

---

Subject: data mismatches

Posted by [jinexoxo](#) on Sat, 01 Nov 2014 10:42:58 GMT

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

---

Hi!

I am using the Individual Recode dataset for Phil 2013. I encoded the code

```
table v044 v024
```

... using STATA. The frequency for woman selected and interviewed for domestic violence module in National Capital Region is 1403.

The result is different from Phil NDHS 2013 for violence against women with 1984 women.

Please help me how the NDHS Report was done.

Thank you ...

---

---

Subject: Re: data mismatches

Posted by [Trevor-DHS](#) on Sat, 08 Nov 2014 02:09:03 GMT

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

---

The NDHS survey is not a self-weighted sample and so your tabulation needs to be weighted.  
Try:

```
tab v024 v044 [iw=v005/1000000]
```

---

---

Subject: Re: data mismatches

Posted by [jinexoxo](#) on Wed, 12 Nov 2014 09:10:02 GMT

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

---

Hi Trevor!

I did the code you told me and it actually produced same results as the final report. Hmmmm, but how would I extract the observations necessary with the use of weights? 'coz I always get the unweighted sample.

Thank you for assisting me. :)

---

---

Subject: Re: data mismatches

Posted by [jinexoxo](#) on Wed, 12 Nov 2014 09:11:50 GMT

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

---

Hi Trevor!

I did the code you told me and it actually produced same results as the final report. Hmmmm, but how would I extract the observations necessary with the use of weights? 'coz I always get the unweighted sample.

Thank you for assisting me. :)

---

---

Subject: Re: data mismatches

Posted by [Trevor-DHS](#) on Wed, 12 Nov 2014 13:54:06 GMT

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

---

I'm not sure I'm following your question.

Whenever you tabulate data, you just need to apply the weights as I did earlier, using:

```
[iw=v005/1000000]
```

or

```
[pw=v005/1000000]
```

or using the svyset command

depending on which is appropriate for your analysis.

---

---

Subject: Re: data mismatches

Posted by [jinexoxo](#) on Thu, 13 Nov 2014 00:10:16 GMT

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

---

Sorry, if I misled you.

I am extracting the observations - I am selecting the observations. Say, respondent numbers 4, 5, 9, 15, ... are to select so I can run an analysis.

Sorry, I'm not good in English.

---

---

Subject: Re: data mismatches

Posted by [jinexoxo](#) on Sun, 16 Nov 2014 14:51:16 GMT

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

---

How would I select weighted cases for a specific region? 'coz the scope of our study covers only one region in the Philippines.

---

---

Subject: Re: data mismatches

Posted by [Trevor-DHS](#) on Mon, 17 Nov 2014 01:00:53 GMT

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

---

You can weight the data in the same way and just select the one region you are interested in.

---

---

Subject: Re: data mismatches

Posted by [jinexoxo](#) on Thu, 20 Nov 2014 00:40:24 GMT

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

---

Thank you.

---

---

Subject: Re: data mismatches

Posted by [jinexoxo](#) on Thu, 20 Nov 2014 04:45:27 GMT

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

---

Hi!

Why is "experienced any physical violence" not in the dataset? And d105g has no answer?

I tried to get the "experienced any physical violence" by identifying if the respondent answered YES from d105a to d105f (and named my variable, physical); however, the total didn't match the final report again.

I used the code:

```
tab physical if v024 == 1 [iw=d005/10000000]
```

I got 53 women who were physically abused, which is 4.66 % of 1144. On the final report, the percentage should be 13.7 % of the respondents got physically abused by their partners.

---

---

Subject: Re: data mismatches

Posted by [Trevor-DHS](#) on Thu, 20 Nov 2014 05:57:37 GMT

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

---

d105g is used when a general question about physical violence is asked, but the indicator for physical violence is not based on just this question, but on the responses to a number of questions. Below is example code to match the physical violence indicator:

```
gen PVEver = (inrange(d105a, 1, 4) | inrange(d105b, 1, 4) | inrange(d105c, 1, 4) | ///  
  inrange(d105d, 1, 4) | inrange(d105e, 1, 4) | inrange(d105f, 1, 4) | ///  
  inrange(d105g, 1, 4) | inrange(d105j, 1, 4) | inrange(d130a, 1, 4) | d115y == 0 | d118y == 0)
```

---

---

Subject: Re: data mismatches

Posted by [jinexoxo](#) on Thu, 20 Nov 2014 12:19:03 GMT

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

---

Will the code you've given me generate "ever experience physical violence" and not "experienced physical violence for the last 12 months"?

---

---

Subject: Re: data mismatches  
Posted by [jinxoxo](#) on Thu, 20 Nov 2014 12:41:18 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

I tried your code and it says invalid name.

---

---

Subject: Re: data mismatches  
Posted by [Trevor-DHS](#) on Thu, 20 Nov 2014 15:02:00 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

The code I gave is for ever experienced physical violence and is used to match the data in the first column of table 15.1. I couldn't find where your numbers were coming from, but I now see that you are referring to the numbers in table 15.10 (please confirm that this is the table you are looking at), and that you are interested in spousal physical violence and not just any physical violence. Sorry, this was not clear at first.

