Biology Education Core Competencies (Vision and Change, 2011)

1. Ability to apply the process of science:
   Biology is evidence based and grounded in the formal practices of observation, experimentation, and hypothesis testing.

2. Ability to use quantitative reasoning:
   Biology relies on applications of quantitative analysis and mathematical reasoning

3. Ability to use modeling and simulation
   (plus 3 more…)

Course Materials
- Statistics: Unlocking the Power of Data by Lock, Lock, Locke Morgan, Lock, and Locke
- WileyPlus necessary; used for homework
  - Details on how to register/purchase WileyPlus
  - WileyPlus comes with ebook; hardcopy is optional
  - WileyPlus includes text, embedded practice problems, videos for every example and learning goal, homework
- iclicker
  - Register by Friday, 9/4, at clickers.psu.edu

Main Course Website
- www.personal.psu.edu/klm47/Courses/STAT250/Fall2015/schedule.html
  - Lecture slides and course documents posted here

Other Course Websites
- Main: lecture slides, labs, documents, info
- WileyPlus: textbook, videos, practice problems, homework
- Piazza: discussion, questions, communication (use Piazza rather than email)
- ANGEL: grades

Keys to Success
- Come to class ready to think and be engaged
- Come to lab ready to think and be engaged
- Do the homework and give it an honest effort
- Read the textbook or watch videos if confused
- Do lots of practice problems
- Stay on top of the material
Introduction to Data

SECTION 1.1
• Data
• Cases and variables
• Categorical and quantitative variables
• Using data to answer a question

Why Statistics?
• Statistics is all about DATA
  o Collecting DATA
  o Describing DATA – summarizing, visualizing
  o Analyzing DATA
• Data are everywhere!
• You will have to make decisions based on data,
or evaluate decisions someone else has made based on data
• (This is particularly true in the health sciences!)

Data
• Data are a set of measurements taken on a set of individual units
• Usually data is stored and presented in a dataset, comprised of variables measured on cases

Cases and Variables
We obtain information about cases or units.
A variable is any characteristic that is recorded for each case.
• Generally each case makes up a row in a dataset, and each variable makes up a column

National Health and Nutrition Examination Survey

<table>
<thead>
<tr>
<th>State</th>
<th>Age</th>
<th>Gender</th>
<th>Race</th>
<th>Height</th>
<th>Weight</th>
<th>Waist</th>
<th>Race</th>
<th>BMI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Africa</td>
<td>20</td>
<td>Female</td>
<td>White</td>
<td>65</td>
<td>125</td>
<td>NA</td>
<td>Male</td>
<td>14.8</td>
</tr>
<tr>
<td>Asia</td>
<td>30</td>
<td>Male</td>
<td>Black</td>
<td>75</td>
<td>150</td>
<td>NA</td>
<td>Male</td>
<td>17.1</td>
</tr>
<tr>
<td>Europe</td>
<td>40</td>
<td>Female</td>
<td>White</td>
<td>80</td>
<td>180</td>
<td>NA</td>
<td>Female</td>
<td>20.3</td>
</tr>
<tr>
<td>North America</td>
<td>50</td>
<td>Male</td>
<td>Black</td>
<td>90</td>
<td>200</td>
<td>NA</td>
<td>Male</td>
<td>23.6</td>
</tr>
<tr>
<td>South America</td>
<td>60</td>
<td>Female</td>
<td>White</td>
<td>100</td>
<td>220</td>
<td>NA</td>
<td>Female</td>
<td>26.9</td>
</tr>
<tr>
<td>Europe</td>
<td>70</td>
<td>Male</td>
<td>White</td>
<td>110</td>
<td>240</td>
<td>NA</td>
<td>Male</td>
<td>29.1</td>
</tr>
<tr>
<td>South America</td>
<td>80</td>
<td>Female</td>
<td>Black</td>
<td>120</td>
<td>260</td>
<td>NA</td>
<td>Female</td>
<td>32.3</td>
</tr>
<tr>
<td>North America</td>
<td>90</td>
<td>Male</td>
<td>White</td>
<td>130</td>
<td>280</td>
<td>NA</td>
<td>Male</td>
<td>35.5</td>
</tr>
</tbody>
</table>

