JIMMY:

Just wanted a more clear-cut answer to what we're doing for this assignment.

I get that we are supposed to replace certain numeric values with their actual names (which is easy enough if time-consuming), but the directions regarding the new table confuse me.

It sounds like every type of response is supposed be separated out of its attribute category to have its own column (as in this male respondents would be one column and female another on its own rather than sharing gender).

So are we supposed to have a table with one row and every entry is the number of respondents for that given column or am I misunderstanding the assignment?

The description leaves out some of the attribute sections like 14 and 15, but they do have entries in the actual data. Are they all just the same as attribute 17 (bar attribute 1 which I figure we ignore)?

Also, how many columns are we going to have at the end? Seems like we have a lot so I'd rather not miss one because of bad math.

MAX:

Hey Jimmy,

Your interpretation is close. You will have the same number of rows as the original data set, but all of your variables are categorical so they will be yes/no i.e. 0 or 1

example

ID   Male   Female

1       0           1

2       1           0

3       0           1

4       0           1

And yes, 14 and 15 are left out of the description. If you look at the data, they should have the same values as the other drugs like CL1, CL2, etc. So each drug variable would become 6 new columns for the six possible classes of user for each drug. But since the values are from {C1,...,C6} you can write a function for handling all the drugs-type data rather than a different function for each drug.

 As far as the psychology scores go, I'm going to recommend that instead of handling each score for those variables individually, that we group them into a few categories like N\_score\_very\_low, N\_score\_low, N\_score\_high, and N\_score\_very\_high. If you notice, the values all range from about -3.5 to 3.5 for each of Nscore, Oscore, Ascore, etc. So we could just bag them together like `if -3.5 <= score <= -1.7 then Nscore\_low = 1`. Doing this also allows you to write one function for these variables.

Let me know if you have further questions.

-- Max

JIMMY:

Ah, that makes more sense. So rows are retained and the columns are expanded, making the data entries more like Boolean values.

You mean 7 regarding drugs right? Or is CL0 ignored?

Hope you don't mind me posting these emails to the groupme. I'm sure I'm not the only one who was confused.

MAX:

Sure seven i guess. I’m not looking at the data. Just however many classes there are.

Yeah go ahead and post them. It’s fine. I’ll run this by Azim and see about getting something posted to icollege.