Unleashing Potentials
Shaping the Future

UNIVERSITI TEKNOLOGI MARA

STATISTICS FOR BUSINESS & ECONOMICS

CHAPTER 1
INTRODUCTION TO STATISTICS

Elementary Statistics
A STEP BY STEP APPROACH
Ninth Edition

ALLAN G. BLUMAN

Prepared by:
Kamarul Ariffin Mansor
Department of Mathematical Sciences
UiTM Kedah Branch Campus
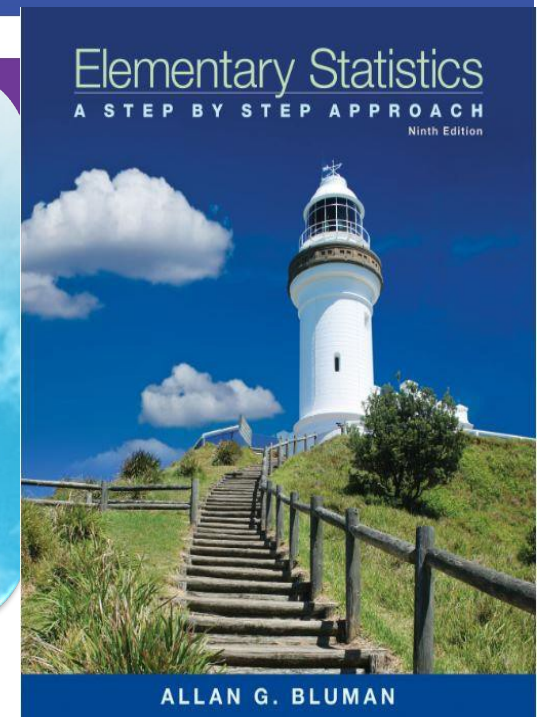
What is Statistics?

Statistics is the science of collecting, organizing, presenting, analyzing and interpreting numerical data, for the purpose of assisting in making a more effective decision.

Types of Statistics

Descriptive Statistics

Inferential Statistics
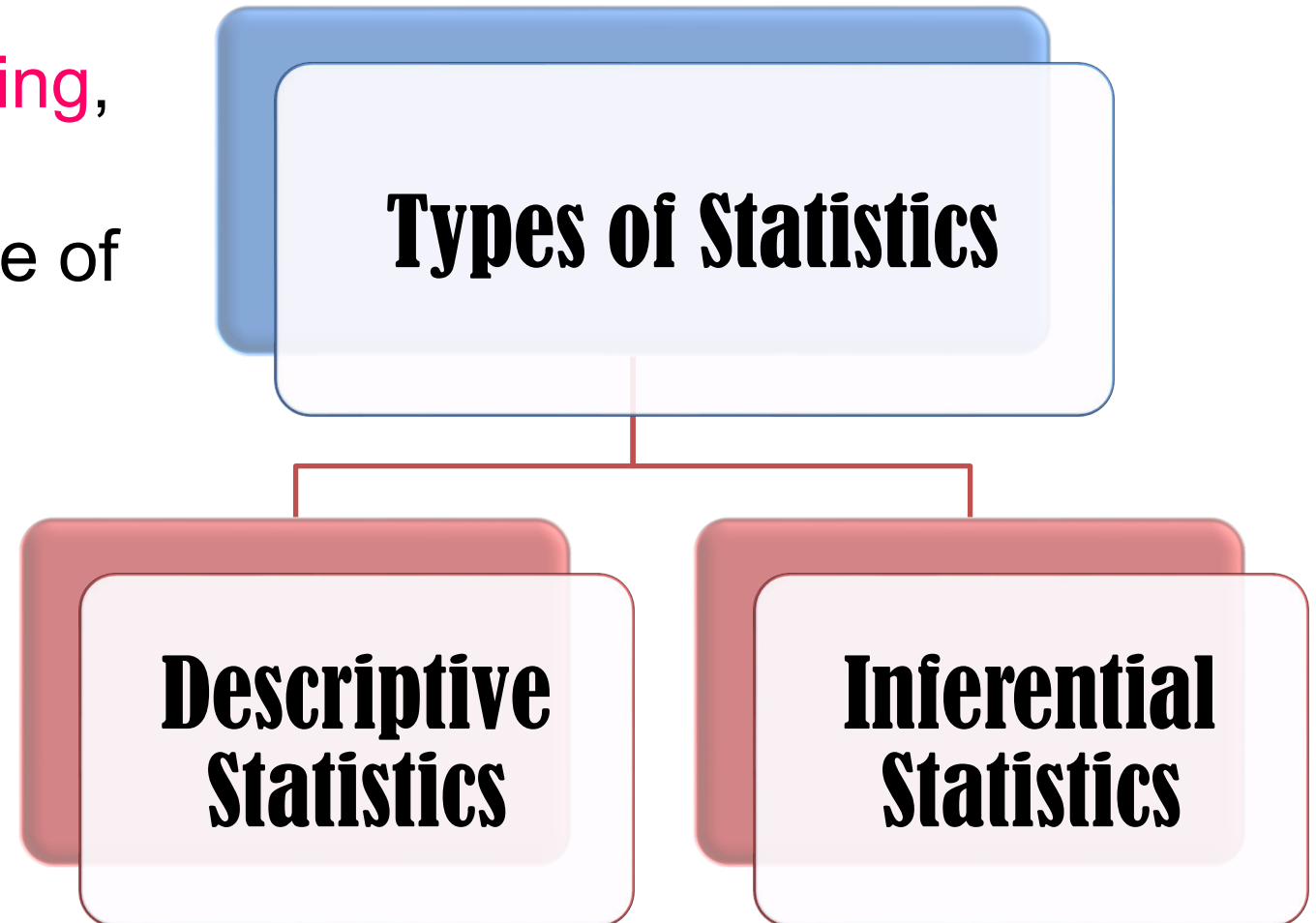
Descriptive Analysis

