Final paper

Project description

* The final project will be to discuss and perform a group of statistical analyses. Both the scientific and statistical background should be clearly presented, and future statistical steps you plan to take.
* These should be written and presented like the results section of a scientific paper, with a **full ABSTRACT**, very brief introduction with a background/hypotheses, methods, and discussion section sufficient for understanding the results.
  + It is acceptable to use the previously written paragraph about your data as background.
  + You **must** include elements of what we have learned in class directly in your paper, incorporating statistical techniques and plotting of figures.
  + Bullet points in the introduction, methods, and discussion section are acceptable. You should think of this an outline of a thesis paper.
  + Usually, we flesh out the results and the figures, then add the rest of the stuff. I write my abstracts as a mini-outline that helps me figure out what should go in each section.

What to turn in:

1) Your full "paper" with figures, tables, captions embedded in the text, and any appendices that you might put in the supplemental online material.

2) All of your code, completely reproducible, committed and pushed to Github.

3) If more than one file is used to make the code, LANGWIG\_FINALPROJECT\_README file that tells me how your code goes together.

Paper template

* Title, Name and authors
* Abstract (full, ~150 – 300 words, think about target journals!)
* Introduction (background, hypotheses; bulleted is okay)
* Methods (scientific and statistical; bulleted is okay)
* Results (fully written like a scientific paper). Please re-visit tips for writing results sections below.
  + Please put figures in this section of paper (minimum 2) with their captions directly below them. Figures for your final paper should be polished and publication-worthy. I expect that you will have explored color, shapes,
  + Include statistical results as 1) either in the text, 2) in main text tables, or 3) as supplemental tables as you would in a real paper.
* Discussion (what do your results mean, future statistical analyses to be conducted, what this tells us about the world; bulleted is okay)

A few notes:

* It is perfectly fine to use a paper you are already writing, but nothing that has already been sent out for peer-review, please.

General suggestions for writing statistical results

1. Always include estimates of coefficients with some kind of standard error or confidence interval
2. Include the exact p-values (e.g. not just p<0.05) unless p<0.0001.
3. It is often helpful to include estimates of means and standard errors for descriptive statistics in describing results
4. You need to include the intercept, not just the slope some place in your manuscript, either in the text, the figure caption, or a supplemental table. You don’t need to report the p-value of the intercept.
5. If you have complicated statistics, I recommend moving the statistical output to tables or figure captions. You can report less complicated statistics in the text.
6. When reporting results, give the direction of an effect, don't just say it is significant.
7. It can be effective to remind a reader of the type of analysis you did, rather than just mentioning it in the Methods.  
     
     
   Examples:  
     
   We found that bee activity increased with elevation (p < 2e-16) with high elevation sites having the highest activity (β = 1.69 ± 0.039), followed by mid elevations (β = 1.10 ± 0.017), and low elevations (β = 0.71± 0.017). Bee activity also increased with date (p = 5.37e-10, β = 0.0012 ± 0.0002), however, the effect was nominal.  
     
   Empirical estimates showed that increasing pathogen challenge dose decreased infection vaccine efficacy (treatment and dose interaction: P < 0.0001), although results varied among studies (random effect of disease: variance = 3.43). For some diseases that examined vaccine protection across five or more different pathogen challenge doses (e.g. poliovirus and malaria), vaccine efficacy decreased in accordance with our predictions from the simulation.  
     
   Firth's penalized‐likelihood logistic regression revealed that sites separated by an oceanic break ≥ 20 km are 31 times more likely to have a ΦST estimate greater than zero, as compared to sites that are not separated by a break [coefficient = 3.43 (odds ratio = 31.00); SE = 0.97; *P* < 0.001]. There was no significant improvement over this ‘oceanic break’ model when ‘Euclidian distance’ was included as a predictor (χ2 = 0.77, d.f. = 1, *P* = 0.38). These results were robust to changing the structure metric to microsatellite‐based *F*ST. Here, the best‐fit model revealed that sites separated by a break were over 18 times more likely to have a nonzero value of FST relative to sites that are not separated by a break [coefficient = 2.92 (odds ratio = 18.52); SE = 1.53; *P* = 0.007] and that adding Euclidian distance as a predictor did not improve model fit (χ2 = 1.12, d.f. = 1, *P* = 0.29). Taken together, these results reject the ideas that SGS is random (H0) or that SGS follows an isolation‐by‐distance pattern (H2), but provide support for H1 that posits that SGS is associated with oceanic breaks.