University of Waterloo STAT 332 – W. H. Cherry

Tutorial 1 Outline Solution

T1 – 1. (a) The *target* population is presumably all Canadian 13- and 16-year-olds.

(b) The *study* population is 13- and 16-year-olds currently attending a recognized Canadian school, except in the province of Saskatchewan.

A *first*-stage frame could be a list of recognized Canadian schools (except in Saskatchewan) attended by 13- and 16-year-olds; a *second*-stage frame could then be a list of the classes containing 13- and 16-year-olds in the selected schools; if sudents are selected with*in* classes (rather than surveying whole classes), a *third*-stage frame could be a list of the students in each selected class.

(c) The *respondent* population is those 13- and 16-year-olds in the study population who *would* respond (*i.e*., write the test) *if* they were selected to participate in the investigation.

The *sample* is those members of the respondent population who are actually selected to write the test.

[Is the sample *size* for the math test the approximately 65,000 students mentioned in the third paragraph of the article? Briefly justify your answer.]

(d) A *response* is a score which measures problem-solving ability or performance on the test.

A *population attribute* is the *proportion* of the 13- and the 16-year-olds in the respondent population who are able to solve the most sophisticated math problems.

Another *population attribute* is the *average* or *median* score on the test of all 13- and all 16-year-olds in the respondent population.

(e) Four categories of error in this sample survey are:

- *study error* arises from incomplete overlap of the target population and the study population – the average mathematical problem-solving ability of Canadian 13- and 16-year-olds who do *not* attend recognized Canadian schools (except in Saskatchewan) may be different from that of those who *do* attend such schools.
- *non-response error* arises from incomplete overlap of the study population and the respondent population – the average mathematical problem-solving ability of Canadian 13- and 16-year-olds in the study population who would be prepared to respond in the investigation (*i.e*., write the test) may differ from that of those who would *not* be prepared to respond.
- *sample error*: the value of the sample attribute used as an estimate of the respondent population attribute is unlikely to be *exactly* equal to the true value of the latter; the difference is the sample error. *Probability* selecting (more specifically *equi*probable selecting) has the advantage of providing a basis for assessing the likely size of sample error. The method(s) of selecting are not specified in *The Globe and Mail* article, a significant weakness of the report of the sample survey.
- *measurement error*: the score obtained on the test on the specific occasion when a student writes the test is not likely to be exactly equal to the student's true mathematical problem-solving ability. [In this context, note the phrase near the bottom of the middle column of the article: ..... *Ontario ..... was especially concerned that its students ..... would suffer if questions were not vetted for cultural and gender bias*.]

NOTE: Answers *other* than those given above *may* also be consistent with the information (or *lack* of information) given in *The Globe and Mail* article.

For example, it can be argued that the article implies that it is the *official* Canadian educational system that is being assessed, and so students not in recognized Canadian schools are not of concern to the investigation and the *target* population is all Canadian 13- and 16-year-olds *attending a recognized Canadian school*. Such a description of the target population may require answers to later parts of the question somewhat different from those given above.

T1 – 2. (a) Different methods of selecting the 200 customers (*i.e*., different *sampling protocols*) are:

(1) *Convenience selecting*: take the first 200 customers who come into the restaurant after the start of the investigation.

(2) *Judgement selecting*: select 200 customers considered to be 'representative' of the restaurant's customers.

(3) *Quota selecting*: select 200 customers whose distributions of sex, age, region of residence, occupation/income, etc., are chosen to match (as closely as feasible) the distributions of these same explanatory variates in the population from which the restaurant's customers are drawn. [Such information could likely be obtained from the most recent population census.]

(4) *Systematic selecting*: Select every k*th* customer (*e.g*., k = 5 or 50 or 100) who comes into the restaurant – this is equivalent to *equiprobable* selecting if all possible orderings of customers coming into the restaurant with respect to the response(s) of interest in the sample survey are *equally* probable; the value of k should desirably to be chosen to spread out the selecting process over any hourly, daily (even weekly, monthly?), etc., cycles in the volume of business.

(*continued overleaf*)

Tutorial 1 Outline Solution (continued 1)

T1 – 2. (b)(i) The relative *ease of implementation* of the four schemes is likely as follows:

(1) would be the *easiest* to implement, because it requires little prior planning for how to select the 200 customers.

(4) would be the next easiest to implement – the main additional matter over (1) is designing and implementing a method to keep track of which *is* every k*th* customer.

(2) and (3) are likely to be the *most* trouble to implement in that both require appreciable knowledge about which explanatory variates are relevant in the investigation and about their distributions in the population from which the restaurant draws its customers; there is also the operational difficulty of deciding, for (2), whether 'judgement' indicates that a particular customer should be selected and, for (3), of obtaining information on the values of explanatory variates from customers.

(b) (ii) The source of non-response error is a difference in the values of corresponding attribute(s) of interest in the study and respondent populations. The distinction between these two populations arises *prior to* selecting the sample, but differences in their attribute values result from the incentives for response offered at the (later) time of obtaining responses (or non-responses) from the units selected. This seeking of responses *after* the selecting process suggests that the latter likely has little bearing on the effectiveness of the incentives for response, with the caveat that *non*-probability sampling might be more likely to select (some) units systematically more receptive (or less receptive) to the incentives offered for response; such an effect, if it exists, is likely minor.

- Increasing incentives for response will usually decrease the size of the non-respondent population, which makes the respondent population closer in size to the study population and so makes it more likely that their corresponding attrinbutes will be closer in value. A difficulty is that greater incentives for response will usually be more resource-intensive.

(b) (iii) A key advantage of *probability* (more specifically *equi*probable) selecting to obtain the sample of respondent population units is that it provides a basis for assessing the likely size of sample error. (4) likely comes closest to this goal; *how* close it comes depends on how close to equiprobable are all possible orderings of customers coming into the restaurant with respect to the response(s) of interest in the sample survey.

- There is also the matter of how well the difficulties of proper selecting introduced by any cyclical nature of the volume of business are understood and managed by the value chosen for k.

The *other* three methods – convenience, judgement and quota – of selecting the sample usually provide no quaatitative basis for assessing the likely size of sample error – the history of sample surveys shows that these methods cannot be *relied on* to provide *acceptable* limitation on answers imposed by sample error, even if in some cases they happen to do do.

NOTE: The discussion above in (b) can remind us of the unforgiving nature of error management in data-based investigating. The four error categories in sample surveys – study, non-response, sample, measurement – must *all* be managed adequately; *one* category *in*adequately managed, despite even exemplary management of the other three, can impose limitation on Answers so severe as to make the whole undertakikng a waste of resources.