- The process of collecting, compiling, summarizing and presenting data into graphical forms such us charts, graphs, tables or numerical forms such as average and percentage derived from them so that one can evaluate the data set easily.
- In descriptive statistics, our objective is to describe the properties of a group of scores or data that we have "in hand,"

Inferential Analysis

A decision, estimate, prediction, or generalization about a population based on a sample. It consists of methods that uses ample results to help make decisions or predictions about a population.

Descriptive statistics	Inferential statistics
• Describe the data set • Concerned with describing and summarizing a sample	• Use the data to draw conclusions about the population • Concerned with going beyond the sample to make predictions about the population from which the sample is being drawn

Test of Understanding 1

1. What is statistics?
2. Identify the following as Descriptive or Inferential Statistics:
   a) The number of damaged or cavity teeth by a dentist.
   b) Prediction by a dentist about the teeth that is susceptible to have cavity or damage in future.

Population vs Sample

Population
The complete collection that share certain property, and it is the entire group to be studied.

➡️ **Parameter**
A number that describes a population characteristics

Since we cannot, and sometimes it is impossible to deal with all the elements in a population, a smaller and representative part of that population is considered. Thus that is called a **sample**.

Sample
A sample is a portion, or part or subset of the population of interest.

➡️ **Statistic**
A number that describes a sample characteristics

Other Statistical Terms

- **Census** is a study of the entire population. Data is gathered on every member of the population.

- **Sample survey** is a study on some selected portion of the population.

- **Element (Experimental Units)** is objects (people or things) on which measurements is taken.

- **Pilot study** is a pretest or trial run on a small number of elements (respondent) before conducting the actual survey. The objectives are:
    - To improve the questionnaires
    - To identify the problems that occur during the survey
    - To predict the cost, time and workforce needed.

Why we collect data?

