S1600 #2
Data Presentation #1

January 14, 2016
Outline

1. Data Presentation #1
   - Statistics and Data
   - Variable Types
   - Summarizing Categorical Data
Statistics and Data

- **Statistics**
  - Broader sense: collection of techniques/procedures for analyzing data, i.e., Statistics = Data Analysis
  - Narrower sense: statistics = numbers derived from data

- **Data** = table/spreadsheet like collection of measurements made on a number of subjects
  - row = subject or case or observation
  - column = measurement or variable

CLASS DATA EXAMPLE: there are 10 subjects (students) and six variables: Sex, (Class) Level, GPA, Hours Taken, Transport, and Sleep Hours Last Night. Note: the column 'Student' is for observation labels.
Statistics and Data

**Statistics**

- Broader sense: collection of techniques/procedures for analyzing data, i.e., Statistics = Data Analysis
- Narrower sense: statistics = numbers derived from data

Data = table/spreadsheet like collection of measurements made on a number of subjects

- row = subject or case or observation
- column = measurement or variable

CLASS DATA EXAMPLE: there are 10 subjects (students) and six variables: Sex, (Class) Level, GPA, Hours Taken, Transport, and Sleep Hours Last Night. Note: the column 'Student' is for observation labels.
Statistics and Data

Statistics

- Broader sense: collection of techniques/procedures for analyzing data, i.e., Statistics = Data Analysis
- Narrower sense: statistics = numbers derived from data

Data = table/spreadsheet like collection of measurements made on a number of subjects

- row = subject or case or observation
- column = measurement or variable

CLASS DATA EXAMPLE: there are 10 subjects (students) and six variables: Sex, (Class) Level, GPA, Hours Taken, Transport, and Sleep Hours Last Night. Note: the column ‘Student’ is for observation labels.
Statistics and Data

Statistics

- Broader sense: collection of techniques/procedures for analyzing data, i.e., Statistics = Data Analysis
- Narrower sense: statistics = numbers derived from data

Data = table/spreadsheet like collection of measurements made on a number of subjects

- row = subject or case or observation
- column = measurement or variable

CLASS DATA EXAMPLE: there are 10 subjects (students) and six variables: Sex, (Class) Level, GPA, Hours Taken, Transport, and Sleep Hours Last Night. Note: the column ‘Student’ is for observation labels.
Statistics and Data

Statistics

- Broader sense: collection of techniques/procedures for analyzing data, i.e., Statistics = Data Analysis
- Narrower sense: statistics = numbers derived from data

Data = table/spreadsheet like collection of measurements made on a number of subjects

- row = subject or case or observation
- column = measurement or variable
- CLASS DATA EXAMPLE: there are 10 subjects (students) and six variables: Sex, (Class) Level, GPA, Hours Taken, Transport, and Sleep Hours Last Night. Note: the column ‘Student’ is for observation labels.
Statistics and Data

Statistics

- Broader sense: collection of techniques/procedures for analyzing data, i.e., Statistics = Data Analysis
- Narrower sense: statistics = numbers derived from data

Data = table/spreadsheet like collection of measurements made on a number of subjects

- row = subject or case or observation
- column = measurement or variable

CLASS DATA EXAMPLE: there are 10 subjects (students) and six variables: Sex, (Class) Level, GPA, Hours Taken, Transport, and Sleep Hours Last Night. Note: the column ‘Student’ is for observation labels.
Statistics and Data

Statistics

- Broader sense: collection of techniques/procedures for analyzing data, i.e., Statistics = Data Analysis
- Narrower sense: statistics = numbers derived from data

Data = table/spreadsheet like collection of measurements made on a number of subjects

- row = subject or case or observation
- column = measurement or variable

CLASS DATA EXAMPLE: there are 10 subjects (students) and six variables: Sex, (Class) Level, GPA, Hours Taken, Transport, and Sleep Hours Last Night. Note: the column ‘Student’ is for observation labels.
Statistics and Data

Statistics
- Broader sense: collection of techniques/procedures for analyzing data, i.e., Statistics = Data Analysis
- Narrower sense: statistics = numbers derived from data

