

Unit 8

L1 Introduction to Statistical Surveys

Population: the complete set of objects or persons that are being considered in a survey

Sample: a subset of the population.

Variables (data): What is being surveyed. (annual income, eye color, age, height, shoe size...)

➤ **Qualitative variables:** Ones that don't have a numerical value. Ex: Eye color

➤ **Quantitative variables:** Ones that can be measured numerically. Ex: Age, height, weight.

i) **Discrete variables:** Variables that have separate values (often integer values)

(# of goals in a hockey game, Shoe size 9, 9 ½ , 10...)

ii) **Continuous variable:** Ones that can be any real number within an interval

(time it takes to run 100m)

Census: Statistical survey where all of the population is being surveyed.

– Canada has a census every ten years.... Every person in Canada is counted.

Poll: Survey where a sample is studied to get information about the population being studied.

– A group of students at a university are asked which night of the week they party the most often

Study: A statistical survey where experts in the field being studied are surveyed

– (“4 out of 5 Dentists prefer X brand of toothpaste” is a study.)

Bias: Any error involved in a statistical survey. There are several sources... basically... why is it not a good survey....

Types of Bias:

– **Choosing a sample:** Don't poll a Seniors retirement home about what activities Senior Citizens partake in. (Seniors in a home are probably less active than those who still live at home)

– **Non-Random Sampling:** Time and place. Asking random spectators at the Olympic Games in Vancouver if they smoke. (perhaps people interested in sports are less likely to smoke)

– **Bad Questions:** Vague or misleading such as, “Are you a heavy drinker?” Define Heavy.

– **Errors in processing data:** tabulating results can lead to an error.

– **Errors in Analysis of data:** Making assumptions, (like discounting undecided voters before an election ... maybe they are decided... just don't want to tell you)

Where and how can we gather the data?

- 1. **Direct observation** (record events or behaviours of people)
- 2. **Face to face interview** (in a public place or shopping center)
- 3. **Telephone interview** (usually faster, efficient)
- 4. **Written questionnaire** (to be filled out and returned)
- 5. **Documentary observation** (pull from existing data base)
- 6. **Mechanical or electronic instrumentation** (scanner in supermarkets, electronic counters)

How and who to choose for a sample population?

- 1. **Simple Random Sampling:** Randomly choose the individuals (Choose names out of a hat)
- 2. **Systematic Random Sampling:** Choose a starting individual and then choose every nth one after that. (Inspecting every 100th TV on an assembly line for quality control)
- 3. **Cluster Sampling:** The population is made up of several similar clusters with a lot of variation within each cluster (ex: gr.9 homeroom), we randomly choose some clusters to survey.
- 4. **Stratified Sampling:** The population is divided into subgroups (strata) by personalities, gender, interests, grade level, age etc. We want to represent all proportions of the population in our sample.

Ex 1: A small school with 400 students has 160 grade 9's, 140 grade 10's, and 100 grade 11's .

A sample of 60 students are to be chosen for a survey. How many of each grade should be included?

Ex 2: The following table shows the distribution of the 1200 students in a school.

	# of girls	# of boys
First cycle	360	345
Second cycle	240	255

A sample of 180 students is required, it must be representative of the population. How many girls from the second cycle should be in the sample?

Practice:
8.1 # 1-4