Countries of the World

<table>
<thead>
<tr>
<th>Country</th>
<th>Land Area</th>
<th>Population</th>
<th>Rural</th>
<th>Urban</th>
<th>Health</th>
<th>Internet</th>
<th>Birth Rate</th>
<th>Life Expectancy</th>
<th>HIV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Afghanistan</td>
<td>652230</td>
<td>29021099</td>
<td>75</td>
<td>3.7</td>
<td>1.7</td>
<td>14.6</td>
<td>43.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Algeria</td>
<td>2381740</td>
<td>343734248</td>
<td>34.8</td>
<td>10.6</td>
<td>10.2</td>
<td>20.8</td>
<td>72.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>American Samoa</td>
<td>200</td>
<td>66107</td>
<td>7.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Andorra</td>
<td>470</td>
<td>83810</td>
<td>11.1</td>
<td>21.3</td>
<td>70.5</td>
<td>10.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Angola</td>
<td>1246700</td>
<td>18020668</td>
<td>43.3</td>
<td>6.8</td>
<td>3.1</td>
<td>42.9</td>
<td>47.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Antigua and Barbuda</td>
<td>440</td>
<td>86634</td>
<td>69.5</td>
<td>11</td>
<td>11.7</td>
<td>75.3</td>
<td>0.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Argentina</td>
<td>2738690</td>
<td>39882980</td>
<td>12.7</td>
<td>28.1</td>
<td>17.3</td>
<td>75.3</td>
<td>0.5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Diet Coke and Calcium**

<table>
<thead>
<tr>
<th>Drink</th>
<th>Calcium Excreted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diet cola</td>
<td>50</td>
</tr>
<tr>
<td>Diet cola</td>
<td>62</td>
</tr>
<tr>
<td>Diet cola</td>
<td>48</td>
</tr>
<tr>
<td>Diet cola</td>
<td>55</td>
</tr>
<tr>
<td>Diet cola</td>
<td>58</td>
</tr>
<tr>
<td>Diet cola</td>
<td>61</td>
</tr>
<tr>
<td>Diet cola</td>
<td>58</td>
</tr>
<tr>
<td>Diet cola</td>
<td>56</td>
</tr>
<tr>
<td>Water</td>
<td>48</td>
</tr>
<tr>
<td>Water</td>
<td>46</td>
</tr>
<tr>
<td>Water</td>
<td>54</td>
</tr>
<tr>
<td>Water</td>
<td>46</td>
</tr>
<tr>
<td>Water</td>
<td>53</td>
</tr>
<tr>
<td>Water</td>
<td>53</td>
</tr>
<tr>
<td>Water</td>
<td>48</td>
</tr>
</tbody>
</table>

**Kidney Cancer**

Counties with the highest kidney cancer death rates


---

**Data Applicable to You**

- Think of a potential dataset (it doesn’t have to actually exist) that you would be interested in analyzing:
  - What are the cases?
  - What are the variables?
  - What interesting questions could it help you answer?

---

**Kidney Cancer**

If the values in the kidney cancer dataset are rates of kidney cancer deaths, then what are the cases?

(a) The people living in the US
(b) The counties of the US

---

**Kidney Cancer**

If the values in the kidney cancer dataset are yes/no, then what are the cases?

(a) The people living in the US
(b) The counties of the US
Categorical versus Quantitative

- Variables are classified as either **categorical** or **quantitative**:
  - A **categorical** variable divides the cases into groups
  - A **quantitative** variable measures a numerical quantity for each case

Kidney Cancer

If the cases in the kidney cancer dataset are counties, then the measured variable is...

(a) Categorical
(b) Quantitative

Kidney Cancer

If the cases in the kidney cancer dataset are people, then the measured variable is...

(a) Categorical
(b) Quantitative

Variables

For each of the following situations:
- What are the variables?
- Is each variable categorical or quantitative?
1. Are children with higher exposure to pesticides more likely to develop ADHD?
2. Does exercise make you smarter?
3. Can dogs detect cancer?
4. Do males find females more attractive if they wear red?

(We’ll explore all of these questions during the course!)
To Do

• Read Section 1.1
• Due Friday, 8/28: Take the pretest
• Due Friday, 9/4: Section 1.1 HW
• If you haven’t already...
  ○ Get WileyPlus
  ○ Get a clicker and register by 9/4

Why Statistics?
http://www.youtube.com/watch?v=nTBZuQR7dRc&feature=youtu.be