First, I checked the code that I provided earlier and it works just fine for "ever experienced physical violence" matching table 15.1. Here is the Stata output:

```
. use "PHIR61FL.DTA", clear  
  
.   
. gen PVever = (inrange(d105a, 1, 4) | inrange(d105b, 1, 4) | inrange(d105c, 1, 4) | ///  
> inrange(d105d, 1, 4) | inrange(d105e, 1, 4) | inrange(d105f, 1, 4) | ///  
> inrange(d105g, 1, 4) | inrange(d105j, 1, 4) | inrange(d130a, 1, 4) | d115y == 0 | d118y == 0)  
  
.   
. tab PVever [iw=d005/1000000]
```

PVever	Freq.	Percent	Cum.
0	8,814.2297	80.40	80.40
1	2,148.7701	19.60	100.00
-----+			
Total	10,963	100.00	

I don't get any error message, and I can't tell from your message where the error is. Did you copy the code or re-type it? If you re-typed it, perhaps one of the variables is mis-spelled or is incorrectly upper or lower case. I would check that first.

Second, to match table 15.10 you use almost the same code, but without some of the last few conditions, however, you should limit the code to ever married women as follows:

```
. gen PVEver_spouse = (inrange(d105a, 1, 4) | inrange(d105b, 1, 4) | inrange(d105c, 1, 4) | ///
> inrange(d105d, 1, 4) | inrange(d105e, 1, 4) | inrange(d105f, 1, 4) | ///
> inrange(d105g, 1, 4) | inrange(d105j, 1, 4)) if v502 > 0
(5512 missing values generated)
```

```
. tab v024 PVEver_spouse if v502>0 [iw=d005/1000000],row
```

```
+-----+
| Key      |
|-----|
| frequency |
| row percentage |
+-----+
```

Region	PVEver_spouse		Total
	0	1	
National Capital Regi	987.791213	156.481776	1,144.273
	86.32	13.68	100.00
Cordillera Admin Regi	99.587833	15.859699	115.447532
	86.26	13.74	100.00
I - Ilocos Region	291.607249	45.173311	336.78056
	86.59	13.41	100.00
II - Cagayan Valley	212.098394	53.819234	265.917628
	79.76	20.24	100.00
III - Central Luzon	698.653727	83.820854	782.474581
	89.29	10.71	100.00
IVA - CALABARZON	918.00612	99.513806	1,017.52
	90.22	9.78	100.00
IVB - MIMAROPA	141.750067	30.69048	172.440547
	82.20	17.80	100.00
V - Bicol	305.319453	54.026097	359.34555
	84.97	15.03	100.00
VI - Western Visayas	396.651668	54.116535	450.768203
	87.99	12.01	100.00
VII - Central Visayas	425.376509	43.214089	468.590598
	90.78	9.22	100.00

VIII - Eastern Visaya	232.304843	41.147731	273.452574
	84.95	15.05	100.00
IX - Zamboanga Penins	265.596913	38.752526	304.349439
	87.27	12.73	100.00
X - Northern Mindanao	273.245834	47.268561	320.514395
	85.25	14.75	100.00
XI - Davao	356.346056	57.555294	413.90135
	86.09	13.91	100.00
XII - SOCCSKSARGEN	287.88012	51.42066	339.30078
	84.85	15.15	100.00
XIII - Caraga	168.792793	33.588978	202.381771
	83.40	16.60	100.00
ARMM	206.651887	8.217312	214.869199
	96.18	3.82	100.00
Total	6,267.661	914.666943	7,182.328
	87.27	12.73	100.00

---

Subject: Re: data mismatches

Posted by [jinexoxo](#) on Fri, 21 Nov 2014 13:40:48 GMT

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

Yes, I just overlooked at the tables and mistaken ever as got abused for the last 12 months.

Now, I understand how the variable was generated. I only retyped the code you gave me and got an error message. Maybe because I'm reading your replies only in my mobile phone and I accidentally omitted some characters in the code.

Another thing, to generate spousal age, educational difference, etc., do I always have to use STATA commands just like what you gave me (on ever experience spousal physical violence)? Are there any easy and possible ways I can do that in Excel? 'Coz I'm just new to STATA and this is actually my first time to handle secondary data.

Thank you for always accomodating me!

---

Subject: Re: data mismatches  
Posted by [Trevor-DHS](#) on Fri, 21 Nov 2014 16:45:23 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

I strongly recommend that you use something like Stata, or alternatively SPSS or another statistical software. While it is possible to produce some of the variables that you are interested in in Excel it is awkward and more complicated, and Excel isn't really designed for this kind of thing.

---

Subject: Re: data mismatches  
Posted by [jinexoxo](#) on Sat, 22 Nov 2014 04:03:21 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

Oh, I see.

Do you have any modules or lecture notes that I could use to guide me such as generating all the variables I need for my analysis?

