Class 2

Daniel B. Rowe, Ph.D.

Department of Mathematics, Statistics, and Computer Science
Agenda:
Briefly Review Math
Briefly Review Chapter 1
Lecture Chapter 2
Review Math
1. Summation Notation
\[ \sum_{i=1}^{n} f(x_i) = f(x_1) + f(x_2) + \ldots + f(x_n) \]

2. Factorials
\[ n! = n \times (n-1) \times (n-2) \times \ldots \times 2 \times 1 \]

3. Computations
\[ x = 20, \ y = 14, \ s = 16, \ w = -2, \ m = 15, \ n = 10 \]
Compute \[ x + y \cdot \frac{\sqrt{s}}{n} = 25.6 \]

4. Simple Linear Equations
\[ 2 - 2x = 3x + 3 \quad x = -1/5 \]
Recap Chapter 1
Chapter 1: Statistics

Daniel B. Rowe, Ph.D.

Department of Mathematics, Statistics, and Computer Science
1: Statistics
1.1 Americans Here’s Looking at you

Statistics is all around us!

How much time between Internet usage?

Figure from Johnson & Kuby, 2012.
1: Statistics
1.1 What is Statistics?

**Population:** A collection, or set, of individuals, objects, or events whose properties are to be analyzed.

**Sample:** Subset of the population.

**Variable:** A characteristic of interest about each individual element of a population or sample.

**Data value:** The value of the variable associated with one element of a population or sample.

**Parameter:** A numerical value summarizing all the data of an entire population.

**Statistic:** A numerical value summarizing the sample data.
1: Statistics
1.1 What is Statistics?

**Data**: The set of values collected from the variable from each of the elements that belong to the sample.
Lecture Chapter 2
Chapter 2: Descriptive Analysis and Presentation of Single-Variable Data

Daniel B. Rowe, Ph.D.

Department of Mathematics, Statistics, and Computer Science
2: Descriptive Analysis and Single Variable Data

2.1 Graphs - Qualitative Data

**Pie charts (circle graphs) and bar graphs**: Graphs that are used to summarize **qualitative**, or attribute, or categorical data.

<table>
<thead>
<tr>
<th>Type of Operation</th>
<th>Number of Cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thoracic</td>
<td>20</td>
</tr>
<tr>
<td>Bones and joints</td>
<td>45</td>
</tr>
<tr>
<td>Eye, ear, nose, and throat</td>
<td>58</td>
</tr>
<tr>
<td>General</td>
<td>98</td>
</tr>
<tr>
<td>Abdominal</td>
<td>115</td>
</tr>
<tr>
<td>Urologic</td>
<td>74</td>
</tr>
<tr>
<td>Proctologic</td>
<td>65</td>
</tr>
<tr>
<td>Neurosurgery</td>
<td>23</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>498</strong></td>
</tr>
</tbody>
</table>

Figures from Johnson & Kuby, 2012.
2: Descriptive Analysis and Single Variable Data

2.1 Graphs - Qualitative Data

**Pie charts (circle graphs) and bar graphs**: Graphs that are used to summarize qualitative, or attribute, or categorical data.

<table>
<thead>
<tr>
<th>Type of Operation</th>
<th>Number of Cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thoracic</td>
<td>20</td>
</tr>
<tr>
<td>Bones and joints</td>
<td>45</td>
</tr>
<tr>
<td>Eye, ear, nose, and throat</td>
<td>58</td>
</tr>
<tr>
<td>General</td>
<td>98</td>
</tr>
<tr>
<td>Abdominal</td>
<td>115</td>
</tr>
<tr>
<td>Urologic</td>
<td>74</td>
</tr>
<tr>
<td>Proctologic</td>
<td>65</td>
</tr>
<tr>
<td>Neurosurgery</td>
<td>23</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>498</strong></td>
</tr>
</tbody>
</table>

Figures from Johnson & Kuby, 2012.
2: Descriptive Analysis and Single Variable Data
2.1 Graphs - Qualitative Data

Example:

Enter and highlight data
2: Descriptive Analysis and Single Variable Data

2.1 Graphs - Qualitative Data

Example:

Select Insert

Select column bar
2: Descriptive Analysis and Single Variable Data
2.1 Graphs - Qualitative Data

Show Example in Excel!
2: Descriptive Analysis and Single Variable Data

2.1 Graphs – Quantitative Data

**Dotplot Display:** Displays the data of a sample by representing each data value with a dot positioned above the scale.