Data = table/spreadsheet like collection of measurements made on a number of subjects
- row = subject or case or observation
- column = measurement or variable

CLASS DATA EXAMPLE: there are 10 subjects (students) and six variables: Sex, (Class) Level, GPA, Hours Taken, Transport, and Sleep Hours Last Night. Note: the column ‘Student’ is for observation labels.
Levels of Measurement

in increasing order of complexity

- nominal: pure labeling, no ordering
- ordinal: ordering exists, but not distance
- interval: distance exists, but not ratios, zero is arbitrary
- ratio: ratios exist, zero indicates the absence of such measurement
Levels of Measurement

in increasing order of complexity

- nominal: pure labeling, no ordering
- ordinal: ordering exists, but not distance
- interval: distance exists, but not ratios, zero is arbitrary
- ratio: ratios exist, zero indicates the absence of such measurement
Levels of Measurement
in increasing order of complexity

- nominal: pure labeling, no ordering
- ordinal: ordering exists, but not distance
- interval: distance exists, but not ratios, zero is arbitrary
- ratio: ratios exist, zero indicates the absence of such measurement
Levels of Measurement
in increasing order of complexity

- nominal: pure labeling, no ordering
- ordinal: ordering exists, but not distance
- interval: distance exists, but not ratios, zero is arbitrary
- ratio: ratios exist, zero indicates the absence of such measurement
Levels of Measurement

in increasing order of complexity

- nominal: pure labeling, no ordering
- ordinal: ordering exists, but not distance
- interval: distance exists, but not ratios, zero is arbitrary
- ratio: ratios exist, zero indicates the absence of such measurement
# CLASS DATA EXAMPLE

<table>
<thead>
<tr>
<th>Sex</th>
<th>Level</th>
<th>GPA</th>
<th>Hours</th>
<th>Transport</th>
<th>Sleep Hrs. Last Night</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>Sophomore</td>
<td>3.10</td>
<td>32</td>
<td>Car</td>
<td>7.0</td>
</tr>
<tr>
<td>M</td>
<td>Junior</td>
<td>3.20</td>
<td>66</td>
<td>Car</td>
<td>8.0</td>
</tr>
<tr>
<td>F</td>
<td>Senior</td>
<td>3.49</td>
<td>94</td>
<td>Bus</td>
<td>8.0</td>
</tr>
<tr>
<td>M</td>
<td>Senior</td>
<td>2.68</td>
<td>89</td>
<td>Walk</td>
<td>10.0</td>
</tr>
<tr>
<td>F</td>
<td>Junior</td>
<td>3.73</td>
<td>69</td>
<td>Bicycle</td>
<td>8.0</td>
</tr>
<tr>
<td>:</td>
<td>:</td>
<td>:</td>
<td>:</td>
<td>:</td>
<td>:</td>
</tr>
<tr>
<td>M</td>
<td>Senior</td>
<td>3.10</td>
<td>96</td>
<td>Walk</td>
<td>3.0</td>
</tr>
</tbody>
</table>
## CLASS DATA EXAMPLE

<table>
<thead>
<tr>
<th>Sex</th>
<th>Level</th>
<th>GPA</th>
<th>Hours</th>
<th>Transport</th>
<th>Sleep Hrs. Last Night</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>Sophomore</td>
<td>3.10</td>
<td>32</td>
<td>Car</td>
<td>7.0</td>
</tr>
<tr>
<td>M</td>
<td>Junior</td>
<td>3.20</td>
<td>66</td>
<td>Car</td>
<td>8.0</td>
</tr>
<tr>
<td>F</td>
<td>Senior</td>
<td>3.49</td>
<td>94</td>
<td>Bus</td>
<td>8.0</td>
</tr>
<tr>
<td>M</td>
<td>Senior</td>
<td>2.68</td>
<td>89</td>
<td>Walk</td>
<td>10.0</td>
</tr>
<tr>
<td>F</td>
<td>Junior</td>
<td>3.73</td>
<td>69</td>
<td>Bicycle</td>
<td>8.0</td>
</tr>
<tr>
<td>:</td>
<td>:</td>
<td>:</td>
<td>:</td>
<td>:</td>
<td>:</td>
</tr>
<tr>
<td>M</td>
<td>Senior</td>
<td>3.10</td>
<td>96</td>
<td>Walk</td>
<td>3.0</td>
</tr>
</tbody>
</table>

