Ten common statistical mistakes to watch out for when writing or reviewing a manuscript

by Tamar R Makin and Jean-Jacques Orban de Xivry | Presented by Melanie Desroches

About the Paper and it's Authors

About the Authors

Tamar R Makin

- Neuroscientist studying brain plasticity (how the brain reorganizes itself)
- She is currently a professor at the University of Cambridge teaching cognitive neuroscience

Jean-Jacques Orban De Xivry

- Also a Neuroscientist studying movement control
- He is a professor in the Movement Control & Neuroplasticity research group of KU Leuven

The Inspiration for the Paper

- While the list was inspired by issues seen in neuroscience journals, the authors argue that this can be applied to any field of science
- There are 10 mistakes and for each, the authors provide an explanation of the problem, how reviewers can identify the problem, and how researchers can solve the problem
- There tends to be some overlap in the mistakes so it important to be aware of all these mistakes since one can impact the other

1

Ten Most Common Statistical Mistakes

Absence of an adequate control condition/group

The Problem

- In cases where an effect is being studied over time, it is important to have a control group to compare the results to in order to have a baseline
- Sometimes, there either isn't a control group or there is some bias introduced into the control group that makes the control group inadequate
- The control and experimental group should be as similar as possible, the only difference being the variable that is being manipulated

Absence of an adequate control condition/group

Reviewers should look out for:

- Conclusions being drawn without being compared to the control group
- Not accounting for factors that may impact the control group

Solutions for researchers:

- If you cannot separate the effects of time, then the results should be considered tentative

Interpreting comparisons between two effects without directly comparing them

- Researchers tend to determine the effect of some treatment/intervention by comparing the results of the experimental group to the control group
- The problem is when researchers conclude results based on the fact that the experimental group has significant results while the control group does not
- The two groups need to be directly compared to each other in one statistical test

Interpreting comparisons between two effects without directly comparing them

Reviewers should look out for:

- Conclusions are drawn without statistically comparing the two effects

Solutions For Researchers:

- Compare groups directly (for example, Correlations between two groups can use Monte Carlo simulation or When doing group comparisons, use ANOVA)

Inflating the units of analysis

The Problem

- The experimental units used in analysis is important to be clearly defined because it determines the degrees of freedom used
- A common mistake is when doing group analysis, the experimental unit is determined to be the number of observations made within each subject (this is incorrect), rather than the number of subjects (this is the correct unit)
- This can increase the degrees of freedom and therefore increase the statistical power. Thus, it is easier to observe falsely significant results

Inflating the units of analysis

Reviewers should look out for:

- Verify that degrees of freedom reflect the number of independent subjects, not measurements.

Solutions For Researchers:

- Consider using a mixed-effects linear model
- Calculate correlations for each observation separately
- Average values across observations
- Calculate correlations separately for pre and post, then average the resulting R values

Spurious correlations

The Problem

- In parametric correlation, Pearson's correlation coefficient comes with a set of assumptions: normality, linearity, pairs of values, continuity, and no outliers
- Spurious correlation is when a relationship is detected that is not causal. This can often occur when there is a violation of one of the assumptions, the most common being the presence of outliers
- Outliers that did not occur by mistake should not simply be discarded. Instead, try another statistical approach

Spurious correlations

Reviewers should look out for:

- Correlations being reported without a scatter plot
- Outliers that have been removed without justification

Solutions For Researchers:

- Try using alternative correlation metrics such as bootstrapping (resampling with replacement), data winsorizing (set outliers to a certain percentile), or skipped correlations (calculate correlations after removing outliers).
- If using parametric statistics, always consider assumption violations.