Figures from Johnson & Kuby, 2012.

Rowe, D.B.
2: Descriptive Analysis and Single Variable Data

2.1 Graphs - Quantitative Data

**Distribution**: The Pattern of variability displayed by the data of a variable. The distribution displays the frequency of each value of the variable.

2.2 Frequency Distributions and Histograms

**Frequency distribution**: A listing, often expressed in chart form, that pairs values of a variable with their frequency.
2: Descriptive Analysis and Single Variable Data
2.2 Frequency Distributions and Histograms

1. Identify the high score \((H=98)\) and the low score \((L=39)\).

\[
\text{range} = H - L = 98 - 39 = 59
\]

Figure from Johnson & Kuby, 2012.
2. Select the number of classes \((m=7)\) and a class width \((c=10)\) (These are subjective and depend on how you feel. But the larger \(n\), the more classes and smaller \(c\) you should have, the smaller \(n\), the fewer classes you should have and larger \(c\).)

\[mc = 70\] a little larger than the range = 59.

Figure from Johnson & Kuby, 2012.
2: Descriptive Analysis and Single Variable Data

2.2 Frequency Distributions and Histograms

3. Pick a starting point and set up class boundaries

\[35 \leq x < 45, \quad 45 \leq x < 55, \quad 55 \leq x < 65\]

\[65 \leq x < 75, \quad 75 \leq x < 85, \quad 85 \leq x < 95, \quad 95 \leq x < 105\]

Figure from Johnson & Kuby, 2012.
2: Descriptive Analysis and Single Variable Data
2.2 Frequency Distributions and Histograms

Statistics Exam Scores

<table>
<thead>
<tr>
<th>Score</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>60</td>
<td>47</td>
</tr>
<tr>
<td>58</td>
<td>64</td>
</tr>
<tr>
<td>70</td>
<td>64</td>
</tr>
<tr>
<td>72</td>
<td>77</td>
</tr>
<tr>
<td>82</td>
<td>95</td>
</tr>
<tr>
<td>88</td>
<td>50</td>
</tr>
<tr>
<td>72</td>
<td>77</td>
</tr>
<tr>
<td>67</td>
<td>72</td>
</tr>
<tr>
<td>66</td>
<td>86</td>
</tr>
<tr>
<td>68</td>
<td>50</td>
</tr>
<tr>
<td>98</td>
<td>86</td>
</tr>
<tr>
<td>90</td>
<td>77</td>
</tr>
<tr>
<td>77</td>
<td>64</td>
</tr>
<tr>
<td>86</td>
<td>95</td>
</tr>
<tr>
<td>74</td>
<td>70</td>
</tr>
<tr>
<td>72</td>
<td>70</td>
</tr>
<tr>
<td>74</td>
<td>64</td>
</tr>
<tr>
<td>88</td>
<td>70</td>
</tr>
<tr>
<td>74</td>
<td>72</td>
</tr>
<tr>
<td>77</td>
<td>58</td>
</tr>
<tr>
<td>39</td>
<td>89</td>
</tr>
<tr>
<td>90</td>
<td>44</td>
</tr>
<tr>
<td>63</td>
<td>55</td>
</tr>
<tr>
<td>68</td>
<td>85</td>
</tr>
<tr>
<td>97</td>
<td>82</td>
</tr>
<tr>
<td>68</td>
<td>83</td>
</tr>
</tbody>
</table>

