Mat 131 Agenda Day 1: 1/22/03

Data Sheet Fill In
Index Cards:
   Name
   Address
   Phone Number
   Email Address ✓
   Why taking this course
   Something Unique about you
Attendance Roll Sheet ✓
Class Policy Sheet ✓
Assignments and Course Outline
Software/Video Guide
PowerPoint
Lecture Notes Chapter 1
Homework: Triola 1.1-1.4
BlackBoard: http://blackboard.mc3.edu
Your username is: << first letter first name – full last name – last 4 digits of Datatel ID >>
Password is: <<Datatel ID >>
For Example:
Student Name: John Smith  Datatel ID: 1234567
ID: jsmith4567
Password: 1234567
Web page: http://faculty.mc3.edu/rhofman/first.htm

Triola Web: http://www.aw.com/triolaessentials
Chapter 1
Introduction to Statistics

1-1 Overview
1-2 The Nature of Data
1-3 Uses and Abuses of Statistics
1-4 Design of Experiments
Overview

Statistics

Two Meanings

- Specific numbers
- Method of analysis
Statistics

- **Specific number**
  numerical measurement determined by a set of data

Example: **Twenty-three percent** of people polled believed that there are too many polls.
Method of analysis

a collection of methods for planning experiments, obtaining data, and then organizing, summarizing, presenting, analyzing, interpreting, and drawing conclusions based on the data.
Population

the complete collection of all elements (scores, people, measurements, and so on) to be studied. The collection is complete in the sense that it includes all subjects to be studied.
Sample

a subcollection of elements drawn from a population
Parameter

a numerical measurement describing some characteristic of a population

Greek

mean

Sample Statistic

\overline{x}
Parameter

a numerical measurement describing some characteristic of a population

population

parameter
Statistic

a numerical measurement describing some characteristic of a sample
Statistic

a numerical measurement describing some characteristic of a sample
Quantitative data

numbers representing counts or measurements
Qualitative (or categorical or attribute) data can be separated into different categories that are distinguished by some nonnumeric characteristics.
Quantitative data

numbers representing counts or measurements

Qualitative (or categorical or attribute) data

can be separated into different categories that are distinguished by some nonnumeric characteristics
Quantitative data

the incomes of college graduates

Qualitative (or categorical or attribute) data

the genders (male/female) of college graduates
Discrete data result when the number of possible values is either a finite number or a ‘countable’ number of possible values

0, 1, 2, 3, . . .
Continuous

(numerical) data result from infinitely many possible values that correspond to some continuous scale that covers a range of values without gaps, interruptions, or jumps.
Discrete

The number of eggs that hens lay; for example, 3 eggs a day.
Continuous

The amounts of milk that cows produce; for example, 2.343115 gallons a day.
nominal level of measurement

characterized by data that consist of names, labels, or categories only. The data cannot be arranged in an ordering scheme (such as low to high)
nominal level of measurement

characterized by data that consist of names, labels, or categories only. The data cannot be arranged in an ordering scheme (such as low to high).

Example: survey responses yes, no, undecided
ordinal level of measurement involves data that may be arranged in some order, but differences between data values either cannot be determined or are meaningless
 ordinal level of measurement

involves data that may be arranged in some order, but differences between data values either cannot be determined or are meaningless.

Example: Course grades A, B, C, D, or F
Interval level of measurement

like the ordinal level, with the additional property that the difference between any two data values is meaningful. However, there is no natural zero starting point (where none of the quantity is present)

\[ \text{-2 -1 0 1 2 3 -10°C} \]
interval level of measurement

like the ordinal level, with the additional property that the difference between any two data values is meaningful. However, there is no natural zero starting point (where none of the quantity is present)

Example: Years 1000, 2000, 1776, and 1492
ratio level of measurement

the interval level modified to include the natural zero starting point (where zero indicates that *none* of the quantity is present). For values at this level, differences and ratios are meaningful.
ratio level of measurement

the interval level modified to include the natural zero starting point (where zero indicates that none of the quantity is present). For values at this level, differences and ratios are meaningful.

