The Science of Statistics

I. Introduction

The objective of statistics is to make inferences about a population based on information contained in a sample.

Population is a collection of all items (individuals, plants, corporations, …) in a study.

Sample is a portion of the population selected to represent the whole population.

Examples:

1. Cost of Medical Care:

   Population: All patients visiting UCDMC primary care.
   Sample: 510 patients selected randomly.
   Variables: Patient demographics
              Income
              Education
              Health Status
              ......
   Objective: Relate cost of care to patient characteristics such as age, gender, income, … and style of care controlling for patient health status.
2. Projection of Inmate Population and Costs:

Population: All inmates in California.
Variables: Number of inmates (monthly)
Inmate demographics
Type of felony

....
Policy changes
Different costs.

Objective: This is an observational study to predict the inmate population and costs.

3. Environmental Studies:

a. LA Air pollution

Population: Daily air pollution.
Variables: Carbon monoxide

....
Temperature

....
Daily mortality

Objective: Model mortality in terms of pollution levels. Regulations on emission control.

b. Sacramento Delta Water Pollution

Population: Daily water pollution.
Variables: Water levels
Discharges

....
Pollution level.

Objective: Predict the pollution levels. Possible regulations.
II. Parameter and Statistic

**Parameter:** A descriptive measure of a population.
Examples: Population mean, population median, …

**Statistic:** A descriptive measure of a sample.
Examples: Sample mean, sample median, …

III. Two Types of Statistical Analysis

a. **Descriptive Statistics:** Statistical methods used to develop tabular, graphical and/or numerical summaries of data

Example: Cost of care.

Primary care charges:

- Smallest = 0
- Q1 = 174.5
- Median = 303.5
- Q3 = 487.875
- Largest = 1929.7
- IQR = 313.375
- Outliers = (1929.7, 1893.1, 1429, 1215.5, 1168.5,)

![Primary Care Cost Frequency Chart](chart.png)
b. **Inferential Statistics**: The process of using data from a *sample* to draw conclusions about a *population*.

Examples: Cost of health care, projection of inmate population, environmental studies, …

**IV. Computers**

- Use computers to prepare numerical and graphical summaries of large amounts of data.
- Use statistical packages to model the data (process of estimation and hypothesis testing)
- Some popular packages are: JMP, MINITAB, SAS, SPSS, STATA, and R.
- Use word processors such as MS Word to produce reports.