**nominal: Sex, Transport**
## CLASS DATA EXAMPLE

<table>
<thead>
<tr>
<th>Sex</th>
<th>Level</th>
<th>GPA</th>
<th>Hours</th>
<th>Transport</th>
<th>Sleep Hrs. Last Night</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>Sophomore</td>
<td>3.10</td>
<td>32</td>
<td>Car</td>
<td>7.0</td>
</tr>
<tr>
<td>M</td>
<td>Junior</td>
<td>3.20</td>
<td>66</td>
<td>Car</td>
<td>8.0</td>
</tr>
<tr>
<td>F</td>
<td>Senior</td>
<td>3.49</td>
<td>94</td>
<td>Bus</td>
<td>8.0</td>
</tr>
<tr>
<td>M</td>
<td>Senior</td>
<td>2.68</td>
<td>89</td>
<td>Walk</td>
<td>10.0</td>
</tr>
<tr>
<td>F</td>
<td>Junior</td>
<td>3.73</td>
<td>69</td>
<td>Bicycle</td>
<td>8.0</td>
</tr>
<tr>
<td>:</td>
<td>:</td>
<td>:</td>
<td>:</td>
<td>:</td>
<td>:</td>
</tr>
<tr>
<td>M</td>
<td>Senior</td>
<td>3.10</td>
<td>96</td>
<td>Walk</td>
<td>3.0</td>
</tr>
</tbody>
</table>

*nominal: Sex, Transport  Ordinal: Level*
## CLASS DATA EXAMPLE

<table>
<thead>
<tr>
<th>Sex</th>
<th>Level</th>
<th>GPA</th>
<th>Hours Taken</th>
<th>Transport</th>
<th>Sleep Hrs. Last Night</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>Sophomore</td>
<td>3.10</td>
<td>32</td>
<td>Car</td>
<td>7.0</td>
</tr>
<tr>
<td>M</td>
<td>Junior</td>
<td>3.20</td>
<td>66</td>
<td>Car</td>
<td>8.0</td>
</tr>
<tr>
<td>F</td>
<td>Senior</td>
<td>3.49</td>
<td>94</td>
<td>Bus</td>
<td>8.0</td>
</tr>
<tr>
<td>M</td>
<td>Senior</td>
<td>2.68</td>
<td>89</td>
<td>Walk</td>
<td>10.0</td>
</tr>
<tr>
<td>F</td>
<td>Junior</td>
<td>3.73</td>
<td>69</td>
<td>Bicycle</td>
<td>8.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>Senior</td>
<td>3.10</td>
<td>96</td>
<td>Walk</td>
<td>3.0</td>
</tr>
</tbody>
</table>

**nominal**: Sex, Transport  
**Ordinal**: Level  
**interval**: GPA
# CLASS DATA EXAMPLE

<table>
<thead>
<tr>
<th>Sex</th>
<th>Level</th>
<th>GPA</th>
<th>Hours</th>
<th>Transport</th>
<th>Sleep Hrs. Last Night</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>Sophomore</td>
<td>3.10</td>
<td>32</td>
<td>Car</td>
<td>7.0</td>
</tr>
<tr>
<td>M</td>
<td>Junior</td>
<td>3.20</td>
<td>66</td>
<td>Car</td>
<td>8.0</td>
</tr>
<tr>
<td>F</td>
<td>Senior</td>
<td>3.49</td>
<td>94</td>
<td>Bus</td>
<td>8.0</td>
</tr>
<tr>
<td>M</td>
<td>Senior</td>
<td>2.68</td>
<td>89</td>
<td>Walk</td>
<td>10.0</td>
</tr>
<tr>
<td>F</td>
<td>Junior</td>
<td>3.73</td>
<td>69</td>
<td>Bicycle</td>
<td>8.0</td>
</tr>
<tr>
<td>:</td>
<td>:</td>
<td>:</td>
<td>:</td>
<td>:</td>
<td>:</td>
</tr>
<tr>
<td>M</td>
<td>Senior</td>
<td>3.10</td>
<td>96</td>
<td>Walk</td>
<td>3.0</td>
</tr>
</tbody>
</table>

