

## Understanding Society User Support - Support #843

### w\_englang is not part of xwavedat, but available for wave a, e, f

08/26/2017 03:13 PM - Nico Ochmann

|  |                 |                    |            |
|--|-----------------|--------------------|------------|
| <b>Status:</b>   | Closed          | <b>Start date:</b> | 08/26/2017 |
| <b>Priority:</b>   | Normal          | <b>% Done:</b>     | 100%       |
| <b>Assignee:</b>   | Nico Ochmann    |                    |            |
| <b>Category:</b>   | Data management |                    |            |
| <b>Description</b><br>Dear Alita,<br><br>if I understand the variable description correctly, englang is unfortunately not part of xwavedat. Whether English is first language is asked both immigrants and natives in wave one, and the IEMB in wave 6.<br>For wave 5, the englang variable appears as well. I want to use all six waves in USoc, for the years 2009-2015. For that purpose, I need to impute the information I have on englang to the missing waves/years.<br>So far, I have used the information from all three waves and generated a new variable englang_dv which I then merged m:1 pidp with the rest of the data set.<br>In broad terms, am I doing this correctly?<br>Thanks a lot!<br>Best wishes.<br><br>Nico |                 |                    |            |

#### History

##### #1 - 08/27/2017 05:05 PM - Alita Nandi

- Status changed from New to In Progress
- Assignee changed from Alita Nandi to Nico Ochmann
- % Done changed from 0 to 90

Hello Nico,

This was asked:

In Wave 1 - Everyone

In Wave 5 - Only those in the EMBoost, GP Comparison or LDA sample, or a recent immigrant, who did not have a valid response for this variable

In Wave 6 - IEMBS members only.

I am assuming you are doing the following, in which case it is correct:

```
use pidp a_englang using a_indresp, clear
merge 1:1 pidp using e_indresp,keepus(e_englang) nogen
merge 1:1 pidp using f_indresp,keepus(f_englang) nogen
```

```
generat emglang_dv=-9
replace emglang_dv=a_englang if a_englang>0 & a_englang<.
replace emglang_dv=a_englang if e_englang>0 & e_englang<. & (englang_dv==.|englangv<0)
replace emglang_dv=f_englang if f_englang>0 & f_englang<. & (englang_dv==.|englang_dv<0)
```

```
keep pidp englang_dv
merge 1:m pidp using indresp_long
```

Best wishes,  
Altia

##### #2 - 08/27/2017 05:05 PM - Alita Nandi

- Private changed from Yes to No

##### #3 - 08/27/2017 06:55 PM - Nico Ochmann

Dear Alita,

thank you very much for your reply.

In order to learn, I would like to share with you how I have done it.

The devil lies in the detail as usual, so I do not think our two codes match 1:1. I still would love to hear your opinion on the following:

```

foreach w in a e f {
use pidp `w'_englang using "$Stata11_se/^w'_indresp_protect",clear
mvdecode _all, mv(-1 -2 -10 -11 -20)
save $Mergeddata_master2016/^w', replace
}

use $Mergeddata_master2016/a, clear
foreach w in e f {
merge 1:1 pidp using $Mergeddata_master2016/^w'
drop _merge
}
generate englang_dv=-9
foreach w in a e f {
replace englang_dv=`w'_englang if `w'_englang>-7 & `w'_englang<. & englang_dv==9
}

merge 1:m pidp using $Mergeddata_master2016/englang_dv, force generate(_wemerge_9)

**later on I try to impute by making every UK born person a native speaker

gen adjustlang = englang_dv
replace adjustlang = 1 if adjustlang==9 & ukb==1 // where ukb ==1 if ukborn

```

As usual, short comments from you are very welcome.

Thanks.

Nico

#### #4 - 09/04/2017 09:23 AM - Stephanie Auty

- Status changed from In Progress to Feedback

Dear Nico,

We have noticed a couple of typos in the syntax above. The second section of the code should have been:

```

generate englang_dv=-9
replace englang_dv=a_englang if a_englang>0 & a_englang<.
replace englang_dv=e_englang if e_englang>0 & e_englang<. & (englang_dv==. | englang_dv<0)
replace englang_dv=f_englang if f_englang>0 & f_englang<. & (englang_dv==. | englang_dv<0)

```

However, our remit at the User Forum is to answer queries related to Understanding Society data and provide general advice about how to manage the data. Given the number of users we have I'm afraid we cannot advise on individual users' analysis syntax specifically. If you can validate the syntax yourself then using loops will mean it is more expandable in the future.

Whether you decide to impute by making every UK born person a native speaker will depend on your research question and whether this imputation is valid for your analysis and research purposes. For example, a person could be born here but brought up with parents speaking another language so they may not have learned much English until they started nursery or school.

Best wishes,  
Stephanie Auty - Understanding Society User Support Officer

#### #5 - 09/04/2017 11:50 AM - Nico Ochmann

Dear Stephanie,

I appreciate your reply and your comments on my imputation method. That is a valid point, I might have to reconsider that.

You may go ahead and close this issue.

Thank you very much.

Best wishes.

Nico

#### #6 - 09/04/2017 11:52 AM - Stephanie Auty

- Status changed from Feedback to Resolved

- % Done changed from 90 to 100

#### #7 - 09/22/2017 03:13 PM - Stephanie Auty

- Status changed from Resolved to Closed

**#8 - 09/01/2023 02:15 PM - Understanding Society User Support Team**

- *Category changed from Data analysis to Data management*