STA 291
Lecture 2

• Course web page: Updated: Office hour of Lab instructor.

• Statistics is the Science involving Data
• Example of data:

<table>
<thead>
<tr>
<th>Item Name</th>
<th>Price</th>
<th>In Stock?</th>
<th># in stock</th>
</tr>
</thead>
<tbody>
<tr>
<td>Silver cane</td>
<td>43.50</td>
<td>Yes</td>
<td>3</td>
</tr>
<tr>
<td>Top hat</td>
<td>29.99</td>
<td>No</td>
<td>0</td>
</tr>
<tr>
<td>Red shoes</td>
<td>35.00</td>
<td>No</td>
<td>0</td>
</tr>
<tr>
<td>Blue T-shirt</td>
<td>5.99</td>
<td>Yes</td>
<td>15</td>
</tr>
<tr>
<td>……</td>
<td>……</td>
<td>……</td>
<td>……</td>
</tr>
</tbody>
</table>

• More complicated data (time series): many of those tables over time….every quarter company have their financial report.

• A single variable value over time: Stock price over the time period of 20 years.
Basic Terminology

• **Variable**
  - a characteristic of a unit that can vary among subjects in the population/sample
  - Examples: gender, nationality, age, income, hair colour, height, disease status, grade in STA 291, state of residence, voting preference, weight, etc.

There are 4 variables displayed in the table on previous slide

Type of variables

• Categorical/Qualitative and Quantitative/numerical

• **Recall:**
  - A Variable is a characteristic of a unit that can vary among subjects in the data

• Within numerical variables: continuous or discrete.

• Within categorical variables: nominal or ordinal.

• Examples (ordinal): very satisfied, satisfied, unsatisfied.....
Qualitative Variables
(=Categorical Variables)

**Nominal** or Ordinal

- **Nominal**: gender, nationality, hair color, state of residence
- Nominal variables have a **scale of unordered categories**
- It does not make sense to say, for example, that green hair is greater/higher/better than orange hair

---

Qualitative (Categorical) Variables
Nominal or **Ordinal**

- **Ordinal**: Disease status, company rating, grade in STA 291. (best, good, fair, poor)
- Ordinal variables have a scale of ordered categories. They are often treated in a quantitative manner (GPA: A=4.0, B=3.0,…)

---

Quantitative Variables
(=numerical variables)

- **Quantitative**: age, income, height, price
- Quantitative variables are measured numerically, that is, for each subject, a number is observed
Example 1

• Vigild (1988) "Oral hygiene and periodontal conditions among 201 institutionalized elderly", Gerodontics, 4:140-145
• Variables measured
  – Nominal: Requires Assistance from Staff? Yes / No
  – Ordinal: Plaque Score
    No Visible Plaque - Small Amounts of Plaque - Moderate Amounts of Plaque - Abundant Plaque
  – Quantitative: Number of Teeth (discrete)

Example 2

• The following data are collected on newborns as part of a birth registry database
  • Ethnic background: African-American, Hispanic, Native American, Caucasian, Other
  • Infant’s Condition: Excellent, Good, Fair, Poor
  • Birthweight: in grams
  • Number of prenatal visits

Why is it important to distinguish between different types of data?

• Some statistical methods only work for quantitative variables, others are designed for qualitative variables.
You can treat variables in a less quantitative manner. (but lose information/accuracy.....sometimes for security reason).

- Examples include income, [20k or less, 20k to 40k, 40k to 60k, 60k and above] and
  - Height: Quantitative variable, continuous variable, measured in cm (or ft/in)
  - Can be treated as ordinal: short, average, tall
  - Can even be treated as nominal
    180cm-200cm, all others

Sometimes, ordinal variables are treated as quantitative: the quality of the photo prints rated by human with a score from 1 to 10.

**Discrete** and Continuous

- A variable is discrete if it can take on a finite number of values
- Examples: gender, nationality, hair color, disease status, company rating, grade in STA 291, state of residence
- Qualitative (categorical) variables are always discrete
- Quantitative variables can be discrete or continuous
Discrete and **Continuous**

- Continuous variables can take an *infinite continuum* of possible real number values
- Example: time spent on STA 291 homework
  - can be 63 min. or 85 min.
  - or 27.358 min. or 27.35769 min. or ...
  - can be *subdivided*
  - therefore *continuous*

---

Discrete or Continuous

- Another example: number of children
- can be 0, 1, 2, 3, ...
- can not be 1.5 or 2.768
- can not be *subdivided*
- therefore not *continuous* but *discrete*

---

- Data are increasingly getting larger. A few gigabyte is considered large 5 years ago
- Microsoft Excel often not enough. (64k rows by 256 columns)
- Data base software SQL etc.
- Data mining
Where do data come from?

- Two types of data collection method covered in this course:
  1. experiments
  2. polls

Second hand, from internet.....

Simple Random Sampling

- Each possible sample has the same probability of being selected. [no discrimination, no favoritism.]
- The sample size is usually denoted by \( n \).

Example: Simple Random Sampling

- Population of 4 students: Adam, Bob, Christina, Dana
- Select a simple random sample (SRS) of size \( n=2 \) to ask them about their smoking habits
- 6 possible samples of size \( n=2 \):
  1. A B
  2. A C
  3. A D
  4. B C
  5. B D
  6. C D
How to choose a SRS?

- Each of the six possible samples has to have the same probability of being selected
- For example, roll a die (or use a computer-generated random number) and choose the respective sample
- Online Sampling Applet

How not to choose a SRS?

- Ask Adam and Dana because they are in your office anyway
  - “convenience sample”
- Ask who wants to take part in the survey and take the first two who volunteer
  - “volunteer sampling”

Problems with Volunteer Samples

- The sample will poorly represent the population
- Misleading conclusions
- BIAS
- Examples: Mall interview, Street corner interview
Homework 1

• Due Jan 28, 11 PM.
• homework assignment:
  Log on to MyStatLab and create an account for this course. Complete one question with several multiple choices.

Attendance Survey Question

• On a 4”x6” index card (or little piece of paper)
  – write down your Name and 291 Section number
  – Today’s Question: (regarding prereq.)
  You have taken
  A. MA123, B. MA113, C. both, D. equiv.