
Subject: Which recode file to choose IR-KR-BR
Posted by [Alanood](#) on Mon, 27 Feb 2023 16:53:51 GMT
[View Forum Message](#) <> [Reply to Message](#)

Hi

I repst my question here in " Dataset use in Stata" b/c my question is related to dataset and stata , I hope i will get answered.

I am a PhD student in economics, I am writing my first paper in the impact of women employment on child nutrition - for Egypt - 2008 and 2014

My outcome variable is child health indicators (HAZ - WHZ - WAZ), I produced binary variables: (stunted wasted overweight underweight)
The main independent variable is women employment , yes/no
Also I have controled for other variables such as child characteristics , mother characteristics, Household characteristics.

In the begging I started using the KR file since the child is my outcome variable.I cleaned the data and started analyzing it, however, when I tried to replicate the tables and analysis for one paper which is published in Q1 journal.

The authors used Egypt 2014 DHS dataset.

The result for the replicating frequency tables were not the same, so I tried to use the IR file. and then I replicated these tables successfully.

So to confirm, The frequency tables for maternal employment comes from IR file.

For the OLS, I couldn't replicate their work yet but i got significant coefficient as them.

```
*****stata comand*****  
gen strata=v023  
gen psu=v021  
gen sampwt=v005/1000000  
svyset psu, strata(strata) weight(sampwt) vce(linearized)  
gen id = sgov
```

```
*****OLS*****  
svy: reg stunted maternalemployed ib1.gender bage100 bagesq100 i.twin i.birthsize mage100  
magesq100 mothertotedu100 mothertotedusq100 i.marital nofchildren i.wealth nofadult i.id
```

They stated " We include a governorate dummy to control for regional differences and governorate fixed effect" so I include the i.id (the governate fixed effect)

```
*****Result*****
```

They have 12,502 observation , R2= 0.0760 , mother employed coefficient : 0.0316 (10% significant)

for my analysis i got

In the IR - Number of obs = 9,809 R-squared = 0.0825 , mother employed
coefficient : .0336397 (5% significant)

In the BR - Number of obs = 13,415 R-squared = 0.0751 , mother employed
coefficient : .0248322 (10% significant)

In the KR - Number of obs = 11,774 R-squared = 0.0751 , mother employed
coefficient : -.1192583 (1% significant)

I used the same variables as they used and i did the weighting, strata, cluster , and fixed effect for the governate.

as they stated in their paper

" all the regression analyses have corrected for the survey design, i.e. the sampling weight, the cluster, and the strata were all taken into account."

My questions here are:

1- - what do you think which file should i go with ? As you can see the coefficient of the Maternal employment in KR file is negative where it's positive in the IR file. the relationship between the dependent variable and independent variable will be different according to the which recode file i am using.

2- is it possible to do the Description table from one file (IR) and the analysis from another file (KR)

or the paper should use only one kind of recode file. (IR BR KR)

I appreciate your feedback and assistance.

Thank you

Subject: Re: Which recode file to choose IR-KR-BR

Posted by [Bridgette-DHS](#) on Mon, 27 Feb 2023 21:00:24 GMT

[View Forum Message](#) <> [Reply to Message](#)

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

I think you are not clear on the differences between the files. The IR file has women as units. The children of the women are identified with subscripts. The children born in the past five years--these are the only children whose height and weight are measured--are indicated with subscripts 1, 2, 3, 4, 5, where 1 refers to the youngest child. In these two surveys, 5 is the maximum number of children born in the past 5 years. For example, the HAZ scores are given by

hw70_1, hw70_2, hw70_3, hw70_4, and hw70_5. Most of these will be NA, depending on how many children the woman actually had in the past 5 years.

The IR file is not well-suited to the analysis of data for individual children. In fact, I don't know why or how someone would use the IR file to analyze data for children. I don't know how you did that. Instead, we use the BR and/or KR files, which have one child per record. For example, if a woman had two children in the past 5 years, there would be two records in the KR file, one for each child, and virtually all of the woman's information would be repeated for each child.

The BR file has a record for all children in the birth histories, i.e. for all children ever born. The KR file is a subset of the BR file, restricted to those children in the BR file who were born in the past 5 years. Because the HAZ, etc., are only coded for children under 5, they are supposed to be NA for children over 5. You should not get any difference between an analysis using the KR file and an analysis using the BR file. The only reason I can think of for why you could get any difference at all is that the BR file may include the HAZ, etc., for some children whose age was determined to be greater than 5, and they would be in the BR file but not the KR file. But that should not happen.

For your analysis you should use the KR file.

Subject: Re: Which recode file to choose IR-KR-BR
Posted by [Alanood](#) on Tue, 28 Feb 2023 02:45:51 GMT
[View Forum Message](#) <> [Reply to Message](#)

Thank you Tom Pullum for your reply.

I understand that the IR file is not well-suited to the analysis for children, so for the analysis I will use either KR or BR.

how about the discription part??

I noticed that the authors used different type of file depend on what is the main interest. for example, If we are looking at the ditribution of working women across education they used IR file. However, if we are looking at the frequency of stunted childern in regions they used BR file.

Is this okay that we used different recode file in the same paper. or this will data misspresentive.

Table 1. Child malnutrition (stunted- wasted- underweight- overweight) by socio-economic characteristic.>>>>>> they used BR

Table 2. Working mothers by socio-economic characteristics.>>>>>>> they used IR

Table 3. Percentage of mothers by socio-economic background in various occupations >>>> they used IR

note

I noticed that the description tables in this paper matched exactly the tables in egypt2014 report. but how the authors sample 12,888 child where in the report is 13,601 child (Table 12.9 Nutritional status of children page 175)

no exclusion conditions have been done in their paper, so how they got 12,888 ?

Also how they get the same numbers as in the report and they used different number of sample.

I think I need to let it go , this paper cant be replicated. I tried to do so b/c its the most related paper to my topic, and i thought if i replicate it i will know how to do a high quality paper. However now I know that If i publish my paper in the future, I will state this information. (what recode file i used , how many children in this file , and what is the exclusion conditions or specifications)

Your reply is valuable. Thank you again

Subject: Re: Which recode file to choose IR-KR-BR

Posted by [Bridgette-DHS](#) on Wed, 01 Mar 2023 16:03:43 GMT

[View Forum Message](#) <> [Reply to Message](#)

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

I think it would be appropriate to have a section in which the mothers are the units, and another section in which the children are the units. For some topics, such as this one, both perspectives are valid. So, even if the analysis mainly uses the KR / BR files, some description using the IR file, with mothers as the units, and describing their education and employment and fertility, say, could be useful. You just need to be clear for the reader. Good luck with this topic!
