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 314 times