GER1000 QUANTITATIVE REASONING
TUTORIAL 3

Please work on the problems before coming to class. In class, you will engage in group work.

All questions refer to the report "National Survey of American Attitudes on Substance Abuse VIII: Teens and Parents". Besides reading the specific passages closely, you might need to browse through other relevant parts.

Question 1

(A) What is the study about (in not more than 20 words)? For the following questions, you need to refer to Appendix B Survey Methodology on page 29. What is the population of interest? What is the sampling frame? Does the frame cover the population completely?

It is about understanding the attitude of US teenagers and their parents on substance abuse. The researchers were interested in all US teenagers between 12 and 17 years (page 27).

The only practical way to reach the teenagers is through households, which are in turn reached by landline telephone numbers. It is reasonable to view all possible US telephone numbers as the frame. A probability sample of size 94,184 was chosen by a commercial survey sample vendor (page 29).

The sampling frame may not cover the population completely. Some households with teenagers between 12 and 17 may not own a landline phone, though such households may not be too common. If these households are quite different those who own landline phones with regard to substance abuse, then there can be bias. For example, if such households are more to substance abuse, then the survey will underestimate the rate of substance abuse in the population.

(B) Look at the data presented in Table B.1 on page 31. Someone suggests that among all the potential respondents, the response rate is 3.4%. How did he calculate this figure? Is this about right, or too low, or too high?

1,987/58,288 x 100% ~ 3.4%. Since 25,471 households are ineligible, the denominator should be 58,288 – 25,471 = 32,817. A better estimate is 1,987/32,817 x 100% ~ 6.1%.

In this survey, it is challenging to calculate the response rate, because it is hard to determine the sample size. The 20,260 (equal to 78,548 – 58,288) households belonging to the 5 lines above "Subtotal, Potential Respondents", include both eligible and ineligible households. So one can reasonably argue that the sample size is somewhat larger than 32817. In the unlikely case that all the 20,260 are eligible, then the response rate is 1,987/(32,817+20,260) = 3.7%.

(C) In Units 5 and 6 of Chapter 4, you learnt that

(i) With a probability sample, an estimate can be devised which fluctuates around the parameter, provided there is no bias.

(ii) The random error in such an estimate gets smaller as the sample size increases.

On page 38, in response to Question 8, 80% of the teenagers stated they lived with their biological fathers. Suppose we may regard the number of teenagers interviewed as so large that the random error involved is negligibly small. Would you agree that 80% is very near the

parameter, namely, the percentage of all US teenagers between 12 and 17 who lived with their biological fathers?

There are two likely sources of bias: (1) The frame is unlikely to cover the population; (2) The response rate seems quite low. Hence, it is quite likely that 80% is not close to the parameter. Disagree.

It is common for researchers to overlook substantial bias and pay too much attention to margin of error, especially if the response size is large. Refer to the Literary Digest poll in Chapter 4 Unit 4.

Question 2

(A) Please refer to Appendix E on page 53, which contains questions directed at parents. For the following questions determine if the presented data constitute a categorical or numerical variable? You may ignore the "DON'T KNOW/NO RESPONSE" option.

Question 2

Question 9

Question 12

Question 29

They are all categorical. Nominal: 2, 12. Ordinal: 9, 29. Q12 might seem ordinal, but when in doubt, it is fine to treat a categorical variable as nominal.

(B) Look at Figure 3.1 presented on page 13. Explain how stress is being measured and identify the type of stress variable. Can you pin-point the question in Appendix D that is doing the stress measurement? Can you think of a more objective way to measure stress?

Scale from 0 (no stress) to 10 (very high stress). Low: 0-3, Moderate: 4-6, High: 7-10. Ordinal variable: both original scale and the one in figure. Q16 on page 40.

It is subjective as the teenagers may interpret the scale differently. Two teenagers who both report "7" might have different stress levels. More objective ways may include using a professional to assess stress level, or to measure blood cortisol levels, though these will be challenging to implement in phone interviews.

Note: On page 9, the report has "This survey was conducted by telephone with a randomly selected sample of 1,987 teenagers (ages 12 to 17) living in the United States." This is misleading, giving the impression that the sample size is 1,987 and that the response rate is 100%. It is a real challenge to report a complex study such as this one, so it might be understandable that a slip-up occurs.

Optional:

(1) In this section we read: *"Once a household was qualified as the residence of an eligible teenager, age 12 to 17, permission for survey participation of the teen was sought from the teen's parent or guardian."* (i) If the teenager or the guardian refuses to participate in the survey, is this a problem? (ii) If we call the household but no one picks up on all 4 times, is this a problem?

(i) Yes, this is non-response bias and it can potentially cause problem.

(ii) It is very likely that we are missing a sub-set of our population of interest. Maybe those who are not picking up are those with higher chance of substance use rate. Obtaining a representative sample in a survey is truly hard work.

(2) Look at question 5 at page 37 and the way interviewer is asking this question. Do you see any issue of asking the question in this way?

The first few options may be less likely to picked because the respondent is trying to make sense of the question. Another issue is the inflation of grades.