Chapter 1

General Overview

Introduction

Statistics: a science whereby inferences are made about specific random phenomena on the basis of relatively limited sample material.

- **Mathematical statistics:** concerns the development of new methods of statistical inference and requires detailed knowledge of abstract mathematics for its implementation.

- **Applied statistics:** involves application of mathematical statistical methods to specific subject areas such as economics, psychology, and public health.

Biostatistics: a branch of applied statistics that applies statistical methods to medical and biological problems.

- Standard statistical methods may not necessarily be applicable for all studies
- New biostatistical methods are developed by biostatisticians.

Role of Biostatistics in Medical Research

Observation:

Blood pressure readings of patient X obtained using

- Automatic measuring device = 115 mm Hg;
  highest reading = 130 mmHg
- Standard blood pressure cuff = 90 mm Hg

Why is there a difference in blood pressure readings between an automatic machine vs. a human observer?

Are the two methods of determining blood pressure comparable?

Study Question:

Are the methods of automatic vs. manual determination of blood pressure comparable?

To address this question, we designed and carried out the following small-scale study of blood pressure monitoring machines.

Planning step

Questions:

1. No. of machines to be tested
2. No. of participants for each machine to be tested
3. Order of taking measurements: Manual--> automated or vice versa (For our study, simultaneous readings were logistically not feasible)
4. Critical data to be captured via questionnaire to aid in comparison between the methods
5. Format of recording data to ease future data entry into computers
6. Checking accuracy of computerized data

Determination of test variables

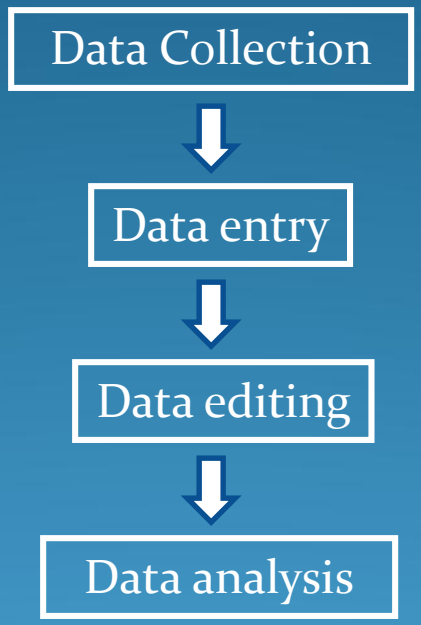
1. No. of machines to be tested = 4 ; since machines may or may not be comparable in quality
2. No. of participants = 100 people at each test location based on sample size determination method
3. To rule out any effects that the measurement method may have, the order of measurement was randomized (flipping a coin, using a table of random numbers, etc)

4. Critical data to be captured via questionnaire:
   - Age
   - Sex
   - Previous hypertension history
   - Body size (since this variable was seen to influence obtaining accurate readings)

5. Format of recording data to ease future data entry into computers
   - Each person assigned a unique identification number (ID)
   - Using a coding form that was keyed in and verified
   - Same coding form entered twice to ensure accuracy of records

6. Checking accuracy of computerized data
   - Using editing programs to check that all values of variables fell within a specific range
   - Outliers or aberrant values were manually checked

Next Steps



Data Analysis

Data obtained from the study can be summarized using descriptive statistics

Descriptive material can be numeric or graphic

- If numeric, data can be tabulated or presented as a frequency distribution

- If graphic, data can be summarized pictorially

Choice of numeric or graphic descriptive statistics is dependent on type of distribution of data.

1. Continuous data:
   - where there are infinite number of possible values (e.g., blood pressure measurements)
   - means and standard deviations may be used
2. Discrete data:
   - where there are only a few possible values (e.g., sex)
   - percentages of people for each value may be considered

Tabulated results of study under consideration

Table 1.1 Mean blood pressures and differences between machine and human readings at four locations

		Systolic blood pressure (mm Hg)					
		Machine		Human		Difference	
Location	Number of people	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation
A	98	142.5	21.0	142.0	18.1	0.5	11.2
B	84	134.1	22.5	133.6	23.2	0.5	12.1
C	98	147.9	20.3	133.9	18.3	14.0	11.7
D	62	135.4	16.7	128.5	19.0	6.9	13.6

Source: By permission of the American Heart Association, Inc.

Note: Meaningful data for all 100 people at each test site could not be obtained on account of a few not valid readings from the machines.

Missing data is common in biostatistics and should be anticipated at the planning stage.

Notice the apparent difference in blood pressure readings between machine vs. manual measurements in locations C and D.

Inferential Statistics

Determining whether the difference in blood pressure readings is “real” or “by chance”

Sample size = 98 people from the general population

Estimated mean difference = 14 mm Hg

Error in estimated mean difference = ?

True mean difference = d = ?

Inferring the characteristics of a population from a sample is the central concern of statistical inference.

To accomplish this aim, we need to develop a **probability model**, which would tell us how likely it is to obtain a 14-mm Hg difference between the two methods in a sample of 98 people if there were no real difference between the two methods over the entire population of users of the machine.

A small enough probability would indicate that the difference between the two methods is real.

For our study, we used a probability model based on *t* distribution.

The probability was found to be < 1 in 1000 for each of the machines at locations C and D.

The low probability indicated that there is a real difference between the automatic and manual method of blood pressure determination.

Further data analyses were carried out using a statistical package.

- A statistical package is a collection of statistical programs that describe data and perform various statistical tests on the data.
- A few statistical packages include SAS, SPSS, Stata, MINITAB, and Excel.

Final Step

Publication of manuscript

Essential data obtained in the data analysis step is included in the final manuscript for publishing

The End