

---

Subject: Regression analysis based on 'Couples' Recode'

Posted by [Zoe\\_C](#) on Thu, 18 Mar 2021 08:48:48 GMT

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

---

Dear all,

I'm a bachelor student working on a research project where I investigate the possible linkage between frequent exposure to mass media (stratified for television, radio and print media) and individual attitudes on intimate partner violence against women (IPVAW) in India. Using a binary logistic regression analysis, I'll control for a handful of confounding variables such as age, employment, education,...

Based on my readings of the DHS documentation, I'd be using the 'Couples' Recode' for the analysis. However, I'm still somewhat in doubt.... I know this recode is using the couple as the standard unit of analysis and is based on the merging of men and women from respectively the 'Men's Recode' and the 'Individual Recode'. A horizontal line in the dataset is a case (= couple) where for ex. variable V744A gives information about the wife and MV744A gives information about the husband. Is this interpretation correct?

A second question is whether there exists a variable where men are asked the question whether their father ever beat their mother? This question is asked in the domestic violence module for women, however, I can't find a similar question for men in the DHS-VII Recode Manual, however I know that this question must exist.

Kind regards,  
Zoë C.

---

---

Subject: Re: Regression analysis based on 'Couples' Recode'

Posted by [Bridgette-DHS](#) on Thu, 18 Mar 2021 12:45:49 GMT

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

---

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

The answer to your first question is definitely "yes". You have correctly described the structure of the couples file. However, the men's and women's variables are not always in pairs. There are variable for women that do not have matching variables for men (and vice versa, although less often). To see whether the question about fathers was asked of men, you should go through the men's questionnaire, which is in an appendix at the back of the main report on the survey (<https://www.dhsprogram.com/pubs/pdf/FR339/FR339.pdf>).

---

---

Subject: Re: Regression analysis based on 'Couples' Recode'

Posted by [Zoe\\_C](#) on Fri, 19 Mar 2021 18:52:09 GMT

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

---

Thank you so much Tom!

Indeed, men were asked whether their father ever beat their mother in question 716 of the men's questionnaire. However, I can't find this variable in the recode manual and the couple's recode. I've searched intensely but I'm overlooking something I'm afraid.

All the best,  
Zoë

---

---

Subject: Re: Regression analysis based on 'Couples' Recode'

Posted by [Bridgette-DHS](#) on Fri, 19 Mar 2021 23:13:47 GMT

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

---

The variable you are looking for (question 716) is a country specific question. For country specific questions, the variables are generally named and coded in the same way as they were on the questionnaire. The variable will have a leading "SH" if the question was asked at the household level, a leading "S" if asked at the women's level, and a leading "SM" if asked at the men's level.

Please look in the men's questionnaire for Question 716. In the Men's Recode and Couples Recode datasets, look for SM716. The questionnaires can be found here.

---

#### File Attachments

1) [Q-716.png](#), downloaded 1097 times

---

---

Subject: Re: Regression analysis based on 'Couples' Recode'

Posted by [Zoe\\_C](#) on Mon, 05 Apr 2021 11:43:33 GMT

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

---

Thank you so much for all the help!

I had one last question though, again regarding the various indicators. Frequency of internet use during the last 12 months is indeed asked both at the women's and men's questionnaire in DHS VII (question 121 for women's questionnaire). However, I can't find this variable (should be v171b according to the DHS Guide for the women's file) in the couple's recode.

Thanks in advance!

Kind regards,  
Zoë

---

---

Subject: Re: Regression analysis based on 'Couples' Recode'  
Posted by [Bridgette-DHS](#) on Wed, 07 Apr 2021 18:20:56 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

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

To get variables in the IR and MR files that are not already included in the CR file, you can easily merge the CR file with the IR file and then merge the CR file with the MR file. For the CR/IR merge, you sort and match on v001 v002 v003 v024 and keep if \_merge==3. For the CR/MR merge, you sort and match on mv001 mv002 mv003 mv024 and keep if \_merge==3. In the IR file, only keep the variables needed for the merge and the new variables you want, such as v171b. Similarly, in the MR file, only keep the variables needed for the merge and the new variables you want, such as mv171b. No need to re-copy the variables you already have in the CR file.

---

Subject: Re: Regression analysis based on 'Couples' Recode'  
Posted by [Zoe\\_C](#) on Fri, 09 Apr 2021 11:07:55 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

Dear,

I do not find the variable:

v171b Frequency of using internet last month (women)  
mv171b: Frequency of using internet last month (men)

that I intend to merge with the couple's recode, in any of the datafiles. However, it is asked for both men and women in the survey (DHS-VII) in India (question 121 for women).