Class Number | Class Tally | Boundaries | Frequency |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>35 ≤ x &lt; 45</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>45 ≤ x &lt; 55</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>55 ≤ x &lt; 65</td>
<td>7</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>65 ≤ x &lt; 75</td>
<td>13</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>75 ≤ x &lt; 85</td>
<td>11</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>85 ≤ x &lt; 95</td>
<td>11</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>95 ≤ x ≤ 105</td>
<td>4</td>
</tr>
</tbody>
</table>

Figures from Johnson & Kuby, 2012.
2: Descriptive Analysis and Single Variable Data

2.2 Frequency Distributions and Histograms

<table>
<thead>
<tr>
<th>Boundaries</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>$35 \leq x &lt; 45$</td>
<td>2</td>
</tr>
<tr>
<td>$45 \leq x &lt; 55$</td>
<td>2</td>
</tr>
<tr>
<td>$55 \leq x &lt; 65$</td>
<td>7</td>
</tr>
<tr>
<td>$65 \leq x &lt; 75$</td>
<td>13</td>
</tr>
<tr>
<td>$75 \leq x &lt; 85$</td>
<td>11</td>
</tr>
<tr>
<td>$85 \leq x &lt; 95$</td>
<td>11</td>
</tr>
<tr>
<td>$95 \leq x \leq 105$</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>50</td>
</tr>
</tbody>
</table>

Figures from Johnson & Kuby, 2012.
2: Descriptive Analysis and Single Variable Data
2.2 Frequency Distributions and Histograms

<table>
<thead>
<tr>
<th>Boundaries</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>35 ≤ x &lt; 45</td>
<td>2</td>
</tr>
<tr>
<td>45 ≤ x &lt; 55</td>
<td>2</td>
</tr>
<tr>
<td>55 ≤ x &lt; 65</td>
<td>7</td>
</tr>
<tr>
<td>65 ≤ x &lt; 75</td>
<td>13</td>
</tr>
<tr>
<td>75 ≤ x &lt; 85</td>
<td>11</td>
</tr>
<tr>
<td>85 ≤ x &lt; 95</td>
<td>11</td>
</tr>
<tr>
<td>95 ≤ x ≤ 105</td>
<td>4</td>
</tr>
</tbody>
</table>

Divide all by 50 to get percent

Figures from Johnson & Kuby, 2012.
We were able to present the information contained in this sample of data using graphical methods.

Now let’s summarize the information contained in the sample of data using numerical summary measures.

Describe the measures of central tendency (sample mean, sample median, sample mode), then describe some the measures of dispersion, then use them in a toy example.
2: Descriptive Analysis and Single Variable Data
2.3 Measures of Central Tendency

**Sample Mean:** The usual average you are familiar with. Represented by $\bar{x}$ called "x-bar." p. 63

Simply add up all the values and divide by the number values.

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

Remember the sigma notation we reviewed?

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + \ldots + x_n$$

**Round-off Rule:** When rounding a number, let's keep one more decimal place than the original numbers.

Rowe, D.B.
2: Descriptive Analysis and Single Variable Data
2.3 Measures of Central Tendency

**Sample Median:** The thing in the middle of the road! LOL. Statistics humor.

Middle value when data ordered. 50% above, 50% below Represented by \( \tilde{x} \) called “x-tilde.” p. 64

\[ \tilde{x} = \text{middle value} \]

Order data from smallest to largest. If \( n \) odd, \( d(\tilde{x}) = \frac{n + 1}{2} \) value

If \( n \) even, \( d(\tilde{x}) \) avg. of \( \frac{n}{2} \) and \( \frac{n}{2} + 1 \) values

Rowe, D.B.
2: Descriptive Analysis and Single Variable Data
2.3 Measures of Central Tendency

**Sample Mode:** The value that happens most often in sample.
Represented by \( \hat{x} \) called “\( x \)-hat.” p. 66

Order data from smallest to largest.
Count how many time each value occurs.
Take the one with the highest count for \( \hat{x} \).

