Class 5

Daniel B. Rowe, Ph.D.

Department of Mathematics, Statistics, and Computer Science
Agenda:

University Closed.
Exam 1 on Thursday.
Review slides follow.
Review Chapters 1 – 3.1
(Exam 1 Chapters)

Just the highlights!
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**

\[ 2 - 2x = 3x + 3 \quad x = -1/5 \]
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.

- **Qualitative**
  - Nominal (names)
  - Ordinal (ordered)
- **Quantitative**
  - Discrete (gap)
  - Continuous (continuum)
1: Statistics

1.1 What is Statistics?

**Qualitative variable:** A variable that describes or categorizes an element of a population.

**Nominal variable:** A qualitative variable that characterizes an element of a population. No ordering. No arithmetic.

**Ordinal variable:** A qualitative variable that incorporates an ordered position, or ranking.

**Quantitative variable:** A variable that quantifies an element of a population.

**Discrete variable:** A quantitative variable that can assume a countable number of values. Gap between successive values.

**Continuous variable:** A quantitative variable that can assume an uncountable number of values. Continuum of values.
2: Descriptive Analysis and Single Variable Data

2.1 Graphs - Qualitative Data

Circle (pie) graphs and bar graphs:
Circle is parts to whole as angle.
Bar graph is amount in each category as rectangular areas.

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

2.2 Frequency Distributions and Histograms

Statistics Exam Scores [TA02-06]

<table>
<thead>
<tr>
<th>Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>60</td>
</tr>
<tr>
<td>58</td>
</tr>
<tr>
<td>70</td>
</tr>
<tr>
<td>72</td>
</tr>
<tr>
<td>47</td>
</tr>
<tr>
<td>64</td>
</tr>
<tr>
<td>64</td>
</tr>
<tr>
<td>77</td>
</tr>
<tr>
<td>82</td>
</tr>
<tr>
<td>95</td>
</tr>
<tr>
<td>70</td>
</tr>
<tr>
<td>72</td>
</tr>
<tr>
<td>95</td>
</tr>
<tr>
<td>74</td>
</tr>
<tr>
<td>70</td>
</tr>
<tr>
<td>86</td>
</tr>
<tr>
<td>88</td>
</tr>
<tr>
<td>72</td>
</tr>
<tr>
<td>58</td>
</tr>
<tr>
<td>78</td>
</tr>
<tr>
<td>94</td>
</tr>
<tr>
<td>80</td>
</tr>
<tr>
<td>98</td>
</tr>
<tr>
<td>90</td>
</tr>
<tr>
<td>68</td>
</tr>
<tr>
<td>77</td>
</tr>
<tr>
<td>39</td>
</tr>
<tr>
<td>55</td>
</tr>
<tr>
<td>85</td>
</tr>
<tr>
<td>82</td>
</tr>
<tr>
<td>97</td>
</tr>
<tr>
<td>77</td>
</tr>
<tr>
<td>83</td>
</tr>
<tr>
<td>86</td>
</tr>
</tbody>
</table>

Figures from Johnson & Kuby, 2012.

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

2.3 Measures of Central Tendency

Sample Mean: Usual average, p. 63

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

Sample Median: Middle value, p. 64

- \( n \) odd, \( \tilde{x} = \frac{n+1}{2} \) value
- \( n \) even, avg \( \frac{n}{2} \) & \( \frac{n}{2} + 1 \) values

Sample Mode: Most often, p. 66

\( \hat{x} = \text{most often} \)

Measures of central tendency characterize center of distribution.

Measures of dispersion characterize the variability in the data.
2: Descriptive Analysis and Single Variable Data

2.4 Measures of Dispersion

Range: \( H - L \), p. 74

Deviation from mean: value minus sample mean, p. 74

\(^{i}\text{th deviation from mean} = x_i - \bar{x}\)

Sample Variance: avg squared dev using \( n-1 \) in den, p. 76

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

Sample Standard Deviation: \( s = \sqrt{s^2} \)
2: Descriptive Analysis and Single Variable Data
2.3, 2.4 Measures of Central Tendency and Dispersion

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

\[
\bar{x} = 2.4 \quad \hat{x} = 2 \quad \tilde{x} = 2
\]

\[
s^2 = 1.3 \quad s = 1.1
\]

\[
\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \quad \hat{x} = \text{most often value} \quad \tilde{x} = \text{middle value}
\]

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

Measures of Position: Quartiles - ranked data into quarters

$L = \text{lowest value}$
$H = \text{highest value}$
$Q_2 = \text{median}$
$Q_1 = 25\% \text{ smaller}$
$Q_3 = 75\% \text{ smaller}$
$IQR = Q_3 - Q_1$
2: Descriptive Analysis and Single Variable Data
2.5 Measures of Position

Measures of Position: percentiles - rank data into $100^{th}$s

$L = $ lowest value

$H = $ highest value

$P_k = $ value where $k\%$ are smaller

\[
\frac{nk}{100}
\]

$P_k$ halfway between value and next one average of $A^{th}$ and $(A+1)^{th}$ values

$p_k$ is value in next largest position, $B+1$ value

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

2.5 Measures of Position

**Standard score, or z-score:** The position a particular value of $x$ has relative to the mean, measured in standard deviations.

$$z_i = \frac{i^{th \text{ value} - \text{mean}}}{\text{std. dev.}} = \frac{x_i - \bar{x}}{s}$$

There can be $n$ of these because we have $x_1, x_2, \ldots, x_n$. 
3: Descriptive Analysis and Bivariate Data
3.1 Bivariate Data: two qualitative

Cross-tabulation tables or contingency tables

Example:
Construct a $2 \times 3$ table.
Know different %ages.