Is there anyone who could check this for me? I've searched everywhere.

Secondly, how would you recommend to merge the couple's recode and the household's recode?

Kind regards  
Zoë

---

Subject: Re: Regression analysis based on 'Couples' Recode'  
Posted by [Zoe\\_C](#) on Fri, 09 Apr 2021 14:10:41 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

UPDATE:

I'm so sorry for the confusion. The question on frequency of internet use during the last month is not asked in the NFHS-4 questionnaire, this explains the absence of the indicator in India for 2015-2016.

I have some last questions though:

- How should I merge the Couple's Recode with the Household's Recode? Do I use the same variables as you suggested for the CR/IR and CR/MR merges? Do I consider the CR as the base file? I want to keep a couple as the unit analysis in the merged file.

- I need the number of sons and daughters for each couple in the CR file as well. In other words, I think I need to merge again for CR/KR. Is this correct?

- I'm confused as to when to use which weights. In the merged file (Couples recode with household and children recode) I'd work with two subsamples: husbands and wives. Is it correct to assume that I only need to weight the data for women with V005 and for men with MV005, despite the fact that I use variables on the household level. I do ask because there's also a weight for households and for the state (SHV005).

I'm very grateful for all your help up until now!

Kind regards,  
Zoë

---

Subject: Re: Regression analysis based on 'Couples' Recode'

Posted by [Bridgette-DHS](#) on Sat, 10 Apr 2021 19:11:21 GMT

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

---

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

Adding PR variables for the woman and man in the CR file is similar to what you did before. First sort the PR file on hv001 hv002 hvidx. Then sort the CR file on hv001=v001 hv002=v002 hvidx=v003 and merge with the sorted PR file, keep if \_merge=3. Then sort the CR file on hv001=mv001 hv002=mv002 hvidx=mv003 and merge with the sorted PR file, keep if \_merge=3.

The number of children the woman has had is given by the v20\* variables, and for the man by the mv20\* variables. Those variables are already in the CR file. I don't believe there is a foolproof way to find the number of children the couple had TOGETHER.

We recommend using mv005 as the weight for the couples file. The adjustment for nonresponse is typically more important for men than for women. Otherwise you use whatever version of \*v005 is in the file you are using.

Subject: Re: Regression analysis based on 'Couples' Recode'

Posted by [Zoe\\_C](#) on Mon, 12 Apr 2021 09:32:32 GMT

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

---

Thank you very much!

Your explanation of the merging applies to the merging of the HR and CR datafiles right? Because you are mentioning PR or the Person's Recode (file for household members).. I need to merge data for the household and assign it to a certain couple in the couples file, for ex. for each couple (case) I want an extra variable with the associated wealth index of the household they're part of.

Concerning the children of the couples, thank you very much. I'll play around with various variables and determine whether to incorporate it as an explanatory variable or not.

1) I found myself a bit confused about the variables that provide information for the partner or husband. In the case of women, her highest educational attainment is given by V106, for the associated husband, that would be V701. Similarly, employment status of the women is given by V731, that of her husband by V704. BUT, starting from the men's point of view, educational attainment and employment status are given by MV106 and MV73. What are the associated variables for their wives? I've searched through the DHS-VII Recode Manual and the CR datafile of India (NFHS-4) but only come to the conclusion I would again need V701 and V704, but that doesn't seem right is it?

2) I need to determine the marital duration of the couples in the CR file. Both husbands and wives are asked about the year of (first) marriage (resp. S308Y and SM223Y), however this can apply to a marriage other than the current one. Is there a foolproof way to determine the marital age of the couple?

Thank you very much for all your help up until now, truly!

Kind regards,  
Zoë Carette

---

---

Subject: Re: Regression analysis based on 'Couples' Recode'

Posted by [Bridgette-DHS](#) on Mon, 12 Apr 2021 12:56:09 GMT

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

---

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

If you want to merge household-level variables--the ones in the HR file that don't have subscripts--then it's the HR file that you would merge with the CR file. But if you want to attach characteristics of individuals, it's better to use the PR file than the HR file. Most of what you want is probably already in the CR file. I suggest that you look carefully and just add what's not there already.

I don't think there is complete symmetry between women and men in what's asked about the

partner. I believe the woman is asked more about her partner than the man is asked about his partner.

If you want the best estimate of education, for example, I recommend using what is reported by the person about herself/himself. An estimate from the partner is only useful if (a) you don't have it directly from the person or (b) you specifically want to analyze differences between the two responses, e.g. to assess perceptions or communication.

