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...)

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
- Do lots of practice problems
- Read the textbook or watch videos
- Stay on top of the material
Why Statistics?

- Statistics is all about DATA
  - Collecting DATA
  - Describing DATA – summarizing, visualizing
  - 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

Countries of the World

<table>
<thead>
<tr>
<th>Country</th>
<th>Land Area</th>
<th>Population</th>
<th>Rural</th>
<th>Health</th>
<th>Internet</th>
<th>Birth Rate</th>
<th>Life Expectancy</th>
<th>HIV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Afghanistan</td>
<td>652230</td>
<td>29021099</td>
<td>76</td>
<td>3.7</td>
<td>1.7</td>
<td>46.5</td>
<td>43.9</td>
<td></td>
</tr>
<tr>
<td>Albania</td>
<td>27400</td>
<td>3143291</td>
<td>53.3</td>
<td>8.2</td>
<td>23.9</td>
<td>14.6</td>
<td>76.5</td>
<td></td>
</tr>
<tr>
<td>Algeria</td>
<td>2381740</td>
<td>34373426</td>
<td>34.8</td>
<td>10.6</td>
<td>10.2</td>
<td>20.8</td>
<td>72.4</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>
</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>
</tr>
<tr>
<td>Angola</td>
<td>1246700</td>
<td>18020688</td>
<td>43.3</td>
<td>6.8</td>
<td>3.1</td>
<td>42.9</td>
<td>47.2</td>
<td></td>
</tr>
<tr>
<td>Antigua and Barbuda</td>
<td>440</td>
<td>86634</td>
<td>69.5</td>
<td>11</td>
<td>75</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Argentina</td>
<td>2736690</td>
<td>39828980</td>
<td>8</td>
<td>13.7</td>
<td>28.1</td>
<td>17.3</td>
<td>75.3</td>
<td>0.5</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>45</td>
</tr>
<tr>
<td>Water</td>
<td>53</td>
</tr>
<tr>
<td>Water</td>
<td>46</td>
</tr>
<tr>
<td>Water</td>
<td>54</td>
</tr>
<tr>
<td>Water</td>
<td>53</td>
</tr>
<tr>
<td>Water</td>
<td>48</td>
</tr>
</tbody>
</table>
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

### 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

Explanatory and Response

If we are using one variable to help us understand or predict values of another variable, we call the former the explanatory variable and the latter the response variable.

Examples:
- Does meditation help reduce stress?
- Does sugar consumption increase hyperactivity?

Variables

For each of the following situations:
- What are the variables?
- Is each variable categorical or quantitative?
- Identify the explanatory and response variables.

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!)

Summary

- Data are everywhere, and pertain to a wide variety of topics
- A dataset is usually comprised of variables measured on cases
- Variables are either categorical or quantitative
- Data can be used to provide information about essentially anything we are interested in and want to collect data on!
To Do

• Read Section 1.1
• Due Friday, 1/16: Take the two pretests
  o Pretest 1
  o Pretest 2
• Due Friday, 1/23: Section 1.1 HW
• If you haven’t already...
  o Get the textbook with WileyPlus
  o Get a clicker and register it on ANGEL by 1/23

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