Common Statistical Mistakes

March 23, 2016
Understanding Inferential Statistics

- Inferential statistics tell what will happen in an entire population based on what did happen in a sample from that population.
- Inferential statistics require a representative and random sample.
- The appropriate analytical technique will depend on the type of inference you want to draw.
Choosing the “correct” test

Analysis selection depends on:
• The type of question you have, and
• The type of data you have.
Nonparametric v. Parametric Analysis

• If your data are non-normal, or violate other assumptions of the parametric analysis you are performing, nonparametric tests are an option to consider.
• You can also transform your data (e.g., take logs or square roots) to meet assumptions.
• Parametric tests still have assumptions!!
Nonparametric vs. Parametric Analysis

- The mean and median are not equal when data are skewed.
- They are equal if data are symmetric OR are normally distributed.
- Skewed data also has larger standard deviations.
- Confidence intervals will not have good coverage in skewed data.
Testing for Normality: Should you?

Tests for normality have low power when sample sizes are small.
  • Most small samples will pass a normality test.
  • Large samples will have a difficult time qualifying as normal.
Another common research myth is that parametric statistical methods are always more powerful than non-parametric methods. According to this myth, researchers increase their chances of finding differences or relationships if they use statistical methods that require normality to analyze nonnormal data. However, parametric methods are more powerful than nonparametric methods only when data fit their assumptions. When their assumptions are not met, parametric methods often miss differences or relationships that would be detected by nonparametric methods.
P-values and Statistical Inference

The p-value does not tell you if the result was due to chance. It tells you whether the results are consistent with being due to chance. That is not the same thing at all.

The American Statistical Association just released a statement on p-values.
ASA’s Statement on P-Values

1. P-values can indicate how incompatible the data are with a specified statistical model.

2. P-values do not measure the probability that the studied hypothesis is true, or the probability that the data were produced by random chance alone.

3. Scientific conclusions and business or policy decisions should not be based only on whether a p-value passes a specific threshold.

4. Proper inference requires full reporting and transparency.
ASA’s Statement on P-Values, continued

5. A p-value, or statistical significance, does not measure the size of an effect or the importance of a result.

6. By itself a p-value does not provide a good measure of evidence regarding a model or hypothesis.

In some sense the p-value offers a first defense line against being fooled by randomness, separating signal from noise...The p-value is a very valuable tool, but it should be complemented – not replaced – by confidence intervals and effect size estimators (as is possible in the specific setting).

--Tal Galili (R-bloggers)
Power

• A power analysis is a way to find either:
  • The effect size you’ll be able to detect given a set sample size, OR
  • The sample size you’ll need to detect a specific effect size.

• Doing a power analysis makes you think critically about your proposed study.

  *It’s better than a post-mortem analysis!*
# Power and Errors: Consequences

<table>
<thead>
<tr>
<th>Conclusion of significance test</th>
<th>Reality</th>
<th>Reality</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Null hypothesis is true</td>
<td>Null hypothesis is false</td>
</tr>
<tr>
<td>Reject null hypothesis</td>
<td>Type I Error</td>
<td>Correct conclusion</td>
</tr>
<tr>
<td></td>
<td>We waste money developing useless drug.</td>
<td>(probability = power)</td>
</tr>
<tr>
<td>Do not reject null hypothesis</td>
<td>Correct conclusion</td>
<td>Type II Error</td>
</tr>
<tr>
<td></td>
<td>(probability = 1 – significance level)</td>
<td>(probability = 1 – power)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>We miss opportunity for profit.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>People don’t get access to an effective treatment.</td>
</tr>
</tbody>
</table>
What impacts power?

1. The design of the study and the type of measurements
   • Paired data? Two groups? Three groups? Continuous data? Nominal or ordinal data?

2. The variability of the data

3. The significance level of the test

4. The effect size of interest
   • Ask yourself “What am I hoping to find?”
   • And “Would it be important if I found half that difference?”

5. The sample size
   • What can you afford to do? (Time or money)
Small Sample Sizes

• Small sample sizes don’t make your statistical analysis invalid.

• They do make it harder to draw inferences about the population of interest.

• You have to be careful to not overstate the strength of your evidence given the size of your sample.
Randomization and Treatment Allocation

• Randomization is VERY important!

• Randomization in treatment allocation ensures:
  • That no systematic differences exist between groups.
  • That estimated treatment effects are not biased by confounding factors.

• You can also stratify (by sex or breed, for example) to control for confounding.

• Randomization and confounding are two reasons that historical controls can be problematic.
Nonindependent Data

• Treating nonindependent data as independent is a huge problem.
  • It will artificially increase your sample size.
  • It will artificially decrease your variances.

• If you have multiple measurements taken on the same subject at one time (e.g., both eyes) or over time (e.g., blood pressure across time), your data is NOT independent.
Pseudoreplication

Pseudoreplication is when your measurements are replicated, but are not independent.

• If you measure blood pressure on 100 subjects 10 times each, you do NOT have 1,000 independent data points!

http://www.statisticsdonewrong.com/pseudoreplication.html
What is the problem?

• Treating nonindependent data as independent:
  • will artificially increase your sample size.
  • will artificially decrease your variances.

What is the effect of an artificially decreased variance estimate?
Nonindependent Data

There are many ways to handle nonindependent data.

• Create a random effect for ID, allowing differences between and within subjects
• Pair differences, then test the independent differences
• Average multiple measurements, then analyze the averages
• Use longitudinal methods to track changes over time (best for lots of measurements over time)
Censored Data

• If you have data lost to follow-up (or a different cause), that data is **censored**.

• Censored data is NOT the same as missing data!
  • Survival analysis is the way to deal with this issue.
    • Proportional hazards regression
    • Kaplan-Meier curves
Outliers!

- Outliers are unusually extreme observations.
  - Sometimes they are a mistake!
  - They definitely warrant further examination
  - Please don’t remove outliers for no reason OR without mentioning that you are doing so.

https://psud63.wordpress.com/2011/10/14/is-it-justified-to-remove-outliers/
Final Thoughts

• There are many ways to analyze your data.
  • Think about what you want to do
  • Pick one of the analyses that will work.
  • Don’t get discouraged!

• Statistics and results should make sense!
  • If something doesn’t make sense, that’s not a good sign.
Coming soon!

• Based off this group’s feedback AND conversations with Duncan, Colleen and I are going to put together articles about best practices for statistical analysis.

• We’ll organize them in a way that makes sense so they can be used:
  • To help you determine which type of analysis you should do, OR
  • As a resource to push back on a review.
Selected References