To repeat, however, not every possible question is actually asked in every survey. If you can't find a variable that you think should be included, go to the questionnaires and see if you can find the relevant question. Also use "lookfor ..." in the data file because sometimes you can find what you are looking for as a survey-specific variable with prefix "s".

---

Subject: Re: Regression analysis based on 'Couples' Recode'

Posted by [Zoe\\_C](#) on Mon, 19 Apr 2021 10:09:24 GMT

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

---

Dear,

Again, thank you so much for all you help. It is such a relief to be able to consult your knowledge on the matter.

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. Furthermore, I reduced the CR file to all couples from which the women participated and successfully completed the questions in the domestic violence module, leaving me with 47514 couples/cases. This reduction was necessary since the variable 'father ever beat mother' (D121 (for women) or SM716 (for men)) is a control variable in my analysis. Another choice is to carry out a binary logistic regression for wives and husbands separately, thus leaving me with two subsamples if you will (one for married wives, and one for married men). I know that the DHS strongly advises to use the 'national men's sample weight' (MV005) for analysis based on the CR file. However, I do still feel some confusion around that topic. Isn't more correct to use the 'national women's sample weight' (V005) for the subsample of wives, and the 'national men's sample weight' (MV005) for the subsample of husbands? Moreover, since I reduced to couples for which the wives participated in the domestic violence module, am I to use the domestic violence weight (D005) as well? More specifically, what weights am I advised to use for the subsample of wives and the subsample of husbands?

Furthermore, I still have missing values for the variable that registers the 'Type of caste or tribe of the household head' (SH36) (3.5%) and quite some 'don't know's' (0.5%). Do I need to narrow down my sample even more to exclude these from the analysis?

Thanks a lot in advance!

Kind regards,  
Zoë Carette

---

---

Subject: Re: Regression analysis based on 'Couples' Recode'  
Posted by [Bridgette-DHS](#) on Mon, 19 Apr 2021 18:51:38 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

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

The weight variables (hv005, etc.) mainly correct for the sample design (the over-sampling of small strata and under-sampling of large strata) but they also incorporate adjustments for nonresponse at different levels--households and individuals. Nonresponse is generally more serious for men than for women, and more serious for the DV module than for other parts of the interview. Here's the general hierarchy in terms of the dependent variable in a regression. For the IR data, use v005--except that if your dependent variables comes from the DV module, use d005. For the MR data, use mv005. For the CR file, use mv005, because you need both the woman and the man, and the man is harder to find, so to speak.

If your subfiles of men and women are pulled from the CR file, then I'd say you should use mv005 for both of them, because the women had to appear in the CR file--and d005 for a DV variable. You could probably invent a synthesis of mv005 and d005 but I would not go down that rabbit hole....

We have done some checking, and the effect of using one weight rather than another, when there would seem to be a choice, is trivial. If you SAY which weight you are using, in your write-up, then your findings can be reproduced. Stan Becker at Johns Hopkins has worked on an alternative weight for the CR file and had an article in the journal Demography (also see <https://grantome.com/grant/NIH/R03-HD068716-01A1>). Mahmoud Elkasabi on the DHS staff is preparing an alternative construction of couple weights and it should eventually appear in the DHS Working Paper series.

---

---

Subject: Re: Regression analysis based on 'Couples' Recode'  
Posted by [Zoe\\_C](#) on Tue, 20 Apr 2021 09:56:28 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

Dear,

Thank you so much for your kind advice. it's truly helping me a lot!

To summarise, I will use mv005 as weight for the subsample of wives, and the subsample of

husbands. Since variable D121 (wives) / SM716 (husbands) is asked in the Domestic Violence Module, and needed as a control variable in the binary logistic analysis, I should use d005 as well right? So I've 2 weights to account for. Moreover, I also incorporate the state variable in the analysis, suggesting I'll also need the 'state men's sample weight' (mv005s). Leaving me with 3 possible weights to account for, how am I to do this in SPSS, that only allows to take 1 weight into account via DATA -> WEIGHT CASES or via the complex sample package.

Kind regards,  
Zoë Carette

---

---

Subject: Re: Regression analysis based on 'Couples' Recode'  
Posted by [Bridgette-DHS](#) on Tue, 20 Apr 2021 13:24:07 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

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

You only use a state weight for an analysis that is confined to a specific state, and even then you will be ok with the national weight. State-specific weights are only included in a few surveys and I never use them. As for choosing between mv005 and d005, I would base the choice on the source of the dependent variable. If it's from the MR file, use mv005. If it's from the DV module in the IR file, use d005.

But I also suggest that you do a few regressions with mv005 and exactly the same regressions with d005 (and maybe with the state weights?) to put your mind at ease. You should get very similar results, except that a case will be dropped if the weight variable is NA.

---