If two or more values in a sample are tied for the highest frequency, we say that there is **no mode**.
2: Descriptive Analysis and Single Variable Data
2.3 Measures of Central Tendency

The measures of central tendency characterize the center of the distribution of data values.

There are other measures called measures of dispersion that characterize the spread or variability in the data.
Range: The difference between the highest data value \((H)\) and lowest data values \((L)\). p. 74

\[
range = \text{high value} - \text{low value}
\]

\[
range = H - L
\]
Deviation from the mean: The difference between the data value $x_i$ and the sample mean $\bar{x}$. p. 74

\[ i^{th} \text{ deviation from mean} = x_i - \bar{x} \]

There can be $n$ of these because we have $x_1, x_2, ..., x_n$. 
2: Descriptive Analysis and Single Variable Data
2.4 Measures of Dispersion

**Sample Variance:** The mean of the squared deviations using \( n-1 \) as a divisor. p. 75

There are two equivalent formulas that can be used.

\[
s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2
\]

and

\[
s^2 = \frac{1}{n-1} \left\{ \sum_{i=1}^{n} x_i^2 - \left[ \left( \sum_{i=1}^{n} x_i \right)^2 / n \right] \right\}
\]

where \( x_i \) is \( i^{th} \) data value, \( \bar{x} \) is sample mean, \( n \) is sample size.
2: Descriptive Analysis and Single Variable Data
2.4 Measures of Dispersion

**Sample Standard Deviation:** Square root of the sample variance. Has same units data values and sample mean.

\[ s = \sqrt{s^2} \]

**Note:**
Sample variance \( s^2 \) uses the entire sample and a denominator \( n-1 \)!
Population variance \( \sigma^2 \) the entire population and a denominator \( N \)!

\[ s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 \]

\( n \) items in the sample
\( N \) items in the population
\( n \leq N, n < N \) for a sample

Rowe, D.B.
2: Descriptive Analysis and Single Variable Data

2.4 Measures of Central Tendency

Example:
Data values: 1, 2, 2, 3, 4

Sample Mean = ?

\[ \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \]
2: Descriptive Analysis and Single Variable Data
2.4 Measures of Central Tendency

Example:
Data values: 1, 2, 2, 3, 4

Sample Median = ?

\[ \tilde{x} = \text{middle value} \]

Order data from smallest to largest.
If the number of data values is odd, take the middle value as the median.
If the number of data values is even, take the average of the middle two.
2: Descriptive Analysis and Single Variable Data
2.4 Measures of Central Tendency

Example:
Data values: 1, 2, 2, 3, 4

Sample Mode = ?

\[ \hat{x} = ? \]

Order data from smallest to largest. Count how many times each value occurs. Take the one with the highest count.
2: Descriptive Analysis and Single Variable Data

2.4 Measures of Dispersion

Example:
Data values: 1,2,2,3,4

Sample Variance = ?

\[ s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 \]
2: Descriptive Analysis and Single Variable Data
2.4 Measures of Dispersion

Example:
Data values: 1,2,2,3,4

\[ s^2 = 1.3 \]

Sample Standard Deviation=?

\[ s = \sqrt{s^2} \]
2: Descriptive Analysis and Single Variable Data

Example: In Excel:

![Excel Insert Data](image)

Insert Data.

Select a cell.
2: Descriptive Analysis and Single Variable Data

Example: In Excel:

=AVERAGE(A1:A5)
=MEDIAN(A1:A5)
=MODE(A1:A5)
=VAR(A1:A5)
=STDEV(A1:A5)

Type in What you want. Answer appears.
2: Descriptive Analysis and Single Variable Data

Example: In Excel:

- Type in What you want.
- Answer appears.

```
=AVERAGE(A1:A5)
=MEDIAN(A1:A5)
=MODE(A1:A5)
=VAR(A1:A5)
=STDEV(A1:A5)
```
2: Descriptive Analysis and Single Variable Data

Questions?

Homework: Read Chapter 2 including 2.6 and 2.7. Chapters 2 # 8, 35, 39 (try), 75, 97, 105