- **nominal**: Sex, Transport
- **ordinal**: Level
- **interval**: GPA
- **ratio**: Hours Taken, Sleep Hrs. Last Night
Numerical Versus Categorical Variables

- Categorical: nominal & ordinal
- Quantitative (or numerical): interval & ratio

Categorical ≠ non-numerical. E.g., SSN
iClicker Question 2.1

Which of the following is nominal?

1. Final exam score
2. Student’s Gender
3. Class level (Freshman, Sophomore, ...)
4. Student’s height
5. Student’s high school GPA
For each of the following variables, identify the type of variable (categorical, numerical).

I. Temperature (in Fahrenheit) of office building
II. Duration (in minutes) of flight between two locations.

1. I. Numeric, and II. Categorical
2. I. Categorical, and II. Numeric
3. There is no correct match
4. I. Categorical, and II. Categorical
5. I. Numeric, and II. Numeric
Dependent Versus Independent Variables

terms used in cause-and-effect studies

- independent variable: a probable cause
- dependent variable: outcome being affected or caused

For the vaccine-autism study, the amount of mercury in the preservative in vaccine is the independent variable, whereas the 42 measures of brain function are dependent variables. Note: Likert items & Likert scale.
Relative Frequency Table
American Community Survey Data

Payment Type is nominal which can be summarized by a relative frequency table:

<table>
<thead>
<tr>
<th>Type</th>
<th>Frequency</th>
<th>Rel. Freq. (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mortgage</td>
<td>44</td>
<td>57.9</td>
</tr>
<tr>
<td>Rent</td>
<td>20</td>
<td>26.3</td>
</tr>
<tr>
<td>None</td>
<td>12</td>
<td>15.8</td>
</tr>
</tbody>
</table>

Total          76  100.0

Frequency = count
Relative Frequency Table

American Community Survey Data

Payment Type is nominal which can be summarized by a relative frequency table:

<table>
<thead>
<tr>
<th>Type</th>
<th>Frequency</th>
<th>Rel. Freq. (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mortgage</td>
<td>44</td>
<td>57.9</td>
</tr>
<tr>
<td>Rent</td>
<td>20</td>
<td>26.3</td>
</tr>
<tr>
<td>None</td>
<td>12</td>
<td>15.8</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>76</strong></td>
<td><strong>100.0</strong></td>
</tr>
</tbody>
</table>

Frequency = count
Rel. Freq.(%) = (Frequency / Total) × 100%
Bar Chart

a plot of frequencies/relative frequencies

bar height = frequency (or rel. freq)
Pie Chart

- Mortgage
- Rent
- None

A bad graphical tool
Pie Chart

- Mortgage: 57.9%
- Rent: 26.3%
- None: 15.8%

A bad graphical tool.
Pie Chart

Mortgage: 57.9%
None: 15.8%
Rent: 26.3%

a bad graphical tool
Pie Chart, continued

impossible to discern the order of wedges’ sizes
Pie Chart, continued

impossible to discern the order of wedges' sizes until adding actual rel. freq.
impossible to discern the order of wedges’ sizes until adding actual rel. freq. A simple (sorted) relative frequency table will do:

<table>
<thead>
<tr>
<th>Category</th>
<th>rel. freq.</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>36%</td>
</tr>
<tr>
<td>C</td>
<td>33%</td>
</tr>
<tr>
<td>B</td>
<td>31%</td>
</tr>
</tbody>
</table>
iClicker Question 2.3

In a frequency/relative frequency table, the total of the relative frequencies is

1. 50%
2. 75%
3. 100%
4. 150%
5. Cannot determine