FISH 6003: Statistics and Study Design for Fisheries - Assignment Guide

Dr. Brett Favaro

10% of your course grade is earned by participation. Just show up, be yourself, and participate!

The remainder will be earned by completing assignments. These fall into three categories:

- Presentation on a special topic (20%)
- Major assignment (50%)
- Minor assignment (20%)

This document outlines the details and grading criteria for each assignment.

Presentation on a special topic

We will cover a lot of ground on this course, but it is only the tip of the proverbial iceberg that respresents the whole of possible approaches used to quanitatively assess data. In this presentation, students will identify a statistical technique that we will not be covering as part of the core course.

Students will teach us about the technique - what it is used for, why it is useful, how to use it, and how to assess its validity. Students will produce a 10-15 minute presentation.

Deliverable:

Prepare a 10-15 minute presentation (plus 5-10 minutes for audience questions) that accomplishes these objectives:

1. Describe the statistical technique. What does it do? What is it for?
2. How do you assess the fit of this technique with your data? How can you tell if it worked or not?
3. Show an example of the technique used in a scientific paper
4. Demonstrate how to use the technique and interpret its results, with real or simulated data
   - This should be a worked example, with fully commented R code that explains what is being done at each step
5. Share **reproducible** R code with the class

Code that is shared for this assignment must be fully reproducible, extensively commented, and contained within an R project.

Timeline

By Week 3: Select topic. Review with instructor. Determine presentation schedule.

Starting Week 7: Present. One presenter per week.

Grading scheme:

Part 1: The presentation

The presentation is worth 30/40 marks of the assignment grade, and will be assessed as follows:

Describe the statistical technique. What does it do? What is it for?

Value: /10

- Statistical technique should be clearly and accurately described. Literature explaining the technique is cited.
- Demonstrate deep understanding of the technique. Description is not superficial. Student constructively engages with questions on the topic.
- Contrast this technique with other methods. What does it add that would be missed with something simpler? How is it better than something more complex?

How do you assess the fit of this technique with your data? How can you tell if it worked or not?

Value: /10

- Any diagnostic plot techniques are explained
- An example of the statistical technique not working, and the technique working properly, are demonstrated
- This may be expressed as a checklist. What steps should be followed to ensure that the output of the technique is valid?

Show an example of the technique used in a scientific paper

Value: /5

- Correctly identify this technique in a paper
- Explain WHY the technique was used, and describe the data on which the technique was applied
- The results and output are explained, along with any diagnostic figures
- The results are interpreted. They found X - how did it allow them to conclude Y?

Demonstrate how to use the technique and interpret its results, with real or simulated data

Value: /5 - Show how to implement this technique in R. Present R code that you have written for Part 2.

Part 2: A reproducible example

The remaining 10/40 marks will be based on assessment of R code that students write, which will demonstrate how to use the technique in R. This may be done with real data (i.e. taken from a paper) or simulated data.

Note that many techniques or packages come with worked examples. It is not sufficient to just copy the exact example included with the package. While you may be writing code that is pre-packaged, you should apply the technique to different data, even if you just make up imaginary data yourself.

A worked example, with reproducible, fully commented R code that explains what is being done at each step

Value: /10

- Code is organized into an R project (see https://mifisheriesscience.github.io/courses/6002Data/6002Week11/)
- Code is fully reproducible (i.e. anyone can open the project and run it without changing any code).
- Code is well-commented, organized, and easy to follow

How to Submit

The presentation will take place in-class, on a Tuesday. Students should create a zip file contianing all sub-folders of their R project. The Zip file should be emailed to the course instructor **on or before the day of the presentation**.

Value: 20% of course grade

Major Assignment

These data may be something that you have collected yourself for your research, from a public database, or as attached by another publication.

Essentially, the major assignment is to do all the basic groundwork that would need to be done to prepare an analysis for publication.

Part 1: Obtaining and Describing Data

The goal of part 1 is to obtain and describe data. Key questions, which will be answered in two pages or less (single-spaced, 12-point Times New Roman):

- Where are the data from?
- What are the data? (i.e. how many points? What are they measuring? How are they measured?)
- If you got the data from someone else, are there any constraints around using them?
- What are **three key research questions** you will ask of these data? In other words: *What are your biological hypotheses* that you will be testing?

The purpose of this section is to determine whether the data you obtained will be sufficient for the major project.

Timeline and Submission

Part 1 is due by the end of week 5 and should be submitted by email to the instructor.

Grading

- Accurately describes the data source: /1
- Accurately describes the data itself (number of points, what they are measuring, etc.): /1
- Three research questions, answerable by these data, are stated: /3

While these are essentially completion marks, the instructor may require at this state that more or different data be incorporated into your project. If the dataset is too limited it can be challenging to produce an interesting project. Further, if the data are too complex, it may represent a challenge that is beyond the scope of this course.

Part 2: Data Exploration

In Part 2, you will conduct a data exploration, following a slightly modified version of the procedure explained in Zuur et al. 2009. You should report the results of each step. These steps include:

A. Re-stating the biological hypotheses (in case they changed from Part 1)

B. Visualize the experimental design, with a diagram

C. Conducting the data exploration:

1. Outliers Y and X
2. Homogeneity Y
3. Normality Y
4. Zeroes Y
5. Collinearity X
6. Relationships Y and X

7. Interactions
8. Independence of Y

D. Specify the statistical model that you intend to use, and why.

Timeline and Submission

Part 2 is due by the end of week 9 and should be submitted by email to the instructor. We will review R projects in class.