---

Subject: Re: data mismatches  
Posted by [Trevor-DHS](#) on Tue, 25 Nov 2014 23:37:46 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

There are several useful tutorials for Stata available online. One is at [http://www.cpc.unc.edu/research/tools/data\\_analysis/statatutorial](http://www.cpc.unc.edu/research/tools/data_analysis/statatutorial)

---

Subject: Re: data mismatches  
Posted by [jinexoxo](#) on Thu, 27 Nov 2014 13:39:00 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

Thank you! :)

---

Subject: Re: data mismatches  
Posted by [jinexoxo](#) on Fri, 26 Dec 2014 08:57:03 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

Hi. Good day, sir!

I am now in the middle of an analysis. I thank you so much for always accommodating and answering all my questions.

I have another problem with my study.

I generated "wife experience physical violence for the last 12 mos" with this:gen victim =

---

(inlist(d105a, 1, 2, 4) | inlist(d105b, 1, 2, 4) | inlist(d105c, 1, 2, 4) | inlist(d105d, 1, 2, 4) | inlist(d105e, 1, 2, 4) | inlist(d105f, 1, 2, 4) | inlist(d105g, 1, 2, 4) | inlist(d105j, 1, 2, 4)) if v502 >0 | recoded 0 as No and 1 as Yes.

Next, I drop some unnecessary observations: drop if v502 == 0

I tried to run a logistic regression with a single independent variable and I got this:

```
. [i]logistic victim v012 if v024 == 1 [iw=d005/1000000][/i]
```

```
Logistic regression                Number of obs =    942
                                LR chi2(1)   =    15.23
                                Prob > chi2   =    0.0001
Log likelihood = -215.9921          Pseudo R2    =    0.0341
```

```
-----+-----
victim | Odds Ratio  Std. Err.   z   P>|z|   [95% Conf. Interval]
-----+-----
v012 | .9359899   .0163611  -3.78  0.000   .9044659   .9686127
_cons | .435881   .2389604  -1.51  0.130   .1488397   1.276489
-----+-----
```

Question: Why do I always get 942 observations when I actually used weights in the data? I expected that the number of observations would be 1144.

Again, thank you ... and merry Xmas! :)

---

Subject: Re: data mismatches  
Posted by [Trevor-DHS](#) on Fri, 26 Dec 2014 16:34:57 GMT  
[View Forum Message](#) <> [Reply to Message](#)

Two points:

1) You need to limit your analysis to women who were asked the domestic violence module using v044, e.g.

```
drop if v044 != 1
```

Without this you have women coded as 0 on your victim variable who were never asked the questions and these may be part of your analysis.

2) Once you drop these cases if you just tab victim

```
tab victim if v024==1
```

you will see that there are 942 unweighted observations in your analysis.

The logistic regression was giving this value as it ignored cases of missing values on d005.

---

Subject: Re: data mismatches



Posted by [jinexoxo](#) on Sun, 28 Dec 2014 08:52:47 GMT

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

---

So am I just going to use that 942 observations?

---

Subject: Re: data mismatches

Posted by [Trevor-DHS](#) on Mon, 29 Dec 2014 05:03:22 GMT

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

---

If you are restricting your data to the National Capital Region only (v024==1), then, yes, your analysis will be restricted to 942 observations.

---

Subject: Re: data mismatches

Posted by [jinexoxo](#) on Mon, 29 Dec 2014 14:22:56 GMT

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

---

So it would be really okay if I run logistic regression without weighting the data?

---

Subject: Re: data mismatches

Posted by [Trevor-DHS](#) on Wed, 31 Dec 2014 02:16:29 GMT

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

---

That's not generally recommended. While there is some debate about whether to run regressions weighted, most DHS data users weight regression analyses. Logistic regression in Stata reports the unweighted number of observations, but I would recommend weighting the data.

---

Subject: Re: data mismatches

Posted by [jinexoxo](#) on Thu, 01 Jan 2015 09:23:27 GMT

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

---

I use logistic victim v012 and it gives me 8160 observations. (did not extract the cases by region)

I tried running it on SPSS but it provides me same results as STATA. I weighted the cases by d005 and select cases by v502 > 0. I did not filter out the cases outside the National Capital Region. The software still uses a total of 8160 observations. The classification table produces no weighted observation.

So what software should I use to run logistic regression for weighted data?

---

Subject: Re: data mismatches

---

Posted by [Trevor-DHS](#) on Fri, 02 Jan 2015 03:08:50 GMT

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

---

A couple of notes:

- 1) Your results are weighted, but Stata reports the unweighted number of observations. That is not a problem.
- 2) You shouldn't use importance weights (iw) with a logistic regression, but rather population weights (pw).
- 3) In fact, rather than using the weights on the logistic command you should use svy commands that take into account not only the weighting but the stratification of the sample, and will give correct significance values for the stratified sample.

To do this, use:

```
gen wt=d005/1000000
```

```
svyset v021 [pw=wt], strata(v023)
```

```
svy: logistic victim v012
```

This also reports the weighted number of cases (see Population size) as well as the unweighted number of observations.

---

---

Subject: Re: data mismatches

Posted by [jinexoxo](#) on Fri, 02 Jan 2015 08:18:27 GMT

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

---

Now, my partner and I can proceed to the last 2 chapters of our thesis. Thank you, sir! You're really a big help.

---