1. Obtain Input to a Research Study
2. Measure Performance
3. Assist in Formulating Decision Alternatives.
4. Satisfy Curiosity
5. Knowledge for the Sake of Knowledge

Types of Data

Primary Data

Secondary Data
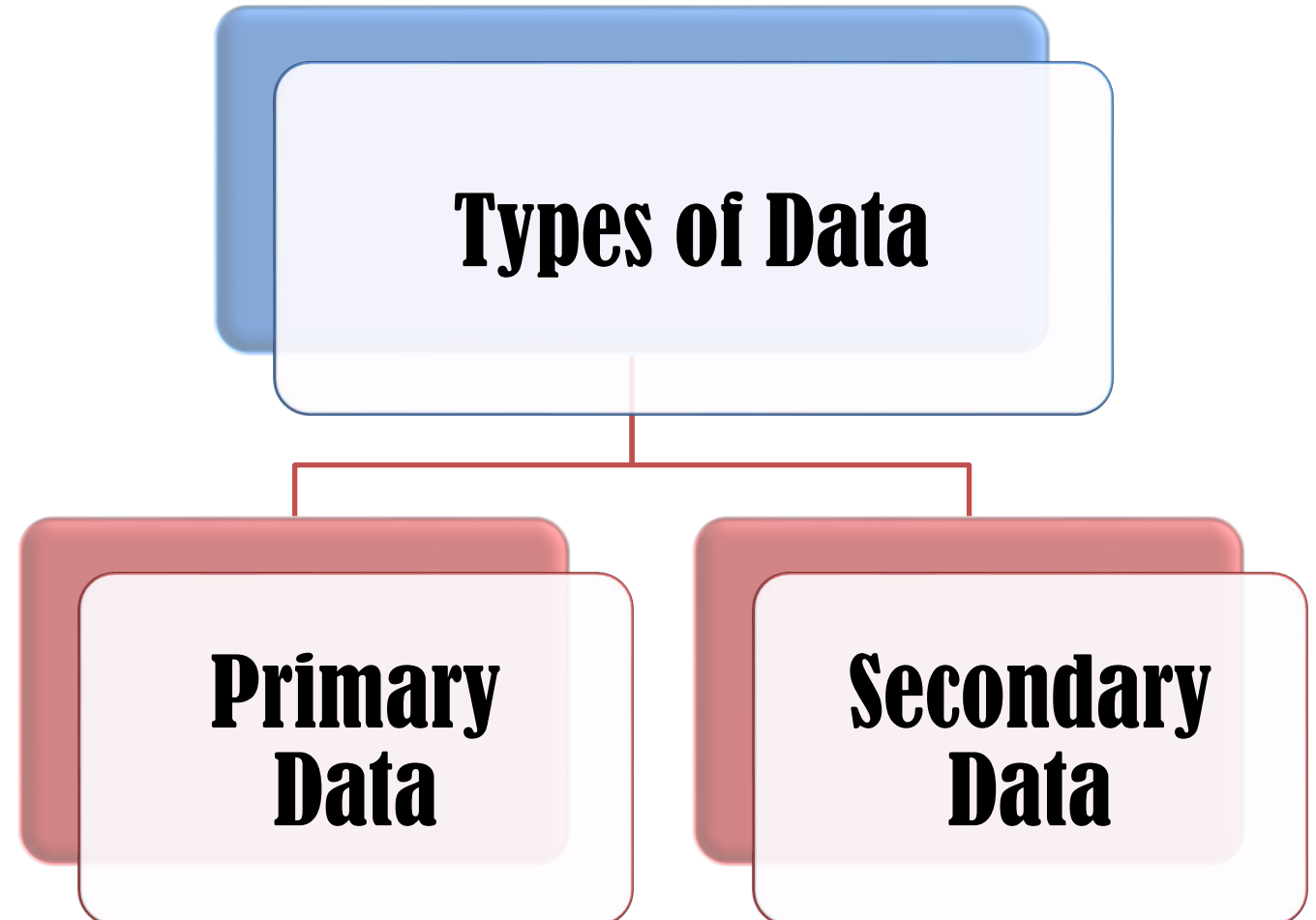
Primary Data

Data gathered and collected by the researcher direct from his respondents is called **Primary Data**

Advantages and Disadvantages of Primary Data

Advantages	Disadvantages
1. More accurate, reliable and up-to- date	1. Very costly
2. If the data needed by decision makers aren't available from other sources (secondary data), primary data has to be gathered	2. Time consuming
3. Primary sources usually explain how the data were gathered and what limitation exists to their use	3. Requires a lot of manpower
4. Usually satisfies the objectives of a research	

Secondary Data

Secondary data is a primary data that has been collected, processed and published for the use of other people.

