BMR 617

Summary

Summary

- This course looked at two aspects of statistics:
- Descriptive statistics
  - How to present data
- Inferential statistics
  - Inferring information about a population from data from a sample
  - Hypothesis testing – automated decision making about the validity of a hypothesis

Using statistics in research

- Statistical techniques should be used to enhance your research program
- Use statistics to understand your data
  - What message is present in your data?
  - What message is not present?
- Use statistics to provide an objective mechanism for testing a hypothesis
  - It is easy to become attached to a particular scientific hypothesis, and consequently difficult to evaluate its validity objectively

Two common misuses of statistics

- Scientists often see statistics as a "hoop through which they have to jump"
  - They know their hypothesis is true, they just need to "get a statistically significant result" to publish
- Conversely, scientists sometime over-rely on statistics, particularly hypothesis testing, to justify a claim
  - "I got a statistically significant p-value, so this hypothesis must be true/interesting/important"
- Remember, a p-value is only a loose measure of the certainty of a result

Scientific integrity

- Scientific integrity is essential to the scientific process
- Peer-review helps ensure scientific integrity, but there is a large reliance on the honor system
- Scientific integrity can be thought of on two levels:
  - Honesty
    - Are the data you present real, correct, and unmanipulated?
  - Truthfulness
    - Do you truly, and objectively, believe the hypothesis you claim to support?
    - Is it consistent with data from other labs? If not, what are the differences between what they have done and what you have done, and why are you "correct"?

Barriers to scientific integrity

- The scientific and academic communities inadvertently place many barriers in the way of scientific integrity
- Requirements to publish
  - Typically you need "statistically significant" results to publish
  - And you need publications to further your career
    - Graduate from graduate school, progress from postdoc, receive grant funding, achieve promotion and tenure, etc. etc.
  - These pressures result in a need to find ways to get small p-values
  - Known as "publication bias"
- Requirements for novelty
  - Grant funding typically requires novel results
  - Paradigm-shifting research is the most highly valued research

Statistics and scientific integrity

- Properly used, statistics can be an enhancement to scientific integrity
- Honesty:
  - Are the data correct and the data actually generated by your experiment?
  - Are the correct hypothesis tests used and the results presented accurately?
  - Are the data presented in a manner that truly reflects the nature of the data?
- Truthfulness:
  - Avoid "p-value hacking"
    - Constantly repeating experiments until the desired result is obtained
    - Trying different statistical tests until one gives the desired result

When you don't get the results you want

- If your experiment doesn't produce the results you want, troubleshoot
  - don't blindly try to fix the experiment without understanding why
- If your p-value is high, why is it high?
  - The difference between values in the different groups is not big enough?
    - Verify the treatments/conditions are as you expect
  - There is too much variability
    - Why? Is a single outlier causing the variability? Or is there generally a lot of "noise" in one or more experimental groups.
    - In either case, investigate why.
  - Always consider the possibility your hypothesis may not be true!

Be mindful

- Use statistics to explore and understand your data and what it is telling you
- Set aside your own personal needs for results, and read your data objectively
- The best scientists:
  - Understand their results in relation to other researchers
  - Interpret their data in a thoughtful and mindful way
  - Admit when their hypotheses are wrong