University of Nebraska - Lincoln
DigitalCommons@University of Nebraska - Lincoln

The Prairie Naturalist | Great Plains Natural Science Society

6-2017

The Pitfalls of Data Analysis

Christopher N. Jacques

Follow this and additional works at: https://digitalcommons.unl.edu/tpn

 Part of the Biodiversity Commons, Botany Commons, Ecology and Evolutionary Biology Commons, Natural Resources and Conservation Commons, Systems Biology Commons, and the Weed Science Commons

This Article is brought to you for free and open access by the Great Plains Natural Science Society at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in The Prairie Naturalist by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

The Pitfalls of Data Analysis

Greetings GPNSS members! I hope summer finds you well and enjoying the Great Plains in some way, whether that be starting a new field season, a long overdue break from the office, or planning your next family vacation. For me, transitioning away from the demands of the Spring 2017 semester to data analyses and preparation of my own manuscripts are particularly exciting (well, as exciting as "office life" can be!). Over the past month, I have initiated a number of grandiose summer objectives, most of which include analysis or in some instances, re-analysis of data to address concerns raised during peer-review. Though the past month has been arduous and much time spend ascending the Program R learning curve, it has provided me with the subject of this editorial: the pitfalls of data analysis.

I have been Editor-in-Chief for nearly 9 years now, over which time I have processed hundreds of manuscripts and considered hundreds of additional reviews. Over the past decade, I have noticed an increasing emphasis on data analyses at the expense of a greater understanding of the biological system under study. I raise this issue not to de-emphasize the statistical advances within the various disciplines of natural resource management….in fact, quite the contrary! I appreciate and greatly admire the role we play in developing novel and rigorous analytical approaches. Nevertheless, I can't help but wonder whether our role as resource managers has somehow been compromised. While I will likely never know if this is the case, it remains a question that continues to fester in my mind.

A hallmark to conducting science requires a knowledge that all parts of a study and resulting manuscript are important, including (but not limited to) formulating appropriate questions, development of rigorous study design, and measuring suitable variables (Block 2012). Whether a researcher frames scientific questions as clear and concise study objectives, hypotheses, or quantitative models, the cornerstone of science demands that they be relevant and addressable (Block 2012). That is, they should be based on knowledge of the biological system under study, focused, well-informed, and intended to move science forward by addressing key information gaps. A well-developed body of literature has been written about study design, so I will not bore you with elaborating on the minutiae of a well-designed study here. In general, studies should be designed to capture spatial and temporal variation in a system, thereby enabling strong inference at appropriate scales of study to address study objectives (Block 2012). Additionally, researchers should consider aspects of randomization, replication, and adequate sample sizes to ensure unbiased and precise parameter estimates and to enable broad inference (Block 2012). I appreciate and remain mindful of logistical and financial constraints imposed on study designs, and the need to work within them. Nevertheless, development and implementation of study designs must be adequate to address study objectives (Block 2012).

In any study exists opportunities to quantify countless variables, though few of us have sufficient resources to measure everything. Consequently, we are faced with the fact that many of us are operating on "shoestring" budgets rather than "Cadillac budgets" so must pare down the list of variables to those that are most biologically relevant. Measuring everything and hoping to identify a "significant" variable or two is comparable to model selection evaluating all possible combinations of variables without developing a priori model sets (Block 2012). Associated with considerations of designing and conducting a study is data analysis, which brings me to my main concern. As I alluded to previously, many statistical tools and options are available for conducting analyses, which are perpetually evolving. During my days as a graduate student in the late 1990s and early- to mid-2000s, I have seen a paradigm shift from frequentist (e.g., univariate and multivariate statistics) approaches to Bayesian (e.g., state-space and integrated population modelling) statistics to model selection and parameter estimation (e.g., information theoretic approaches, AIC). I have met and interacted with practitioners of each new approach, all of which regard "their" approach as the best. Adding to the perils of analytical approaches is the sentiment that if authors do not embrace and employ the latest analysis, their manuscript will not be published (Block 2012). Statistical methods and results sections appear to be increasing in size (and complexity) and tabular presentations of simple descriptive statistics have now been replaced with complex models and associated diagnostic metrics. Synthetic discussions of results often are hyper-focused on statistical limitations of associated analyses rather than on the relevance of results to the biology and management of the system(s) studies. Please do not misinterpret my message here, but enough already!

I noted above that I understand and appreciate the role that natural resource professionals serve in developing novel and more rigorous statistical techniques for analyzing data. All things considered, any one dataset can be analyzed using multiple statistical approaches (Block 2012). As a practitioner of multiple statistical approaches for analyzing datasets, I'm not advocating the use of significant-testing over model-selection, or frequentist over Bayesian statistics. All offer viable alternatives for analyzing a dataset. Instead, I would rather see well-designed studies that use appropriate statistics and convey clear management implications rather than poorly designed studies with complex analyses and confounded or equivocal results (Block 2012). In short, I would rather review well-designed studies that use statistics as tool rather than as the end all. I hope that as natural resource professionals, we can avoid the pitfalls of data analyses at the expense

of sacrificing our understanding of the biology of the species or system being studies and providing viable management options (Block 2012).

As with past issues of *TPN*, I think we have a well-rounded issue with papers representing several taxa, and addressing a number of management and conservation issues. Paul Jones and his colleagues provide an insightful evaluation of pronghorn movement rates. An interesting population study evaluates response of pheasants to lead ingestion. Other studies examine mountain lion capture techniques, Canada goose nest success, predator-prey interactions, birth synchrony in ungulates, trumpeter swans, andn competition among annual and perennial grasses. This issue also features a book review, which was overseen by our Book Review Editor, Larry Igl.

In closing, if you have any questions, comments, or helpful suggestions for improving *TPN*, please feel free to contact me. After all, this is your journal, and I very much appreciate your thoughts about it. Until next time, enjoy your summer everyone!

—Christopher N. Jacques
Editor-in-Chief

LITERATURE CITED

Block, B. 2012. Editor's Message—Analysis Paralysis. Journal of Wildlife Management 76(5):875–876; DOI:10.1002/jwmg.408.