Advantages and Disadvantages of Secondary Data

Advantages	Disadvantages
1. Require less time 2. Require less effort 3. Inexpensive data source	1. May contain errors in printing and transcription from the primary sources 2. Do not know the conditions under which the data were collected and summarized

Test of Understanding 2

1. Define each of the following terms:
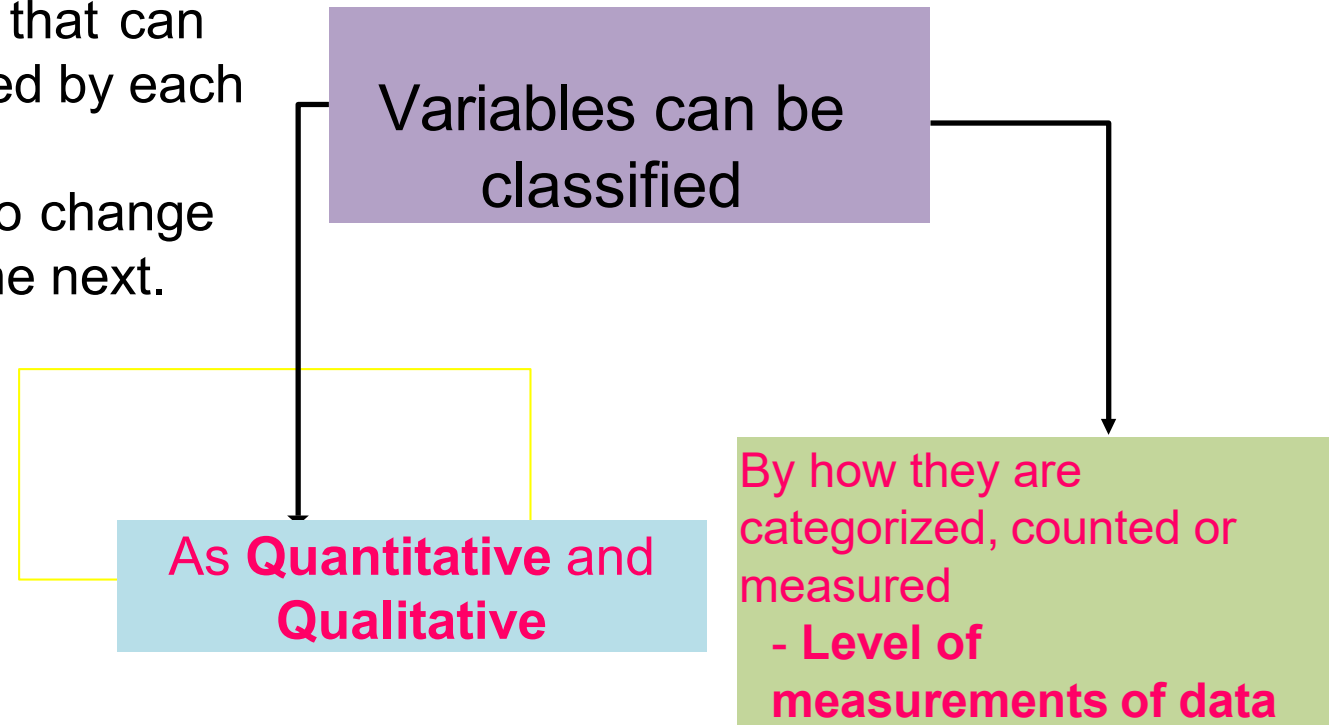   - Census
   - Sample
   - Pilot Test

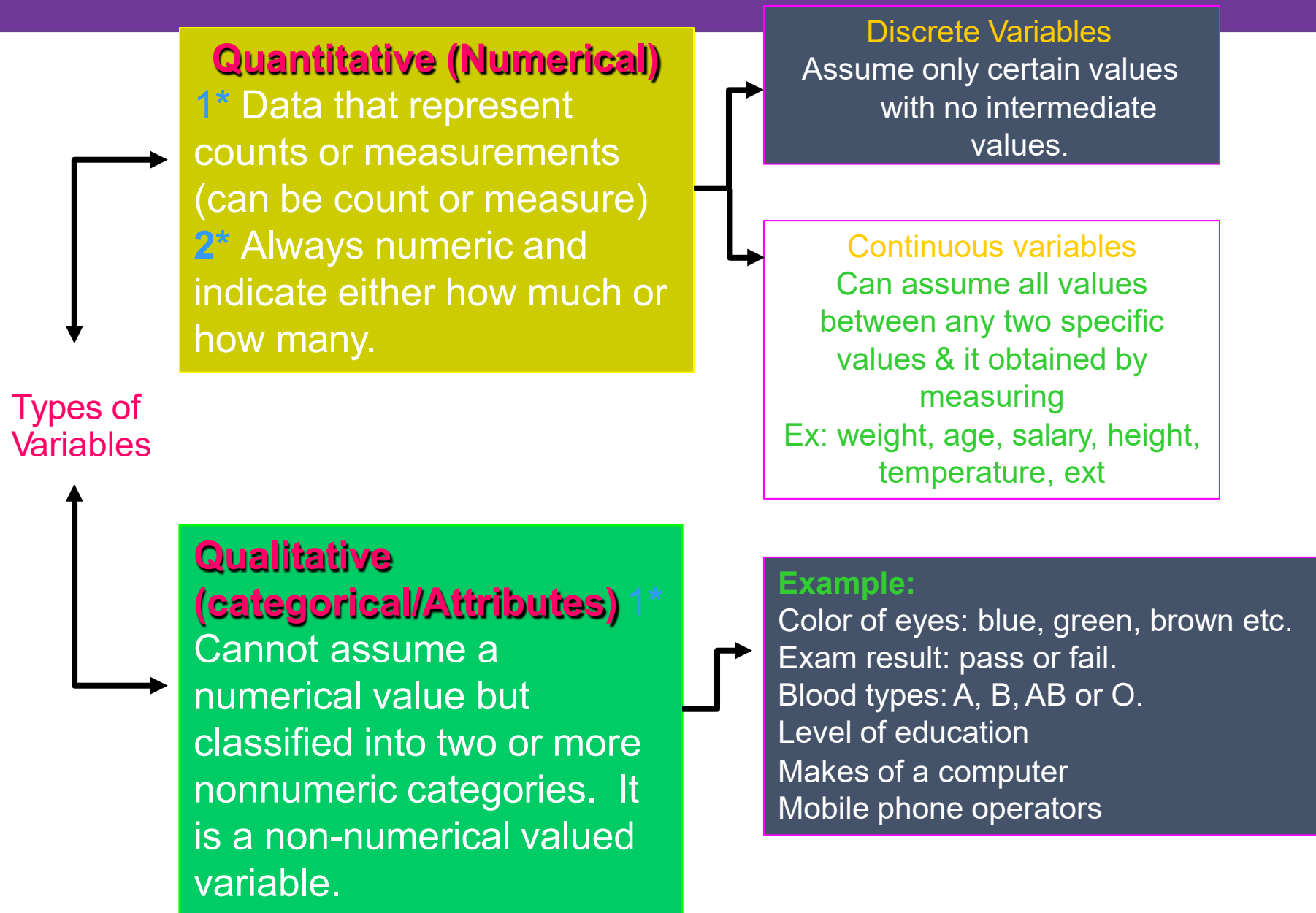2. What is the difference between Primary and Secondary data?

Variables & Types of Data

Statisticians collect information for *variables,* which describe the event.

Variables

❑ is a characteristic or attribute that can assume different values.

❑ is also a characteristics of interest, one that can be expressed as a number that possessed by each item under study.

❑ the value of this characteristics is likely to change or vary from one item in the data set to the next.

Variables can be classified

As **Quantitative** and **Qualitative**

By how they are categorized, counted or measured
- **Level of measurements of data**

Quantitative (Numerical)
1* Data that represent counts or measurements (can be count or measure)
2* Always numeric and indicate either how much or how many.

Discrete Variables
Assume only certain values with no intermediate values.

Continuous variables
Can assume all values between any two specific values & it obtained by measuring
Ex: weight, age, salary, height, temperature, ext

Types of Variables

Qualitative (categorical/Attributes) 1*
Cannot assume a numerical value but classified into two or more nonnumeric categories. It is a non-numerical valued variable.

