Subject: All-women factor in trend analysis
Posted by Serala on Fri, 03 Mar 2023 14:44:20 GMT
View Forum Message <> Reply to Message

Dear forum users,

I am planning to do a difference-in-differences analysis using Nepal DHS 2001, 2006, 2011, and 2016, the women's IR recode. I have appended the data to include all these survey rounds and I have identified a control group and treatment group. I am now trying to do a graphical analysis to see whether the treatment and control group have similar trends in women's educational and employment outcomes.

However, I just noticed that the 2001 NDHS is an ever-married sample while the other rounds are all-women samples. So my question is, how can I make the samples comparable, for example when looking at the trend in the total years of education for women? I assume I have to use the AWFACTE for education, but how? And what about employment outcomes? All help would be very much appreciated!

Best regards,
Serala

Subject: Re: All-women factor in trend analysis
Posted by Serala on Fri, 03 Mar 2023 14:50:18 GMT
View Forum Message <> Reply to Message

Below is a part of my stata code if that helps to understand my problem. AccessAboveMed=1 is my treatment group.

```
*Appending survey rounds*
use [2001 data]
gen sample=2001
append using [2006 data]
replace sample=2006 if missing(sample)
append using [2011 data]
replace sample=2011 if missing(sample)
append using [2016 data]
replace sample=2016 if missing(sample)
egen strata_ID = group(sample v022)
egen cluster_ID = group(sample v001)

*Survey setting*
svyset, clear
gen wt = v005/1000000
svyset [pw=wt], psu(cluster_ID) strata(strata_ID) singleunit(centered)


*Parallel trends test*
```

```
graph twoway (scatter v133 sample if AccessAboveMed==1, msymbol(T) mcolor(blue)) (scatter
v133 sample if AccessAboveMed==0, msymbol(0) mcolor(orange)) (lfit v133 sample if
AccessAboveMed==1 & sample<=2011, lcolor(blue)) (lfit v133 sample if AccessAboveMed==1 &
sample>2011, lcolor(blue)) (lfit v133 sample if AccessAboveMed==0 & sample<=2011,
lcolor(orange)) (lfit v133 sample if AccessAboveMed==0 & sample>2011, lcolor(orange)
xline(2011) legend(order(1 "Treated" 2 "Control"))), yscale(range(0 10))
```

Subject: Re: All-women factor in trend analysis
Posted by Bridgette-DHS on Fri, 03 Mar 2023 17:28:53 GMT
View Forum Message <> Reply to Message

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

I recommend that you not use the all-women factors at all, but instead construct a revised IR file
for the 2001 survey (NP41), which can be handled like any other IR file. A program to do this is
attached, and is set up for this specific survey. It contains many comments. Usually the marital
status variable in the PR file is hv115, but in this survey it is sh08, so there is an exception for
that. Let us know if you have any questions.

The all-women factors are not just a nuisance--they are calculated in an arbitrary way, based on
single years of age, and are different for different covariates. The procedure suggested here is
frankly much more defensible than using an ever-married IR file with all-women factors.

File Attachments
1) convert_EMW_IR_file_to_all_women_file_do_24Apr2023.txt,
downloaded 81 times

Subject: Re: All-women factor in trend analysis
Posted by Serala on Mon, 06 Mar 2023 12:48:43 GMT
View Forum Message <> Reply to Message

Dear Bridgette and Tom,

Thank you so much for the code, this was very helpful! I was able to construct the IR file for all
women. One more question: As there are no employment variables (such as v714, "currently
working") in the PR file, should I just run my analysis and accept the fact that there are missing
values in the employment outcomes, or is there something else that I should still do?

Best regards,
Serala

Subject: Re: All-women factor in trend analysis
Posted by Bridgette-DHS on Mon, 06 Mar 2023 13:19:16 GMT
View Forum Message <> Reply to Message

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

Glad it worked. For a variable such as employment status that is in the women's questionnaire but not in the household questionnaire, probably the only option is to use just the data in the original IR file and say that it refers to ever-married women. Unfortunately, it's unlikely that employment status is independent of marital status, even if you control for everything else, so you cannot safely impute a value for never-married women. You may have to drop the 2001 survey from a trend analysis for all women that includes this variable. You could include this survey if you restrict all the other surveys to ever-married women, and/or you could include all women and omit this survey.

This would be a problem even if you were using the original IR file with all-women factors. Those factors can only be constructed for variables that are in both the (original) IR file and the PR file.

Subject: Re: All-women factor in trend analysis
Posted by Serala on Mon, 06 Mar 2023 13:23:55 GMT
View Forum Message <> Reply to Message

Okay, I will have to restrict the other surveys to ever-married women then. Big thank you for the help!

Subject: Re: All-women factor in trend analysis
Posted by Serala on Tue, 14 Mar 2023 08:32:30 GMT
View Forum Message <> Reply to Message

Hello again,

I am now facing the same issue as before but with the BR file. I was wondering if there is a similar way of constructing a revised BR file from an ever-married sample to an all-women version? Or do I have to just restrict the analysis to ever-married women even in those survey rounds where the sample is all-women?

