Checking data

How many 24h recalls:

Before deciding: if you only use 1 there are many non-consumers of legumes, as legumes are not consumed daily, and it is likely that participants will not have consumed legumes within the past 24h. We restricted sample size to those with 4 24h recalls. Within those who had 4 recalls, if you only take the first questionnaire, how many non-consumers do you have? Then do the same with the first two 24 h recalls, and then the first three, and then with all four. A lot of non-consumers will pass to consumers when adding more recalls. This will give insight to how many webQs should be included. When the number of consumers and non-consumers is stable, that’s how many 24h recalls I’ll need.

It’s just a cross-tabulation

|  |  |  |
| --- | --- | --- |
| Recalls | Consumer = 1 | Consumer = 0 (non-consumer) |
| 1 |  |  |
| 2 |  |  |
| 3 |  |  |
| 4 |  |  |

Code

Filter instances =+4

New variable:

Consumer if 26086!=0 or 26101!=0 or 26136!=0 or 104280!=0

recalls <- instances %<% new data frame?

Date of right censoring

Crosstab of outcome and dates to see if there is a decline and then censor after decline