
Subject: Weighting for binary logistic regression analysis in SPSS

Posted by [Zoe_C](#) on Wed, 28 Apr 2021 12:31:06 GMT

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

Dear all,

My goal is to examine a possible relationship between frequency of mass media exposure (radio, television, print media) and intimate domestic violence attitudes towards married wives in India. Therefore, the CR was the most suitable database, since the unit of analysis is a couple that is currently married and living together. Before carrying out a binary logistic regression analysis, I've successfully merged some indicators (HV024, HV025, HV270, SH34, SH36) from the HR into the CR datafile, so I have some additional variables that provide information about the household the couple is situated in.

But, there seems to be some conflicting information on how to use weights in SPSS. This video link [https://www.youtube.com/watch?v=NNg8HD_IKow] instructs to divide the MV005 (the weight I'm using in the CR file) by 1 000 000 (in my dataset called WGT_men). However, some say this will generate wrong results. I've tried my analysis with both these options and the generated p-values are very different from one another: binary logistic regression with MV005 produces a lot more significant p-values (0,000) compared to an analysis with WGT_men.

Can someone provide some clarification on this topic?

Thanks in advance.

Kind regards,
Zoë Carette

Subject: Re: Weighting for binary logistic regression analysis in SPSS

Posted by [Bridgette-DHS](#) on Wed, 28 Apr 2021 13:48:46 GMT

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

Following is a response from DHS Research & Data Analysis Director, Tom Pullum:

DHS includes an artificial factor of 1 million in the weight variables just to remove the need for a decimal point. If you are using Stata and the pweight version of weights, you do not need to divide by 1 million. Stata automatically re-normalizes the weights to have a mean of 1.

However, apparently in SPSS you do need to remove the factor. Apparently SPSS believes the sample size is 10,000,000,000 when it is actually 10,000 (for example). The standard errors are then incorrectly reduced by a factor of 1,000 (which is the square root of 1,000,000) and that's why you get the very narrow confidence intervals and lots of significant results. You should definitely remove the artificial factor of 1 million.

You have to be very careful with weight options. Some packages, and I think SPSS is one of them, will round or truncate the weight to an integer without telling you.

When trying different weight options, I recommend doing exactly what you did--that is, do some runs that are identical except for the weight option, and compare the results. In Stata, at DHS we almost always use either [pweight=v005] or [iweight=v005/1000000] (here, v005 could be hv005, etc., depending on the file).

Subject: Re: Weighting for binary logistic regression analysis in SPSS

Posted by [Zoe_C](#) on Wed, 28 Apr 2021 14:20:09 GMT

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

Dear Dr. Pollum,

Thank you very much, again for your very informative answer. I have included in attachment a file with the two binary logistic regressions based on the two different weight options. Up until now, I have worked with WGT_men. But what do you suggest to avoid that SPSS will round?

Pic1 = binary logistic regression results using MV005

Pic2 = binary logistic regression results using WGT_men = MV005/1,000,000

File Attachments

1) [pic1.svg](#), downloaded 140 times

2) [pic2.svg](#), downloaded 118 times

Subject: Re: Weighting for binary logistic regression analysis in SPSS

Posted by [Shireen-DHS](#) on Wed, 28 Apr 2021 17:15:52 GMT

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

Hello,

If your analysis involves domestic violence variables you should be using the domestic violence weight d005. To just make tabulations of your variables you would simply use the following syntax in SPSS.

```
compute wt=d005/1000000.  
weight by wt.
```

However, since you are performing analyses that involve producing SEs you need to use the complex sample package in SPSS. This is so you can account for the survey design and so you can supply the strata variable (v023) and the psu (v021). This is package would have to be purchased and is not available with the basic SPSS software.

Please also check our code share library on GitHub (<https://github.com/DHSProgram/DHS-Indicators-SPSS>) to check if you are coding your main variables correctly. You may especially be interested in Chapter 17 on domestic violence

indicators.

Thank you.

Best,
Shireen Assaf
The DHS Program