The deliverables for Part 2 should be organized as follows:

- A properly-organized R project should be created that contains all data and scripts
- A document that includes everything, including plots and relevant R output. You should both show the direct output from R, and interpret it in your own words
- A separate .R script that includes all necessary code to conduct the data exploration

While it is not a requirement, I encourage you to use R Markdown for this assignment, so that diagnostic plots can be embedded directly with the text. See: https://mifisheriesscience.github.io/courses/6002Data/ 6002Week11/

Grading

- Biological hypotheses are clearly stated and sensible, given data: /3
- Experimental design is clearly visualized, including any nested structure: /3
- Steps 1-8: /3 each. Note that many of these steps should include a plot. (1 mark for the plot, 2 for the explanation).

Specifying your intended statistical model:

- Model is identified and explained: /3
- Your choice is defended using literature: /3

Total: /36

Part 3: Conducting and Reporting the Analysis

In Part 3, you will conduct a regression-type data anlaysis, following the reporting procedure explained in Zuur and Ieno, 2016. The steps to this assignment include:

1. Identify the dependency structure in the data
2. Present the statistical model
3. Fit the model
4. Validate the model
5. Interpret and present the numerical output of the model
6. Create a visual representation of the model
7. Simulate from the model

Note that Part 3 comes immediately after Part 2. There is no need to re-do your data exploration in part 3, as it will already be completed in part 3.

Timeline and Submission

Part 3 is due by the end of week 12 and should be submitted by email to the instructor.

The deliverables for Part 3 should be organized as follows:

- A properly-organized R project should be created that contains all data and scripts
- A document that includes everything, including plots and relevant R output. You should both show the direct output from R, and interpret it in your own words
- A separate .R script that includes all necessary code to conduct the data exploration

It is recommended that you simply add to the project folder that you created for Part 2. Place everything into a zip file and submit to the instructor.

While it is not a requirement, I encourage you to use R Markdown for this assignment, so that diagnostic plots can be embedded directly with the text. See: https://mifisheriesscience.github.io/courses/6002Data/6002Week11/

Grading

Steps 1-7:

- Each should include a chunk of R code, output, and intepretation. /5 for each step

Total: /35

Between parts 1, 2, and 3, the Major Assignment is worth a total of 76 marks, which correspond to 50% of the course grade.

Minor Assignment

The minor assigment asks you to take a paper, determine its key hypothesis, and design a model to test that hypothesis.

In pairs select one of the following papers:

1. Virgili, M., Vasapollo, C., & Lucchetti, A. (2018). Can ultraviolet illumination reduce sea turtle bycatch in Mediterranean set net fisheries?. Fisheries Research, 199, 1-7.
2. Levesque, J. C., Hager, C., Diaddorio, E., & Dickey, R. J. (2016). Commercial fishing gear modifications to reduce interactions between Atlantic sturgeon (Acipenser oxyrinchus oxyrinchus) and the southern flounder (Paralichthys lethostigma) fishery in North Carolina (USA). PeerJ, 4, e2192.
3. Utne-Palm, A. C., Breen, M., Løkkeborg, S., & Humborstad, O. B. (2018). Behavioural responses of krill and cod to artificial light in laboratory experiments. PloS one, 13(1), e0190918.
4. Robbins, W. D., Peddemors, V. M., Kennelly, S. J., & Ives, M. C. (2014). Experimental evaluation of shark detection rates by aerial observers. PloS one, 9(2), e83456.

If none of these papers appeal to you, you may also select a different paper that would be analyzable by a regression-type model. If you choose to do this please show the paper to Brett by March 9 for approval.

It is okay for multiple pairs to do the same paper.

Read the paper and do the following:

1. Define the hypothesis in words. /4
2. Describe Y. Is it a count, proportion, etc? /2
   - How was it measured? /1
3. Identify all covariates. What types of data are they? /4
4. Sketch the study design, indicating any dependency structure. /4
   - Note the number of replicates within each treatment group. /2
5. Brainstorm: What sources of uncertainty might there be in the measurements of each covariate, and Y? (At least one per variable) /5
6. Define each of three parts of the GLM:
   - Write the equation for the predictor function. /10

- What distribution defines Y? /4
- What is the link function? /4
7. Write the syntax to run this model in R (e.g. mod <- lm(Y_var ~ Covariate1 + Covariate2)) /10
8. Did the authors run the model you specified? If not, what did they do instead? Critique their approach versus yours. /10

Total: / 60, scaled to 20% of course grade

Assessment Criteria

All of the above papers are analyzable by a regression-type analysis that we will have covered by the end of Week 9 of the class.

Point-form is acceptable where it facilitates clarity. The study design may be hand-drawn and scanned, or done in Powerpoint. Regardless it must be legible.

For the critique (step 8) please focus on the fundamentals. Are assumptions likely to be violated by any of the models? How does YOUR model address that?

Note that I am not asking you to actually run the model or perform a validation (that will be done in the Major Assignment).

Due Date

Please submit by email, by March 26. I recommend you try to complete it earlier so as to allow you to focus entirely on the major assignment by the end of the course.