Simple rules for statistical analyses

Mark D. Scheuerell[1,*] Sarah J. Converse[1]

[1] U.S. Geological Survey Washington Cooperative Fish and Wildlife Research Unit, School of Aquatic and Fishery Sciences, University of Washington, Seattle, WA

[*] Correspondence: Mark D. Scheuerell <scheuerl@uw.edu>

Notes for Sarah & Mark

Journal guidelines

Here is the information from the journal's author guidelines:

> Essays should include ideas, concepts, hypotheses, or opinions to stimulate discussion, debate or research. Essays can be written as opinion pieces that are well-reasoned with well-supported arguments and ideas, but where extensive citations are not needed or available. Essays should be written in clear and general language that are understandable to a broad audience.
>
> - 2,500 words maximum (includes main-body text only; excluding significance statement, acknowledgements, and references) and 1-3 visuals (figures, tables, or boxes).
>
> - Essays should include: Significance statement, text (sub-headings are optional), references. Note that essays do not include an abstract.

Essay example

As an example, here is the first essay of this type for the journal.

Generating references

There are a couple of ways to add references to the `references.bib` file. The first is to copy/paste them from any reference software that will display them in `.bib` format (e.g., EndNote, Zotero). For example,

```
@Article{Smith_2021,
  year = {2021},
  publisher = {The Ecological Society of America},
  volume = {123},
  number = {2},
  pages = {1-10},
  author = {Jane Smith and Joe Johnson},
  title = {Probably the best paper ever written},
  journal = {Ecology},
}
```

The second is to use the `{knitcitations}` package, which will load when the `.Rmd` file is knit. Here is an example of an inline `R` command that will

1) create an entry in `references.bib` for Hotaling (2020) linked to above;
2) display a properly formatted citation in the text; and
3) add a formatted entry to the References section.

```
`r citet("10.1002/lol2.10165")`
```

(Hotaling 2020)

You can see other examples here.

Scientific significance statement

[*This section and heading are required.*]

For example, here's what Scott wrote in his essay:

> One of the most common editorial refrains, regardless of discipline, is "this needs to be tighter." It typically means too many words and ideas are jumbled together and the underlying point is obscure. The writing is not concise. But, improving conciseness is difficult because the problem is caused by a host of factors that are easily overlooked, especially by early career researchers. Here, I describe what it means to write concisely and outline 10 rules, with examples, to help scientists tighten their prose.

Introduction

[*This section was 3 paragraphs in Scott's essay (and it lacked an actual heading).*]

Rules for statistical analysis

[*The above heading is just a placeholder.*]

Here are some possible topics, ordered around the notions of before, during, and after a study/analysis, which we had discussed briefly.

Before beginning

- Identify your primary motivation and scope for inference
  - Is it retrospective understanding? Extrapolation? Forecasting?
- Make sure you understand the assumptions and limitations for any approach
  - eg, could you explain this to both your grandmother and the editor?
- Do not mix frequentist and Bayesian paradigms
  - We're not making a judgment call here, but merely pointing out that there are important philosophical differences that should not be confused/conflated.

- If using a frequentist framework, consider acceptable levels of Type-I and Type-II errors
  - Choose an α-value ahead of time
  - avoid the use of asterisks (eg, $*/**/***$) to indicate significance at varying levels
- When the data are spatial or temporal, account for autocorrelation
  - If not, model assumptions not met; p-values are biased

During analysis

- Make it reproducible
  - provide the data and code necessary to replicate your analyses
  - if there are any proprietary data, provide a minimal work example
- Don't do statistics on statistics
  - reframe your analysis to better match the objective
- Model diagnostics are an important part of any analysis
  - Every model or test involves assumptions; are they met?

Reporting (after)

- General notions of p-values from Wasserstein (2019)
- Correlation is not causation
  - Unfortunately, this bears repeating...
- Effect size is more important than a p-value
  - Highly replicated experiments have large power to detect differences/effect, but are they meaningful?
- Being the best is not always good
  - Model selection exercises, particularly via in-sample methods (eg, AIC), will identify a best model, but the question is how good is it?
  - report other measures like R^2

  - use other methods such as out-of-sample (eg, leave-k-out cross validation) based on metrics that address both bias and precision (eg, RMSE)

- Report all of the information from statistical tests–don't make reviewers and editors guess

  - eg, both df needed for F-test

Conclusion

[*The above heading is just a placeholder and is not actually necessary.*]

Conflict of interest

The authors have no conflicts of interest.

Acknowledgments

We wish to thank...

References

Hotaling, S. 2020. Simple rules for concise scientific writing. Limnology and Oceanography Letters **5**: 379–383. doi:10.1002/lol2.10165

Wasserstein, R. L., A. L. Schirm, and N. A. Lazar. 2019. Moving to a world beyond "p $<$ 0.05." The American Statistician **73**: 1–19. doi:10.1080/00031305.2019.1583913