36-303 Sampling, Surveys & Society
Homework 01 Solutions

February 2, 2011

1   Question 1

See Table 1.

(b) The survey was carried out in regions of the Americas and Europe only. There may be problems generalizing the information found to the entire world since the sampled regions represent mostly developed countries. I couldn't find any information about computer assistance, although I think they probably used computer assistance in randomizing school class. It is generally common to get missing data in questionnaire type of survey, I couldn't find information about how was this problem dealt with in the survey so there might be some bias in the findings.

2   Question 2

See Table 2.

(b) I found all the information needed to fill the table. Overall, I am generally happy about the survey. For example, response rate and cooperation rate were adjusted for calculations. However, It is a common problem that telephone samples may not actually be representative of the US population.

3   Question 3

(a)The target population is US adults (18 years or older). The population in the sample frame is a subset of all individuals using the five largest commercial internet service providers in the US.
(b) Obviously, the sample frame over covers people who uses the five largest commercial ISP. Samples from the target population who does not use emails from one of these five ISP are excluded. Survey were sent through E-mails soliciting however, many email ISP

Table 1: The Health Behavior in School-Age Children Study (HBSC)

Sponsor	The World Health Organization
Collector	collected by an international network of research teams
Purpose	to gain unique insight into the health and behaviour of young people. To study young people's health in its broadest sense - physical, social and emotional well-being, not merely the absence of disease
Year started	1982
Target Population	young people attending school, aged 11,13 and 15 years.
Sampling Frame	students of the age described above who attended school in the thirty-five countries of interest (Americas and Europe), listed within the article
Sample Design	children were selected using a clustered sampling design, where the initial sampling unit was either the school class or the school. The latter was sampled when class lists were not available.
Sample Size	$1500 \times 3 \times 35$
Use of Interviewer	There was no direct interviewer, however, respondents had to fill out a questionnaire
Mode of Administration	trained personnel, teachers and school nurses administered the completion of questionnaires in school classrooms
Computer Assistance	not clear
Reporting Unit	each student
Time dimension	Repeated cross-sectional survey
Frequency	every four years
Interviews per Round of Survey	one
Levels of observation	student
Web Link	http://www.hbsc.org/publications/reports.html

Table 2: Traffic Safety Culture Index

Sponsor	AAA Foundation for Traffic Safety
Collector	Abt SRBI Inc
Purpose	To assess a few key indicators of the degree to which traffic safety is valued and is being pursued
Year started	2008
Target Population	Individuals in the United States who are sixteen years of age or older, presumably those who are allowed to drive
Sampling Frame	various landline and cellular telephone numbers
Sample Design	A list-assisted random digit dial sample of landline telephone numbers, including both listed and unlisted numbers, and a random sample of cellular telephone numbers.
Sample Size	2000 individuals
Use of Interviewer	yes
Mode of Administration	Respondents were asked questions by an interviewer who recorded the relevant information over a phone call
Computer Assistance	used to create a random number list for the stratified landline sample
Reporting Unit	One person age sixteen or older in a household is reported for on behalf of the interviewer. In the case of a household having multiple qualifying individuals, the person with the closet upcoming or past birthday was selected for the interview
Time dimension	two-wave panel of persons
Frequency	once every year
Interviews per Round of Survey	one
Levels of observation	individual
Web Link	http://www.aaafoundation.org/resources/index.cfm?button=research

nowadays use automatic spam filter so the these emails may not have reached the targeted individuals. Measurement error describes behavior that is common and systematic across administrations of the question. Given that the survey is about handheld digital devices, people that actually own blackberries and PDAs are likely to respond fast. Also, the category division is misleading because some people may consider laptops as handheld digital devices. (c)the coverage error can be reduced by simply expanding E-mail soliciting to users outside of the five largest ISP. Also, sampling frame can be extend to people that go to regular stores. Nonresponse error arises when the values of statistics computed based only on respondent data differ from those based on the entire sample data. I think sample frame needs to be extended so that it covers the targeted population. To reduce the measurement error, other sample design other than E-mails soliciting needs to be employed.
(d) Cutting the sample in half increase the sampling error because sampling error is the nonobservational gap between the sampling frame and sample. However, this wouldn't affect the coverage error because our sampling frame is still major ISPs.

4   Question 4

so in section 2.3.6, the book details the non-response bias:
Let $\bar{y}_s$ = Mean of the entire specific sample as selected
$\bar{y}_r$ = Mean of the respondents within the sth sample
$\bar{y}_m$ = Mean of the nonrespondents within the sth sample
n_s = Total number of sample members in the sth sample
r_s = Total number of respondents in the sth sample
m_s = Total number of nonrespondents in the sth sample
The nonresponse bias can be expressed as

$$\bar{y}_r - \bar{y}_s = \frac{m_s}{n_s}(\bar{y}_r - \bar{y}_m)$$