

Better answers to Project 1 Essays:

2A: The General Social Survey contains a standard “core” of demographic, behavioral, and attitudinal questions, plus topics of special interest. Many of the core questions have remained unchanged since 1972 to facilitate time-trend studies as well as replication of earlier findings. The GSS takes the pulse of America, and is a unique and valuable resource. It has tracked the opinions of Americans over the last four decades. The 1972-2008 GSS has 5,364 variables. (cited GSS website)

2B: I would have personally asked the marriage question differently. I wouldn’t have started with “Are you” because that wording makes an implication. I would have said: “Please state your marital status.” The response options are appropriate except that I would have included a section that said “Decline to comment/answer.” The second question on the number of children was appropriate. I like that it included [a request] to include all that were born at anytime and to include those from a previous marriage. This is a numerical value so it was obviously appropriate to measure it numerically. I might have left the question open in terms of the number of children and let them tell me exactly how many children over eight they had. For the last question, age, it is hard to ask this the wrong way.

2C: The coding works well for the questions asked. Age is a pretty simple one and marital status is a nominal variable so the numbers do not really matter anyway. Again I would change sex because I do not think it [the concept sex] is dichotomous.

Note for 3C: Nominal variables cannot be skewed because the categories have no order. Dichotomous variables cannot be skewed because there is no continuous distribution. Skew also refers to nonsymmetrical distributions with a tail, NOT something that is “messed up” or “not quite right.”

3C: Age is skewed because there is an outlier [that skews] the graph. Marital status and sex are not skewed because they are not variables that can be measured on a continuum.

3D: Looking at this sub-sample it seems like the majority of respondents are female and never married. It also looks like respondents are generally between twenty and fifty years of age.

Note for 4B: One may calculate a mode for nominal variables, but the median and mean are meaningless for such variables.

4B: It is not appropriate to [do] the previous step with nominal variables because the numbers would be insignificant and hold no meaning. The numbers are merely placeholders for each category, they are not meant to be used mathematically.

Not for 4C and 5B: Answers must include understanding of the conceptual meaning of the numbers for the people in the population. Merely restating what the statistics are is not helpful.

4c: Within this dataset, the mean for each variable (age=39.76 & education =13.36) let us know that on average most people surveyed are mid-age, late thirties/early forties, and went to college for at least a year.

5B: The range for age (63) shows that there is a large difference in the top and bottom ends of the ages. The standard deviation (17.4) shows that the average deviation from the mean of age is fairly large, so there is a lot of variation between respondents' ages. The range for education (12), for a variable whose values could be from 0 to 20, indicates that in this group people were more similarly educated than they could have been. The standard deviation (2.8) shows not much variation around the mean. Therefore, most people have similar education levels. Comparing age and education, the larger statistics for age make sense given that education is restricted to its 0 to 20 scale.