Example:
Color of eyes: blue, green, brown etc.
Exam result: pass or fail.
Blood types: A, B, AB or O.
Level of education
Makes of a computer
Mobile phone operators

UNIVERSITI TEKNOLOGI MARA

Example

The Lemon Marketing Corporation has asked you for information about the car you drive. For each question, identify each of the types of data requested as either quantitative data or qualitative data. When quantitative data is requested, identify the variable as discrete or continuous.

1. What is the weight of your car?
2. In what city was your car made?
3. How many people can be seated in your car?
4. What's the distance traveled from your home to your school?
5. What's the color of your car?
6. How many cars are in your household?
7. What's the length of your car?
8. What's the normal operating temperature (in degree Fahrenheit) of your car's engine?
9. What gas mileage (miles per gallon) do you get in city driving?
10. Who made your car?
11. How many cylinders are there in your car's engine?
12. How many miles have you put on your car's current set of tyres?

Scales of Measurement

Nominal-level data	Ordinal-level data	Interval-level data	Ratio-level data
consists of names, labels or categories, gender, major at college. There is no natural or obvious ordering of nominal data. (Such as high to low). Arithmetic cannot be carried out on nominal data.	can be arranged in any particular order. However, no arithmetic can be done or performed on ordinal data.	similar to ordinal data, with the extra property that subtraction may be carried out on an interval data. There is no natural zero for interval data.	similar to interval data, with the extra property that division may be carried out on ratio data. There exists a natural zero for ratio data.
Examples: - Gender (male, female) -Eye color (blue, brown, green, hazel) - Political affiliation - Religious affiliation	Examples: - Grade - Judging - Rating scale -Ranking of items or persons - Socio economic status	Examples: - Temperature - IQ - Anxiety score	Examples: - Height - Weight - Time -'Money': Salary, Wages - Age - Shoe size

UNIVERSITI TEKNOLOGI MARA

Test of Understanding 3

1. Determine whether the following variable is qualitative, quantitative discrete or quantitative continuous.
   a) A classification of students by academic program.
   b) Number of hours parents spend time with their children per day.
   c) Speakers of a seminar are rated as excellent, good, average or poor.
   d) The number of cars parked at Stadium Darul Aman parking lot during the football match.

2. State the scale of measurement used for each of the following variables:
   a) Student gender: Male Female.
   b) Annual income of lecturers in a Math/Statistics Department.
   c) Course grades in Statistics subject: A+, A, A-, B+, B, B-,…, F.
   d) Years covered in American History: 1776-1876.

Sampling

Advantages of sampling
- Its cuts the costs of research.
- Suitable for destructive test.
- Accuracy of sample results.
- It cuts the time (period) of research.

Basic statistical terms in sampling:
- **Sampling frame** is a list of all the elements in a population under study (e.g. I.C., name and address)
- **Sampling unit** is the elements listed in the frame.
- The procedure for selecting the samples is known as **sample survey design**.
- **Sampling error** is the difference between a sample statistic and its corresponding population parameter.
- **Non-sampling error**
    i. Occurs in collecting, recording and tabulating of data (Coding and data entry errors, faulty measuring device)
    ii. It is a case of human mistakes, for example response errors (non-response or giving false information from respondents)
    iii. An inadequate sampling frame
    iv. Field errors

Simple Random Sampling (SRS)

❑ Used when true random sampling is essential. A list of all elements of the populations is needed i.e. the sampling frame.

❑ The target population must have the same characteristic (homogenous).

❑ A simple random sample can be selected using

    1) The lottery method (lucky draw)

    2) Random Number

Advantages

❖ Tends to be completely representative, i.e. a pretty good unbiased sample.
❖ Simple to apply.
❖ Analysis of data is reasonably easy and has a sound mathematical basis.

Disadvantages

➢ Complete list is difficult to obtain.
➢ Always a chance of drawing a misleading sample.
➢ Needs a larger sample size.

Systematic Random Sampling

- Appropriate where very **large numbers** are included in the target population and simple random sampling is difficult, or where lists are already grouped into sections or classes.

- A **list of all elements** of the populations is needed i.e. the sampling frame.

