

---

Subject: using svy command and getting out p-values  
Posted by [chichi](#) on Wed, 15 Mar 2017 17:10:28 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

Hello!

I am using the Namibian DHS 2013. I created one file with women's, men's and HIV data. I want to calculate the percentage of people aged 15-49 years, who have heard of AIDS, by background characteristics. I got the same frequencies as in the Namibian final report 2013. But I am not sure if the p-values may be true. For example by residence the p-value for women is significant and for men not, although there are no great differences. I load up a part of my results. It would be great if someone could help me. Thank you!

I used the svy command. Here is also my stata code:

```
svy: tab v013 v751 if gender== "men", missing row  
svy: tab v013 v751 if gender== "women", missing row
```

---

### File Attachments

1) [Heard of AIDS.png](#), downloaded 573 times

---

---

Subject: Re: using svy command and getting out p-values  
Posted by [Mlue](#) on Thu, 16 Mar 2017 08:56:20 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

I think this is supposed to give you the P-values... But, try this...

```
svy: tab v013 v751 if gender== "men", percent format(%4.1f) missing row // Percentages
```

```
svy: tab v013 v751 if gender== "men", count format(%4.0f) missing // Counts
```

```
** ===== **
```

```
svy: tab v013 v751 if gender== "women", percent format(%4.1f) missing row
```

```
svy: tab v013 v751 if gender== "women", count format(%4.0f) missing
```

Now, check your output... the P-value will be the P under your output table...

Example:

Pearson:

Uncorrected  $\chi^2(6) = 14.9190$

Design-based  $F(5.65, 1797.69) = 1.8115 \quad P = 0.0979$

So, here the P-value is  $P=0.0979$ ... Please let me know if it works

---

Subject: Re: using svy command and getting out p-values

Posted by [Bridgette-DHS](#) on Thu, 16 Mar 2017 12:09:47 GMT

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

Following is a response from Senior DHS Stata Specialist, Tom Pullum:

What is the null hypothesis that these p-values refer to? Is it the null hypothesis that, for example, the proportion of women who have heard of AIDS is the same in all age groups? That appears to me to be what you have done, but the tab command as you have shown it would not produce a p-value. I believe Stata will not do a chi-square test with svy. I would like to know what command produced these p values and what null hypothesis you want to test. Note also that these percentages are very close to 100%. Knowledge is nearly universal.

---

Subject: Re: using svy command and getting out p-values

Posted by [chichi](#) on Thu, 16 Mar 2017 13:26:49 GMT

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

Yes, it works. I got the same results as in my attached file. But if you look at "residence" on my attached output table... Does the result mean that there is a significant relationship between "residence" and "ever heard of aids" for women, but not for men? I was confused if the p-values may be true, because there are not much differences in the percentages between women and men...

---

Subject: Re: using svy command and getting out p-values

Posted by [chichi](#) on Thu, 16 Mar 2017 13:48:39 GMT

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

I used the svy command and got out p-values. I want to constitute frequencies by background characteristics and look if there is a relationship between, for example, "age group" and "ever heard of aids". I attached my output for women. There you can see on the top my code and then the results.

## File Attachments

1) [svy command output.png](#), downloaded 645 times

---

---

Subject: Re: using svy command and getting out p-values  
Posted by [Bridgette-DHS](#) on Fri, 17 Mar 2017 12:10:53 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

Following is a response from Senior DHS Stata Specialist, Tom Pullum:

I approached this in a different but equivalent way, using logit regression. I find this:

Women by age group:  $p=0.2626$ , not significant  
Women by residence:  $p=.0002$ , very significant  
Men by age group:  $p=.0000$ , very significant  
Men by residence:  $p=.8207$ , not significant

In terms of statistical significance, the results are different for men and women. However, the substantive implications are really the same for men and women--namely, knowledge is extremely high for all the groups of men and all the groups of women.

Here is how I did this (you must change the paths):

```
use e:\DHS\DHS_data\IR_files\NMIR61FL.dta, clear
svyset v001 [pweight=v005], strata(v022)
```

```
tab v013 v751 [iweight=v005/1000000], row
svy: logit v751 i.v013
svy: logit v751 i.v025
```

```
use e:\DHS\DHS_data\MR_files\NMMR61FL.dta, clear
svyset mv001 [pweight=mv005], strata(mv022)
```

```
tab mv013 mv751 [iweight=mv005/1000000], row
svy: logit mv751 i.mv013
svy: logit mv751 i.mv025
```

---

Subject: Re: using svy command and getting out p-values  
Posted by [chichi](#) on Wed, 22 Mar 2017 17:27:10 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

Thank you! But now I am confused which command to use. As you have seen in my output, the svy command made a Chi-square test and produced as well p-values, but they were different from those you presented with the logic command. What would you recommend to do?

Should the p-values of svy:logit and svy:tab be generally the same?

---

---

Subject: Re: using svy command and getting out p-values

Posted by [Bridgette-DHS](#) on Thu, 23 Mar 2017 12:59:59 GMT

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

---

Following is a response from Senior DHS Stata Specialist, Tom Pullum:

I don't understand the comparison that you are making. Let's be as specific as possible. In order to test the null hypothesis that knowledge of AIDS is the same across all five-year age groups of men, I use this command:

```
use e:\DHS\DHS_data\MR_files\NMMR61FL.dta, clear
svyset mv001 [pweight=mv005], strata(mv022)
svy: logit mv751 i.mv013
```

The output includes this test of the null hypothesis:  $F(9, 516) = 4.18$ ;  $\text{Prob} > F = 0.0000$ .

The p-value is less than .0001; knowledge of AIDS differs very significantly across age groups. This is an F test, not a chi-square test. I do not believe that there is a valid chi-square test of this null hypothesis that includes the svy adjustment. Please show me the lines you used to get a chi-square and p-value using svyset and svy: tab. Please do it just for this specific example, involving mv751 and mv013. Thanks.

---

---

Subject: Re: using svy command and getting out p-values

Posted by [chichi](#) on Thu, 23 Mar 2017 13:19:35 GMT

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

---

Thank you for your efforts!

I want to analyze if there is a significant association between five-year age group and knowledge of AIDS among men. I merged women's and men's data, so my variables are v751 and v013. My analysis just includes ages 15-49 years.

I made it in the following way:

```
svyset [pw=wgt1], psu (v001) strata (v023)
svy: tab v013 v751 if gender=="men", missing row
```

The output includes:

Pearson:

```
Uncorrected chi2(12)    = 17.3136
Design-based F(9.81, 5109.55)= 1.2050    P = 0.2830
```

---

---

Subject: Re: using svy command and getting out p-values  
Posted by [Bridgette-DHS](#) on Thu, 23 Mar 2017 14:01:36 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

Following is a response from Senior DHS Stata Specialist, Tom Pullum:

You cannot "merge" the IR and MR files. You "append" those files. Your line "svyset [pw=wgt1], psu (v001) strata (v023)" will not run. The correct syntax would be "svyset v001 [pw=wgt1], strata (v023)". Your commands should not include the option "missing". What you are calling "missing" cases are actually "not applicable" cases and they are not relevant.

The "uncorrected chi2" value ignores the svyset adjustment. You could get that from a simple "tab a b, chi2" command. The "Design based" model cannot produce a chi-square statistic (neither the Pearson nor the maximum likelihood versions of chi-square), as I said in my previous response. It produces an F statistic from a logit regression (either binary logit or a multinomial logit, depending on the number of categories), in which the svyset adjustment is possible. The p-value from F has the same interpretation that a p-value from a maximum likelihood chi-square would have, if such a chi-square could be calculated.

---