What is Biostatistics?

Math?

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S_{\bar{X}_1 - \bar{X}_2}} = \frac{\bar{X}_1 - \bar{X}_2}{S_{\bar{X}_1 - \bar{X}_2}}$$

$$S_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{(N_1 - 1)s_1^2 + (N_2 - 1)s_2^2}{N_1 + N_2 - 2} \left[ \frac{1}{N_1} + \frac{1}{N_2} \right]}$$
Graphs?

Source
100% of people who use statistics in casual conversations are annoying.

Source

I used to think correlation implied causation. Then I took a statistics class. Now I don’t. Sounds like the class helped. Well, maybe.

Source
Biostatistics

- Theoretical mathematics: building blocks for other researchers
- Applied math: use math to solve a problem like building an estimator
- Biostatistics; using stats with biological data
Brain imaging

Check out this interactive visualization!

Source
Environmental Statistics

Genomics

Source
Activity trackers

Data Science overview

Cleaning data
One type of clean data: tidy data

In tidy data:

1. Each variable forms a column.
2. Each observation forms a row.
3. Each type of observational unit forms a table.

---

<table>
<thead>
<tr>
<th></th>
<th>treatmenta</th>
<th>treatmentb</th>
</tr>
</thead>
<tbody>
<tr>
<td>John Smith</td>
<td>—</td>
<td>2</td>
</tr>
<tr>
<td>Jane Doe</td>
<td>16</td>
<td>11</td>
</tr>
<tr>
<td>Mary Johnson</td>
<td>5</td>
<td>1</td>
</tr>
</tbody>
</table>

messy

<table>
<thead>
<tr>
<th></th>
<th>treatmenta</th>
<th>treatmentb</th>
</tr>
</thead>
<tbody>
<tr>
<td>John Smith</td>
<td>—</td>
<td>2</td>
</tr>
<tr>
<td>Jane Doe</td>
<td>16</td>
<td>11</td>
</tr>
<tr>
<td>Mary Johnson</td>
<td>5</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>name</th>
<th>trt</th>
<th>result</th>
</tr>
</thead>
<tbody>
<tr>
<td>John Smith</td>
<td>a</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Jane Doe</td>
<td>a</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>Mary Johnson</td>
<td>a</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>John Smith</td>
<td>b</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Jane Doe</td>
<td>b</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>Mary Johnson</td>
<td>b</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

Source

You’ll learn Stata

```bash
use afg_worldbank_2016.dta
twoway (scatter SPRURTOTLZG Year), /*
    */ title(Rural population growth over time) /*
    */ subtitle(Afghanistan)
```
Modern biostatistician

- Participates in the experimental design, that is the question the team wants to answer
- Is a full participant in the research
- Learns how to clean data
- Models the data given the question of interest
- Contributes in interpreting the data and suggesting the next steps

Communicating results

- Understand the main concepts so you can explain them: means understanding why we chose a particular method
- Proper graphics
- Interpretation of the results

Dr. McGready

Let's take a look at how John McGready introduces biostatistics

- Open lecture 1 of the course Statistical Reasoning I
Exercise

Divide in two teams and come up with a 2 minute overview of Biostatistics. Imagine that you will record a 2 minute video to get students excited about your Biostatistics course.

Jeff Leek Data Analysis intro

Source

Slides license

[Image] BY-NC-SA