- The target population must have the **same characteristic** (homogenous)

Steps for implementing systematic random sampling:

i. Divide the population size by the sample size and round the result down to the nearest whole number, k.

$$\text{Interval, } k = \frac{Population\ Size}{Sample\ Size} = \frac{N}{n}$$

ii. Use a random number table or a lottery method to obtain a number, i, between 1 and k.

iii. Select for the sample those members of the population that are numbered i, $i+k$, $i+2k$, . . . until the intended sample size is obtained.

Systematic Random Sampling

Advantages

- It is usually quicker/faster than SRS.

- Easier to draw, without mistakes.

- Simple to apply.

- More precise than simple random sampling as more evenly spread over population.

Disadvantages

- It is not perfectly random since the first number chosen pre-determines the other elements.

Stratified Random Sampling

❑ Dividing the population into groups according some characteristic that is important to the study, then sampling from each group.

❑ Elements in each subgroup are homogenous

❑ Strata should be heterogeneous.

❑ Sample are selected in each subgroup.

❑ Elements are selected from each strata(SRS or Systematic)

Cluster Random Sampling

❑ Dividing the population into sections/clusters, then randomly select some of those cluster & then chose all members from those selected cluster

Differences between cluster and stratified sampling

Stratified Sampling	*Cluster Sampling*
• The analysis is done on elements within strata.	• The cluster is treated as a sampling unit so analysis is done on a population of clusters.
• A random sample is drawn from each of the strata.	• Only the selected clusters are studied.
• The main objective is increased precision.	• The main objective is to reduce costs by increasing sample efficiency.

Multi-stage Sampling

❏ This technique is an extension of cluster sampling procedure. Statisticians sometimes refer to cluster sampling as one stage sampling method.

❏ With "large" populations it is often necessary to carry out the sampling in 2 or more stages until the final number of sampling units is reached.

Advantages	*Disadvantages*
• Focuses on important subpopulation but ignore irrelevant ones. • Improves the accuracy of estimation. • Efficient.	● Can be difficult to select relevant stratification variables. ● Not useful when there no homogenous subgroup. ● Can be expensive. ● Requires accurate information about the run.

Quota Sampling

❑ Quota sampling divides the population into subgroups, which are then sampled in proportion to their occurrence in the population.

❑ Sampling frame is not needed.

❑ The researcher may choose any respondent he consider appropriate for his research.

Advantages

❑ Less costly.
❑ Administratively easy.
❑ Does not need any sampling frame.

Disadvantages

❖ Poor or biased judgment can lead to a non-representative sample.
❖ Within quota the sampling may be unrepresentative (e.g. all young, attractive females).
❖ Widely used social class grouping is subjective.

Convenience Sampling

❑ A convenience sample is a group of people who you can conveniently locate and administer the survey. Unfortunately, the convenience sample may not generalize to the target population.

Comparison between random sample and non- random sample

Random Sample

- Each member of the population has a probability of being selected.

- More complex, more time-consuming and usually more costly

Non Random Sampling

- Members are selected from the population in some non-random manner.

- The degree to which the sample differs from the population remains unknown.

1. What is sampling frame?

2. State two reasons why sampling is preferred to census.

3. What is the difference between probability and non-probability sampling?

4. Give four example of probability sampling technique.

Data Collection Methods

❑ Personal Interviews (Face-to-face interview)
❑ Telephone Interviews
❑ Direct (or Self-Administered) Questionnaires
❑ Mail or Postal Questionnaires
❑ Direct Observation

Designing a Questionnaire

✓ Questionnaires should be as short as possible.
✓ Avoid doubt, confusion, and vagueness.
✓ Avoid bias question.
✓ Avoid double-barreled questions.
✓ Avoid leading questions.
✓ Avoid asking questions that are beyond the respondents' capabilities.
✓ Avoid questions that involve calculation.

1. Define questionnaire.
2. Give four example of data collection method.
3. State three characteristics of a good questionnaire.

END OF SLIDES

PRESENTATIONS

THANK YOU

Prepared by:
Kamarul Ariffin Mansor
Department of Mathematical Sciences
UiTM Kedah Branch Campus