Best regards,
Serala

Subject: Re: All-women factor in trend analysis
Posted by Bridgette-DHS on Tue, 14 Mar 2023 17:26:33 GMT

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

You will be ok just using the BR file as it is. Same for the KR file. The aw factors are not used in the analysis of the children's data.

Where EMW surveys are used, it is assumed that never-married women have no children. Certainly there will be exceptions to that generalization, but probably very few. Women with no births do not contribute children to the BR and KR files. The only file that is affected by the EMW restriction is the IR file.

Subject: Re: All-women factor in trend analysis
Posted by Serala on Wed, 15 Mar 2023 08:14:03 GMT

Hi Tom and Bridgette,

I had a feeling this might be the case but could not find a confirmation. Thank you so much once again!

Subject: Re: All-women factor in trend analysis
Posted by Serala on Fri, 21 Apr 2023 09:27:22 GMT

Dear Mr. Pullum,

I am getting back to you because I only now noticed that the "NPIR41FL_all_women" file that is produced with your code does not include weights (v005) for never-married women.
Never-married women from 2001 are therefore still left out of all of my regressions. I tried to figure out what could be wrong with the code, but I cannot seem to find a solution. The "nmw_file" that is created does include weights, but the final all-women file only has missing values in v005 for never-married women. Would you be able to help me solve the problem?

Best regards,
Serala

Subject: Re: All-women factor in trend analysis
Posted by Bridgette-DHS on Fri, 21 Apr 2023 13:23:24 GMT

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

I just looked, and my version of the constructed all-women file does include valid values of v005 for all women.

If nmw_file.dta includes v005 (it does for me and you say it does for you) then you need to check the following lines in the program:

quietly append using nmw_file.dta

* Check after the append to confirm the addition of many cases with NA
tab bidx_01,m

* re-calibrate v005 so the ratio of mean weights in this combined file is the same as in the PR file
* do not change the IR (emw) weights; instead, multiply the nmw weights (hv005) from the PR file by a factor

summarize v005 if nmw==1
scalar mean_wt_nmw=r(mean)

summarize v005 if nmw==0
scalar mean_wt_emw=r(mean)

* the ratio of mean weights, nmw to emw, in the combined file, is mean_wt_nmw/mean_wt_dmw
* need to multiply the nmw=1 weights by a factor so the ratio matches the ratio in the PR file

replace v005=v005*(mean_wt_emw/mean_wt_nmw)*mean_wt_ratio_nmw_to_emw if nmw==1

These lines work ok for me, but perhaps you are getting a warning of some kind or something else is happening with a default. I suggest that you insert "summarize v005" at several points, especially right after the "append" and see whether the number of cases "Obs" in the output stays constant and at the correct value from the beginning to the end of this section of the program. Hope you can figure it out.

Subject: Re: All-women factor in trend analysis
Posted by Serala on Mon, 24 Apr 2023 07:27:43 GMT
View Forum Message <> Reply to Message

Hi,

Thank you for your response! I do not quite understand these rows:

* Select if de facto, female, and age 15-49
keep if hv103==1 & hv104==2 & hv105>=15 & hv105<=49

* Again check eligibility vs marital status
tab hv117 `lmarital_status'

keep if `lmarital_status'==0

* Again check eligibility vs marital status
tab hv117 hv115

summarize hv005 if hv117==0
scalar mean_wt_nmw=r(mean)

summarize hv005 if hv117==1
scalar mean_wt_emw=r(mean)

scalar mean_wt_ratio_nmw_to_emw=mean_wt_nmw/mean_wt_emw
scalar list mean_wt_ratio_nmw_to_emw

The "scalar mean_wt_emw" does not have any values as we have just gotten rid of all the eligible women (at least in the NPPR41FL data it seems that all eligible women are ever-married women). Therefore the "mean_wt_ratio_nmw_to_emw" only includes missing values. I think this is the reason the v005 only has missing values for never-married women in the final file.

Subject: Re: All-women factor in trend analysis
Posted by Bridgette-DHS on Mon, 24 Apr 2023 16:19:29 GMT
View Forum Message <> Reply to Message

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

I apologize for an error in the version of the Stata program with a "3Mar2023" date. Please use the attached version, with today's date, "24Apr2023". The only change in the corrected version is that the two lines

scalar mean_wt_ratio_nmw_to_emw=mean_wt_emw/mean_wt_nmw
replace v005=v005*mean_wt_ratio_nmw_to_emw if nmw==1

should be used instead of the line "replace v005=v005*...... if nmw=1". That line had included mean_wt_ratio_nmw_to_emw but that scalar had not been defined. The program had correctly extracted the two means from the summarize commands and had assigned names to them, but I had not defined another scalar to be the ratio of those two means. Apparently I was doing some final streamlining of the program. The mistake is essentially a typo but a serious one.

If you make that substitution the program should work ok. I'm only posting a revised version, with today's date, in case someone else wants to use the program. Thanks for checking and seeing that v005 was missing for the never-married women. That would have a serious effect on any

analysis!

Subject: Re: All-women factor in trend analysis
Posted by Serala on Tue, 25 Apr 2023 07:09:56 GMT
View Forum Message <> Reply to Message

Hi,

Thank you so much, now everything looks correct in my final file as well!