Ecology Lab Introduction

I thought this was ecology—not statistics

Modern ecology includes statistics

- Overkill or essential?

<table>
<thead>
<tr>
<th>Groups</th>
<th>Count</th>
<th>Sum</th>
<th>Average</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>23</td>
<td>1343.6</td>
<td>58.41739</td>
<td>820.607</td>
</tr>
<tr>
<td>1</td>
<td>13</td>
<td>1234.8</td>
<td>94.98462</td>
<td>136.6497</td>
</tr>
</tbody>
</table>

ANOVA

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>11105.87</td>
<td>1</td>
<td>11105.87</td>
<td>19.17416</td>
<td>0.000108</td>
</tr>
<tr>
<td>Within Groups</td>
<td>19693.15</td>
<td>34</td>
<td>579.2103</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>30799.02</td>
<td>35</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Three points to remember...

- First, look at the data!
- Second, see if you can apply the ‘O test’
- Third, results come first, statistics second

Graphical tips (1)

- Above all else, show the data
- Avoid distortion—don’t lie with numbers
- For written reports, use a figure caption; for oral reports, use a descriptive title

Thanks to Chris Hill, Edward Tufte, and Jack Webster

Today’s lab

- Good graphical techniques
- Basic statistics
- A brief write-up

Channelized streams have more open canopy

<table>
<thead>
<tr>
<th>Channel type</th>
<th>% Open canopy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Natural</td>
<td>20</td>
</tr>
<tr>
<td>Channelized</td>
<td>40</td>
</tr>
</tbody>
</table>

Anscombe’s quartet

- How statistics can unintentionally lie to you
- Four very different datasets, all with the same basic statistics

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>X mean</td>
<td>9</td>
</tr>
<tr>
<td>X variance</td>
<td>11</td>
</tr>
<tr>
<td>Y mean</td>
<td>7.5</td>
</tr>
<tr>
<td>Y variance</td>
<td>4.122</td>
</tr>
</tbody>
</table>

Needmore
Iotla

Graphical tips (1)

- Above all else, show the data
- Avoid distortion—don’t lie with numbers
- For written reports, use a figure caption; for oral reports, use a descriptive title

Thanks to Chris Hill, Edward Tufte, and Jack Webster
Graphical tips (2)

- Use space efficiently
- Draw axes to fit data
- Use 5 to 10 tick marks on the axes (fewer can sometimes be ok)
- Make sure the data points are big enough

Graphical tips (3)

- Label axes clearly and legibly (units!)
- Use labels on the graph itself, rather than putting a legend in the margin

Graphical tips (4)

- Edward Tufte’s tips
  - Maximize the data-ink ratio
  - Erase non-data ink
  - Erase redundant data ink
  - Revise and edit

A test!

Readers see red over low-impact graphics

Sir — Nature authors are increasingly using colour within figures. May I remind them and the graphics editors that a significant proportion of their readers, including me, cannot distinguish red from green? Issue number 7126, for example, was not atypical in having six figures that used these two colours. The impact of their figures — and just possibly the impact factor of the journal — will be improved if all the audience can see all the data.

Chris Miall
School of Psychology, University of Birmingham, Birmingham B15 2TT, UK

445:147

Now, some stats...

- Descriptive statistics
- Comparing groups
- Examining relationships
Descriptive statistics
- Mean or average: central tendency
- Standard deviation: dispersion about a mean

Comparing groups
- Comparing the means of two groups: t-test
- Comparing the means of more than two groups: analysis of variance (ANOVA)
- Recall, complete analysis = graphs AND stats...
  - Comparing groups: bar charts (= column charts in Excel)

Examining relationships
- 'Cause-effect' type of relationship between two variables: regression
  - Also: prediction, explaining variation
  - Recall, complete analysis = graphs AND stats...
  - Relationships: x-y scatter plots

Significance
- In science, you only use the word 'significant' if you have run statistical tests to see if your results are different than what would be expected from chance
- So how do you determine significance from statistical output?
  - Usually through the use of 1 or 2 pieces of statistical information:
    - Calculated and critical test statistics
    - P-values

Interpreting significance: test statistics
- A result is significant if:
  - The calculated test statistic is greater than the critical (or theoretical) test statistic
- Example:
  - \( T_{\text{calc}} = 5.34, T_{\text{crit}} = 1.99; \) Because \( T_{\text{calc}} > T_{\text{crit}} \) there is a significant difference between the two means
- The values for the critical test statistics depend on:
  - the test being run,
  - the sample size,
  - and the probability level being used (usually \( \alpha = 0.05 \))

Interpreting significance: P-values
- A result is significant if:
  - The computer-calculated P-value is less than a pre-defined \( \alpha \)-value (usually 0.05)
- Example:
  - P-value = 0.01; Because the P-value < 0.05, then there is a significant relationship between the two sets of data
  - P-values, or probability values, describe the probability that observed or more extreme differences would be found if the null hypothesis is true
- The pre-defined \( \alpha \)-value can vary depending on the needs of the researcher, but scientists often settle on 0.05 (i.e., reaching the wrong conclusion 5% of the time assuming the null hypothesis is true) as being an acceptable chance of error
Easier to observe significance if…

- You have a well-designed study
- You have many replicates
- You have little variability within the treatments or factors in your study
- You have large differences between the variables of interest (i.e., more variability between the treatments or factors)

Time to collect some data

- On the board, record your:
  - Height (in cm)
  - Arm span (in cm) from tip of left to right hand
  - Academic class (e.g., junior, senior)
  - Gender

Look at the data—Graph it

- In groups, everyone should:
  - Copy the data from the board to an Excel spreadsheet
  - Calculate mean height for each academic class & gender
    - Excel formula: =average(x1:x2)
  - Use a Column chart to compare the mean height of one academic class vs. another (usually known as a bar chart)
    - Hint: first highlight data and labels
      - Next, under Insert menu choose 2D column chart
  - Use a Scatter chart to plot all heights vs. all arm spans for entire class (all data, NOT means)
    - Choose Scatter with only markers
  - Make graphs look as good as you can using Layout

Today’s report

- Each group submits one column chart and one scatter plot
- Copy and paste them into Word
- Briefly (1-2 sentences) summarize the results and identify which statistical test you would use to examine the data shown in each graph