

Hello Shireen,

thank you once more for your help! To be honest, I still have not managed to get to the results in the table of the final report. I have no experience with CSpPro but I tried create a new BMI variable using similar code in R:

```
dataPR$BMI_new3 <- case_when(sh19!=0 & sh231==2 & ha13==0 ~ ha40,  
                             sh19!=0 & sh231==1 & hb13==0 ~ hb40,  
                             shwh==0 ~ shbm,  
                             TRUE~ NA_integer_)
```

Using

```
sdataPR %>% filter(BMI_new<9998 & (sh284a<994 | sh259==1)) %>%  
  group_by(hv104) %>%  
  summarize(total=survey_total())
```

to restrict the analysis to those with both valid glucose or current diabetic medication and valid BMI, I ended up with 3612 observations for men and 3691 observations for women. Based on the final report (table 15.5.1 and 15.5.2 on p244/5), it should be 3812 women and 3721 men. Therefore, I still have too few observations for BMI (and these numbers still include pregnant women).

These are the estimated observations per category:

BMI_cat	hv104	n	n_se
<fct>	<int>	<dbl>	<dbl>
1 underweight	1	1050.	37.0
2 underweight	2	1088.	36.7
3 normal	1	2234.	48.0
4 normal	2	1949.	44.9
5 overweight	1	304.	18.6
6 overweight	2	521.	28.9
7 obese	1	24.8	5.20
8 obese	2	133.	12.8
9 NA	1	108.	14.7
10 NA	2	132.	13.4

Do you have any idea what my mistake might be? I ignored the pidx command in your code as I don't know what exactly it does. I think referring to the household index is not necessary in R?

Secondly, I tried to create an indicator for pregnant women using

```
dataPR$preg <- case_when(dataPR$ha54==1 | dataPR$sh232==1 ~ 1,  
                          TRUE ~ 0)
```

This gave me 1140 pregnant women (unweighted) and 33 of these have valid glucose measurements and BMI. Strangely, all of these are above 50. According to table 15.5.1, there should be 6 pregnant women. Does the table assume a different cutoff than 50 to define a

reasonable age for pregnant women?

Best,
Sarah