Example: Prices of college textbooks
Levels of Measurement

- **Nominal** - categories only
- **Ordinal** - categories with some order
- **Interval** - differences but no natural starting point
- **Ratio** - differences and a natural starting point
*self-selected survey*
(or voluntary response sample)

one in which the respondents themselves decide whether to be included
Abuses of Statistics

- Bad Samples
- Small Samples
- Loaded Questions
- Misleading Graphs
Figure 1-1  Salaries of People with Bachelor’s Degrees and with High School Diplomas

(a) Bachelor Degree: $40,500
High School Diploma: $24,400

(b) Bachelor Degree: $40,500
High School Diploma: $24,400
We should analyze the **numerical** information given in the graph instead of being mislead by its general shape.
Double the length, width, and height of a cube, and the volume increases by a factor of eight.

Figure 1-2
Abuses of Statistics

- Bad Samples
- Small Samples
- Loaded Questions
- Misleading Graphs
- Pictographs
- Precise Numbers
- Distorted Percentages
- Partial Pictures
- Deliberate Distortions
Observational Study

observing and measuring specific characteristics without attempting to modify the subjects being studied
Experiment

apply some *treatment* and then observe its effects on the subjects
Designing an Experiment

- Identify your objective
- Collect sample data
- Use a random procedure that avoids bias
- Analyze the data and form conclusions
Confounding
occurs in an experiment when the effects from two or more variables cannot be distinguished from each other.
Replication

used when an experiment is repeated on a sample of subjects that is large enough so that we can see the true nature of any effects (instead of being misled by erratic behavior of samples that are too small)
Random Sample

members of the population are selected in such a way that each has an \textit{equal chance} of being selected
Simple Random Sample (of size $n$)

subjects selected in such a way that every possible sample of size $n$ has the same chance of being chosen
Random Sampling - selection so that each has an equal chance of being selected
Systematic Sampling - Select some starting point and then select every $K$th element in the population.
Convenience Sampling - use results that are readily available

Hey! Do you believe in the death penalty?
Stratified Sampling - subdivide the population into subgroups that share the same characteristic, then draw a sample from each stratum.
Cluster Sampling - divide the population into sections (or clusters); randomly select some of those clusters; choose all members from selected clusters.
Methods of Sampling

- Random
- Systematic
- Convenience
- Stratified
- Cluster
Definitions

- **Sampling Error**
  the difference between a sample result and the true population result; such an error results from chance sample fluctuations.

- **Nonsampling Error**
  sample data that are incorrectly collected, recorded, or analyzed (such as by selecting a biased sample, using a defective instrument, or copying the data incorrectly).
Chapter 1

Objectives:
1. Know the definition of statistics.
2. Distinguish between a population and a sample; parameter and a statistic.
3. Distinguish between qualitative and quantitative data.
4. Classify data using four levels of measurement: nominal, ordinal, interval, and ratio.
5. Distinguish between types of sampling: Random, Stratified, Systematic, Cluster, Convenience.
After reading the chapter and going through the PowerPoint fill in the following definitions. Make sure that you can apply the correct term in a given situation.

Definitions: Fill in
STATISTICS: Two meanings: Actual numbers; methods of analysis
POPULATION:
SAMPLE:
PARAMETER:
STATISTIC:
Difference between Quantitative Data and Qualitative (attribute) Data
Difference between Discrete and Continuous Data
Abuses of Statistics: Name 5 examples of improper uses of statistics
<table>
<thead>
<tr>
<th>Levels of Measurements</th>
<th>Definition</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominal</td>
<td>Names only</td>
<td>Gender</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Brands of cars</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Responses of Yes or No</td>
</tr>
<tr>
<td>Ordinal</td>
<td>Names with some order</td>
<td>Ratings: Excellent, Good, Bad</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Finishes: first, second, third, honor mention</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Size: Sub-compact, Compact, Mid-size</td>
</tr>
<tr>
<td>Interval</td>
<td>Differences but no &quot;zero&quot;</td>
<td>Temperature</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Military Time</td>
</tr>
<tr>
<td>Ratio</td>
<td>Differences and a zero</td>
<td>Weights Lengths, Distances</td>
</tr>
</tbody>
</table>
Definition: Difference between Observational Study and Experiment

Methods of Sampling:

<table>
<thead>
<tr>
<th>Method</th>
<th>Definition</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stratified</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Systematic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cluster</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conveni</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>