" is the number of household members). You have to do it with the HR file because everyone in the same household has the same value of the continuous index and is in the same quintile.

```
set more off

use "...NGHR7AFL.DTA" , clear

keep hv001 hv002 hv005 hv012 hv013 hv270 hv271

gen mem = hv012

replace mem = hv013 if mem == 0

gen pwt=mem*hv005

gen wt=pwt/1000000

xtile hv270_test1=hv271 [pweight=pwt], nquantiles(5)

tab hv270 hv270_test1 [iweight=wt]
```

Subject: Re: How to arrive at final wealth index score using PCA
Posted by Anonymous on Thu, 31 Oct 2024 01:27:16 GMT
View Forum Message <> Reply to Message

Using the predict command after running PCA will give you the factor scores based on the component you specified. This score can serve as a wealth index, but it is important to ensure that the PCA model is appropriately specified (e.g., the number of components retained). However, simply using the factor scores without additional adjustments (like standardization or weighting) might not fully align with the methodology used in the DHS to create their wealth index.

Subject: Re: How to arrive at final wealth index score using PCA
Posted by tahirandreas5 on Thu, 31 Oct 2024 21:22:34 GMT
View Forum Message <> Reply to Message

If you want to use the predict command after Principal Component Analysis (PCA), you will get

factor estimates based on the specified component. These estimates can be used as an index of wealth.

However, it is important that the PCA model is set up correctly. For example, you need to choose the right number of components to keep. If you choose the wrong number of components, the results will be biased and the wealth index will be incorrect.

If factor scores are used without additional modifications such as standardisation or weighting, this will not be consistent with the methodology used by the household surveys (DHS) to produce the wealth index. The DHS often use more sophisticated methods that take into account different aspects of the data. For example, they use weighted components or additional statistical methods to make the index more accurate.

Subject: Re: How to arrive at final wealth index score using PCA
Posted by Anonymous on Tue, 25 Feb 2025 02:11:24 GMT
View Forum Message <> Reply to Message

When you use the predict command after performing PCA, it generates factor scores based on the components you selected. These scores can be used as a wealth index, but it's essential to make sure the PCA model is correctly set up (e.g., choosing the right number of components to retain). However, simply relying on these factor scores without further adjustments, such as standardization or weighting, may not fully match the approach used by the DHS to calculate their wealth index.

Subject: Re: How to arrive at final wealth index score using PCA
Posted by Anonymous on Tue, 18 Mar 2025 04:14:52 GMT
View Forum Message <> Reply to Message

annabellebronson wrote on Mon, 19 February 2024 05:10
I need to arrive at a final wealth index score for my analysis using PCA in stata. I am working with the Demographic and Health Survey dataset which provides only an asset-based wealth index for measuring socio-economic welfare. I am unsure about the variable normalisation process to arrive at the final wealth index score for each household. I have followed the variable preparation process using the Filmer and Pritchett /Demographic and Health Survey and coded each variable as a dummy. I have always assumed that if I predict the scores using the command predict score1 for example and apply the survey weights to generate break the scores up into quintiles, I arrive at quintiles representing different categories of wealth from the poorest to wealthiest. I, however, have been unable to duplicate the survey provided wealth index and quintile distribution using this approach. They provide the code in SPSS, and I am working in stata. I need to be sure what I am doing is right and not necessarily arrive at the very same index.

I did some more reading used the following document;http://www.psi.org/equity-wealth-quintileguideDrift Boss from the Population Service International (PSI). They show how to construct a wealth index based on a survey. In the document which is a guide they use the PCA command and then generate a factor score using

the predict command, it appears they do not use this as the final wealth index score. They, however, go on to standardise each variable using the mean and standard deviation. They then proceed to create a final wealth index score for each household by multiplying the standardised variable by the factor scores obtained from the PCA and summing these up for each household. I guess my questions are;
1. Is predicting the relevant factor score using the predict command in stata alone adequate to arrive at a wealth Index score based on which to rank households.
2. Is ranking the households using quintiles and survey weight command sufficient to reflect the distribution of wealth based on this index.
3. is the process of generating the index using the standardised variable as described above same or similar to the stata predict command, ie. another way toobtain the score?

In the interest of brevity my commands are as follows;
pca n1, n2, n3 ... ni [aweight = weight], means
predict wealthscore1 wealthscore2
sort wealthscore1
xtile quintile= wealthscore1 [pweight=weight], nq(5)
Some clarity as soon as possible will be greatly appreciated. regards,

Using the predict command after performing PCA will return factor scores for the component you specified. This score can be used as a wealth index, however the PCA model must be properly configured (for example, the number of components kept). However, using the factor scores without any extra changes (such as normalization or weighting) may not fully agree with the approach used by the DHS to produce their wealth index.

Subject: Re: How to arrive at final wealth index score using PCA
Posted by Anonymous on Sat, 23 Aug 2025 14:18:21 GMT
View Forum Message <> Reply to Message

hi, i think in DHS methodology, the wealth index is indeed based on the first principal component (PCA). Using predict in Stata after running PCA will give you the factor scores, which can be used to rank households. Two clarifications:

Yes, taking the first predicted factor score (usually wealthscore1) is sufficient to construct the wealth index.

Households are then ranked and divided into quintiles using survey weights (as you did with xtile).

Subject: Re: How to arrive at final wealth index score using PCA
Posted by emmascott637 on Mon, 25 Aug 2025 08:59:52 GMT
View Forum Message <> Reply to Message

I had the same problem in Stata. The predict scores work for ranking households, but they won't exactly match the DHS wealth index because DHS standardises variables first and applies

loadings differently. Using quintiles with survey weights is fine for your own analysis, just don't expect an identical replication of the DHS index.