<table>
<thead>
<tr>
<th>Name</th>
<th>Gender</th>
<th>Major</th>
<th>Name</th>
<th>Gender</th>
<th>Major</th>
<th>Name</th>
<th>Gender</th>
<th>Major</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adams</td>
<td>M</td>
<td>LA</td>
<td>Feeney</td>
<td>M</td>
<td>T</td>
<td>McGowan</td>
<td>M</td>
<td>BA</td>
</tr>
<tr>
<td>Argento</td>
<td>F</td>
<td>BA</td>
<td>Flanigan</td>
<td>M</td>
<td>LA</td>
<td>Mowers</td>
<td>F</td>
<td>BA</td>
</tr>
<tr>
<td>Baker</td>
<td>M</td>
<td>LA</td>
<td>Hodge</td>
<td>F</td>
<td>LA</td>
<td>Ornt</td>
<td>M</td>
<td>T</td>
</tr>
<tr>
<td>Bennett</td>
<td>F</td>
<td>LA</td>
<td>Holmes</td>
<td>M</td>
<td>T</td>
<td>Palmer</td>
<td>F</td>
<td>LA</td>
</tr>
<tr>
<td>Brand</td>
<td>M</td>
<td>T</td>
<td>Jopson</td>
<td>F</td>
<td>T</td>
<td>Pullen</td>
<td>M</td>
<td>T</td>
</tr>
<tr>
<td>Brock</td>
<td>M</td>
<td>BA</td>
<td>Kee</td>
<td>M</td>
<td>BA</td>
<td>Rattan</td>
<td>M</td>
<td>BA</td>
</tr>
<tr>
<td>Chun</td>
<td>F</td>
<td>LA</td>
<td>Kleeberg</td>
<td>M</td>
<td>LA</td>
<td>Sherman</td>
<td>F</td>
<td>LA</td>
</tr>
<tr>
<td>Crain</td>
<td>M</td>
<td>T</td>
<td>Light</td>
<td>M</td>
<td>BA</td>
<td>Small</td>
<td>F</td>
<td>T</td>
</tr>
<tr>
<td>Cross</td>
<td>F</td>
<td>BA</td>
<td>Linton</td>
<td>F</td>
<td>LA</td>
<td>Tate</td>
<td>M</td>
<td>BA</td>
</tr>
<tr>
<td>Ellis</td>
<td>F</td>
<td>BA</td>
<td>Lopez</td>
<td>M</td>
<td>T</td>
<td>Yamamoto</td>
<td>M</td>
<td>LA</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gender</th>
<th>LA</th>
<th>BA</th>
<th>T</th>
<th>Row Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>18</td>
</tr>
<tr>
<td>F</td>
<td>6</td>
<td>4</td>
<td>2</td>
<td>12</td>
</tr>
<tr>
<td>Col. Total</td>
<td>11</td>
<td>10</td>
<td>9</td>
<td>30</td>
</tr>
</tbody>
</table>

M = male  
F = female  
LA = liberal arts  
BA = business admin  
T = technology

Figures from Johnson & Kuby, 2012.
3: Descriptive Analysis and Bivariate Data

3.1 Bivariate Data: one qualitative and one quantitative

Example:

<table>
<thead>
<tr>
<th>Design A (n = 6)</th>
<th>Design B (n = 6)</th>
<th>Design C (n = 6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>37 36 38 34 40 32</td>
<td>33 35 38 34 42 34</td>
<td>40 39 40 41 43</td>
</tr>
</tbody>
</table>

Figures from Johnson & Kuby, 2012.

Vertical box-and-whiskers

Rowe, D.B.
3: Descriptive Analysis and Bivariate Data
3.1 Bivariate Data: two quantitative, Scatter Diagram

Example: Push-ups

<table>
<thead>
<tr>
<th>Student</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Push-ups, $x$</td>
<td>27</td>
<td>22</td>
<td>15</td>
<td>35</td>
<td>30</td>
<td>52</td>
<td>35</td>
<td>55</td>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td>Sit-ups, $y$</td>
<td>30</td>
<td>26</td>
<td>25</td>
<td>42</td>
<td>38</td>
<td>40</td>
<td>32</td>
<td>54</td>
<td>50</td>
<td>43</td>
</tr>
</tbody>
</table>

**Input variable:** independent variable, $x$.

**Output variable:** dependent variable, $y$.

**Scatter Diagram:** A plot of all the ordered pairs of bivariate data on a coordinate axis system. 

$(x,y)$ ordered pairs.

Figures from Johnson & Kuby, 2012.

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

Questions?

Homework: Chapter 1 # 7, 9, 11, 41, 49a
vocabulary on page 27.
Chapter 2 # 8, 35, 75, 97, 105,
115, 123c-d, 129, 137
Chapter 3 # 3, 7, 15