

Approaching analyses for the infrequent useR

Kim Cressman

Grand Bay NERR

kimberly.cressman@msstate.edu

May 19-20, 2022

Analysis Planning

1. What is your research question?

2. What are your variables?

- Dependent (response)
- Independent (explanatory/predictor)
- Anything else that could impact the relationship between explanatory and response variables

3. For each variable:

- What type of data? E.g. continuous; categorical; counts; proportions; percentages
- What general range of values do you anticipate?
- Do you expect many outliers?
- Do you expect many 0s (or “below detection”s)?
- Can you (reasonably) expect the values to be normally distributed?

4. What is your sampling design, and why?

- How many stations, and where?
 - Draw a map!
- What are you sampling at each station (and why)?
- How frequently are you sampling for each variable?
- What are your replicates? (beware pseudoreplication)
- How are your data points related to each other?
 - Time
 - Space

5. How do you want to summarize your data and patterns?

- Summary stats: mean, standard deviation, range? By group?
 - Make an empty table that you can put your values into when you get them
- Figures: sketch out, conceptually, the ones you plan to make

6. What do you want to test statistically?

- Go back to your research questions
 - Is [thing] different between group A and group B?
 - What are these groups?
 - How many samples per group?
 - Do you expect anything other than [thing being tested] to differ between groups?
 - Is there a relationship between [x and y] variables?
 - Are all of your data points independent, or might they be related (in space and/or time)?
 - Do you want to include more than one potential explanatory variable? (how is Y predicted by X and Z?)
 - Are you most interested in describing the direction and strength of a relationship? Or in making predictions in new situations?
 - Are there other variables that could impact the relationship you want to examine?

6. What do you want to test statistically? (cont'd)

- Go back to your research questions
 - Community analyses
 - What do you mean by “community”? How will you represent that in tabular form?
 - What are your variables: counts? Relative abundance? Presence/absence?
 - Think about transformations
 - Think about best similarity matrix (Bray-Curtis is typical in ecology)
 - Will you also calculate richness and/or diversity metrics for each sample?
- What statistical analyses are typically used for questions like this in your field?
 - And what decisions are common along the way?

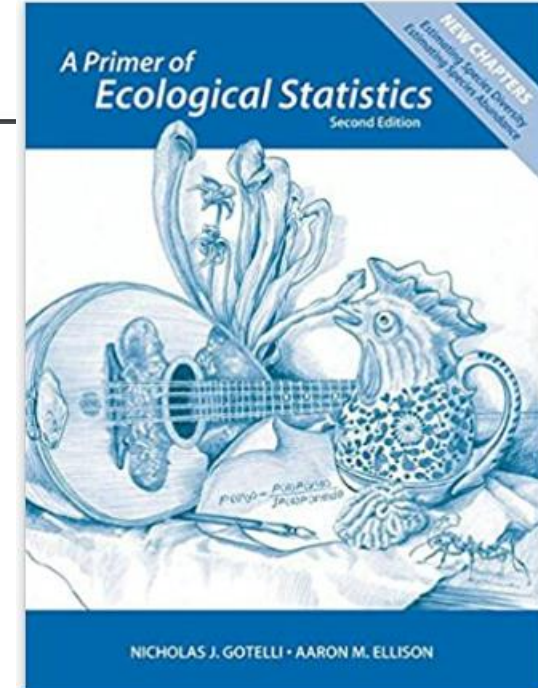
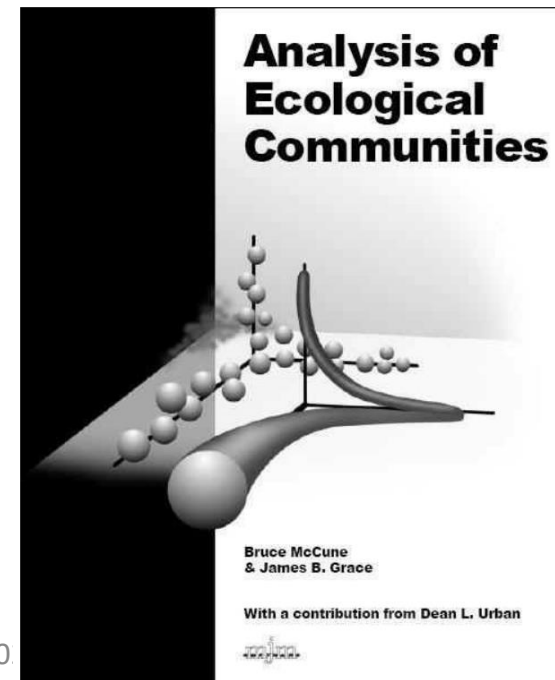
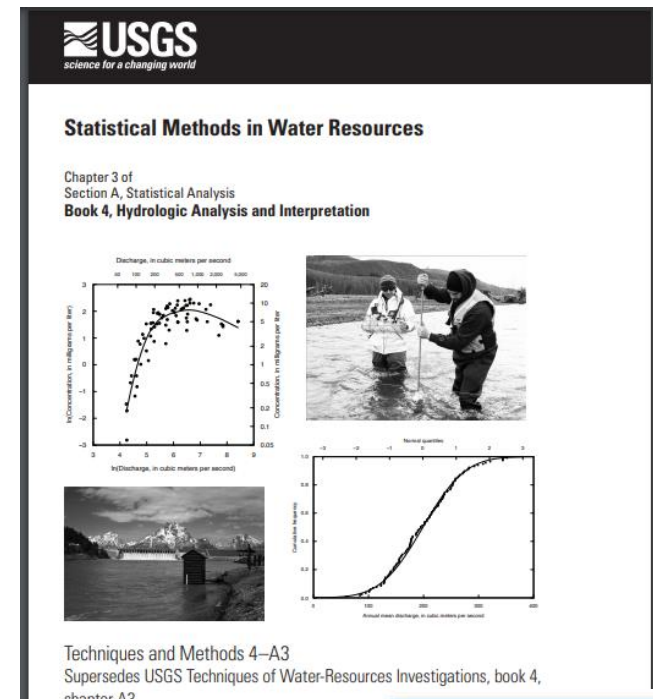
7. What do you want to present from the statistical results?

- **More than just a p-value**
- Estimate of effect
- 95% confidence interval
- Sample size
- Test statistic (F, t, chi-squared, etc.)
- Slope and R^2 (regression)

Choosing a test

Get guidance

- What's common in your field?
 - Scientific publications
 - Books
 - Your advisor or mentor
- Many universities have a statistical consulting center



Choosing a function

Deciding which function to use

- Try to see what *other* people in your situation are using:
 - CRAN task views: <https://cran.r-project.org/web/views/>
 - online class notes
 - blog posts
 - social media: #rstats on twitter; “Ecology in R” group on Facebook
 - scientific publications
- Look for packages and functions with good documentation
 - blog posts
 - vignettes
 - github READMEs
 - help files
- Use what *you* can understand (and test it first)

Some good, solid packages

- stats (base): all the basics – t.test, lm (regression and anova)
- vegan: community analyses
- EnvStats: lots of WQ stats
- lme4: mixed models
- mgcv: GAMs
- unmarked: occupancy analysis

Running it

Things to consider as you code

- Figure out what assumptions that test and function are making
 - data format: columns/rows; number vs. character; etc.
 - normality? (not always needed)
 - how to handle 0s or NAs
- Decide whether your data meets those assumptions
- Look at the function's inputs: what do you feed into it?
- Look at the function's default values; do you want to change them?
- Document all the choices you make along the way, and why!