Use of small samples

The Problem

- With small sample sizes, it's easier to detect only large effects, which creates a lot of uncertainty around the actual strength of the relationship between variables
- With small samples, false positives can appear large, leading to what's called the "significance fallacy": the mistaken belief that a large effect in a small sample is definitely real.
- In reality, a larger correlation in small samples isn't due to a stronger relationship; it's just an overestimate
- Small sample sizes can also make it difficult to test for normality

Use of small samples

Reviewers should look out for:

- Strong claims made using a small sample size
- It is the reviewers discretion on whether or not the sample size is too small

Solutions For Researchers:

- Provide evidence that the sample size is reasonable through performing a statistical power analysis or through replicating the study
- If the sample size is limited, use Crawford t-test or try to isolate the effect of the primary factor being studied

Circular analysis

- Circular analysis happens when you use the same data to define or select something and then test or measure that same thing. This creates a "circle" because you're using the data both to set up the analysis and to draw conclusions, which leads to biased or misleading results.
- A hypothesis is tested on the same data set that led to the generation of the hypothesis in the first place. This creates a feedback loop where hypotheses seem supported simply because they were derived from the same data.

Circular analysis

Reviewers should look out for:

- Data that was selected because it showed the effect of interest
- An effect size (such as a correlation or a difference between groups) is much larger than what is theoretically or practically plausible

Solutions For researchers:

- Before looking at the data, decide the analysis criteria to avoid introducing bias
- Use a different dataset to test your predictions

Flexibility of analysis: p-hacking

The Problem

- P-hacking is when researchers manipulate the procedure in order to get a significant result
- This can come from removing data points, trying out multiple tests to get a significant result, removing variables, etc.
- The authors argue that if the experiment is well designed, there should be some tolerance towards borderline significant results

Flexibility of analysis: p-hacking

Reviewers should look out for:

- If a large bulk of variables were taken but only a few statistically significant ones were reported
- If the analysis plan is predetermined before testing and if that plan was not followed

Solutions For researchers:

- Avoid changing hypothesis before testing
- Be transparent
- Use pre-registration of the experiment design and analysis

Failing to correct for multiple comparisons

The Problem

- When testing a set of hypotheses, as the number of hypotheses tested increases, so do the chances of seeing a false positive.
- The more variables tested, the greater the chances of seeing a significant correlation with the dependent variable simply by chance
- Without adjusting for multiple comparisons, researchers risk publishing findings that reflect statistical noise rather than real effects.

Multiple comparisons

Reviewers should look out for:

- Many variables that are included do not correlate with the dependent variable.
- Interpretation of results from exploratory analyses without applying corrections for multiple comparisons.

Solutions For researchers:

- Researchers should be transparent about the exploratory nature of their analysis and disclose all measured variables.
- Different methods exist for this correction, and it's crucial to use well-accepted ones to avoid inaccurate findings

Over-interpreting non-significant results

The Problem

- Researchers may try to use the fact that a result is deemed non-significant means that there is no effect
- Non-significant results can be due to small effects, an underpowered study (such as from a small sample size), etc.
- Just because you fail to reject the null hypothesis does not mean that the null hypothesis is true
- Absence of evidence is not evidence of absence

Over-interpreting non-significant results

Reviewers should look out for:

- Describing a non-significant p-value as indicating that an effect was not present

Solutions For Researchers:

- Report the effect size along with the p-value. This allows readers and reviewers to understand not only if there is an effect, but to what extent there is an effect.
- Always be mindful of how you interpret non-significant results.

Correlation and causation

The Problem

- Researchers often want to examine the relationship between two variables using correlation
- When the correlation is calculated between variables, some may assume this means that one directly causes the other
- Correlation does mean not causation. There may be other variables at play or it may just be a coincidence

Correlation and causation

Reviewers should look out for:

- Using correlation as evidence of causation

Solutions For Researchers:

- Causation should only be determined from a direct manipulation of one variable causing a change in the other - and even then there should be caution
- Always consider other confounding variables that may impact two variables correlation

Conclusion

Summary

This presentation has covered:

- The ten most common statistical mistakes made in manuscripts
- How reviewers can spot each mistake
- How researchers can fix each mistake

THANK YOU!!!