
Subject: Doubts regarding water related variables NFHS-4

Posted by [Varsha](#) on Mon, 27 May 2024 21:41:59 GMT

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

Hello,

I want classify households into two groups based on whether the source of drinking water is on premises or not. The variables hv201 and hv235 seem to be quite contradictory. For instance, when hv201 is public tap/standpipe, it can't be located in one's dwelling/yard/plot. Please help we with this.

Thank you.

Subject: Re: Doubts regarding water related variables NFHS-4

Posted by [Bridgette-DHS](#) on Tue, 28 May 2024 14:28:35 GMT

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

Following is a response from Senior DHS staff member, Tom Pullum:

I opened the HR file (IAHR74FL.dta) in Stata and entered "tab hv201 hv235, m". I think the discrepancies can be explained by a distinction between water for general purposes, with hv235, and water specifically for drinking, with hv201. Otherwise, many combinations, in addition to what you mentioned, would make no sense.

I see that hv235 is NA if hv201 is "piped into dwelling" or "piped to yard/plot". Some recoding of hv235 is called for if you want it to include those cases.

Subject: Re: Doubts regarding water related variables NFHS-4

Posted by [Varsha](#) on Tue, 28 May 2024 18:17:15 GMT

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

Thank you for the clarification, Tom.

Subject: Re: Doubts regarding water related variables NFHS-4

Posted by [Bridgette-DHS](#) on Thu, 30 May 2024 16:06:58 GMT

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

Please see the following response from Senior DHS Staff Member Blake Zachary:

We standardly ask where the drinking water source (hv235) is located (hv201) for public

tap/standpipe and it can be located in the dwelling/yard. The only water sources not asked about location are piped into dwelling/yard/plot and in NFHS-4, rainwater.

A separate question on water source for cooking and handwashing (hv202) is standardly asked for those that use bottled water for drinking but the location of that source is not asked. However, NFHS-4 did not include the question on water for handwashing/cooking.

NFHS-4 QRE

DHS-8 Standard QRE

File Attachments

- 1) [NFHS-4 QRE.png](#), downloaded 611 times
- 2) [DHS-8 Standard QRE.png](#), downloaded 642 times

Subject: Re: Doubts regarding water related variables NFHS-4

Posted by [Anonymous](#) on Fri, 31 May 2024 02:34:34 GMT

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

I understand your dilemma regarding the classification of households based on the source of drinking water being on premises or not, especially when dealing with potentially contradictory data from variables hv201 and hv235. Let's break down the issue and work towards a solution.

Step-by-Step Solution:

1. Understanding Variables:

- hv201: This variable typically indicates the main source of drinking water. Examples might include:

- + 11: Piped water into dwelling
- + 12: Piped water to yard/plot
- + 13: Public tap/standpipe
- + etc.

- hv235: This variable indicates whether the water source is on premises or not. Common codes could be:

- + 1: Yes (on premises)
- + 0: No (not on premises)

2. Identifying Contradictions: As you pointed out, a source like "Public tap/standpipe" (hv201 = 13) is unlikely to be located on the premises (hv235 = 1). We need to identify and address such contradictions.

3. Data Cleaning and Classification: To classify the households accurately, we need to create a logic that cross-references hv201 and hv235 and resolves contradictions.

Pseudo-code / Algorithm:

1. Load Data: Load your dataset into your preferred data analysis tool (e.g., Python, R, Excel).

2. Create a New Classification Variable: Create a new variable (e.g., water_on_premises) to classify households based on the drinking water source being on premises or not.

3. Apply Logic to Classify:

- If hv201 indicates a water source that is inherently on premises (e.g., hv201 = 11 or 12), classify

as on premises.

- If hv201 indicates a public or communal source (e.g., hv201 = 13), classify as not on premises.
- Otherwise, use hv235 to determine the classification.

Example in Python: image

Explanation:

1. Loading Data: The dataset is loaded using pandas.

2. Defining Classification Logic: A function `classify_water_source` is defined to classify the water source based on hv201 and hv235.

3. Applying Classification: The function is applied to each row of the dataset to create a new column `water_on_premises` that holds the classification result.

Addressing Contradictions:

If you find that there are still contradictions (e.g., hv201 = 13 and hv235 = 1), you might need to review those specific records manually or set rules to handle these exceptions based on the context of your study.

Final Step:

After classification, you can analyze the newly created `water_on_premises` variable to group households and derive insights.

Feel free to adjust the logic based on the specific codes and context of your dataset. If you have further questions or need more detailed assistance, let me know!

Best regards,

kneesplatter

File Attachments

1) [f.PNG](#), downloaded 310 times

Subject: Re: Doubts regarding water related variables NFHS-4

Posted by [Varsha](#) on Fri, 31 May 2024 09:52:43 GMT

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

Thank you, Blake.

Subject: Re: Doubts regarding water related variables NFHS-4

Posted by [Varsha](#) on Fri, 31 May 2024 09:53:19 GMT

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

Thank you, Kneesplatter.
