# ESCI 502 Final Exam

2023

This exam is worth 52 points and is worth 25% of the grade for the class. You will turn in a Word/PDF file and your R code. Not all questions will require R work. To complete your exam, you will upload your files to Canvas by the deadline.

This exam is open book. As such, I expect precise language in your answers, which should be concise and to the point.

**Do your own work**

This exam is to be completed by you and you alone. You may not work with your classmates or with your data analyst partner. You may consult your own notes and labs, books, Canvas, R documentation, and internet help. You may contact me with questions. If you are uncertain about what it means to turn in your own work, please see the [WWU Academic Honesty](https://catalog.wwu.edu/content.php?catoid=15&navoid=3320#Academic_Honesty_Policy) policy.

**Tips for full credit**

Use clear, concise, and precise language.

Be specific about what you find. Think about what you are testing and what exactly your findings refer to; provide statistical evidence.

Copy plots to your Word document using the *export* function on the plots pane of R Studio so all elements are legible (and make them big enough to read on a laptop). Label plot axes for all final plots (not needed for exploratory plots).

Read the full problem before diving in so you know where you are headed and can get there efficiently, with a logical workflow.

**Code**

Provide clear code. Use comments and line breaks. Use logical object names that are not too long. Make sure that your code will run from top to bottom without error. To check this, restart R once you are done and run your entire script, checking the console for output errors. *In addition to turning in your R file, please copy/paste code at the end of your document (this helps with a quick scan while grading).* If your code does not run, you will lose points.

**Tips if you get stuck**

Check your object labels.

Check capitalization and punctuation.

Run ?*function* (e.g. ?lm) to make sure you have specified the arguments correctly.

Save your code, restart R, and re-run your code again. Sometimes an object will be saved in the environment that was a relict of old code.

Check any subsets and intermediate analysis as you go to make sure they are what you expect.

**Grading**

Point values for each question are listed below next to the questions. Partial credit will be given. However, answers without associated code (where needed to produce an answer) will not be given credit. Your code counts for 10 points on this exam—make sure it runs and is legible.

You’ve got this!

**1. Describe (5 pts):**

Define the three components of a generalized linear model and describe how you would choose a proper distributional family.

**2. Draw (5 pts):**

Draw (roughly, by hand is fine) two plots. Each plot will have a continuous response variable and a continuous explanatory variable as well as a categorical explanatory variable with 3 levels. Draw the predicted line for 3 separate groups (levels) where:

Plot (A) will show the 3 groups with no interactions

Plot (B) will show an interaction effect between the categorical and continuous predictors, with respect to the response variable.

In short, demonstrate what an interaction effect would look like and what a plot without an interaction effect would look like, and describe an interaction effect in words.

**3. Define (3 pts):**

Random Effects

Collinearity

Maximum Likelihood Estimation

**4. Discuss (9 pts)**

a. Why using a generalized linear model is preferable to transforming your lognormal data and using a linear model.

b. Model goodness of fit versus model selection.

c. Why partial regression plots are important to interpreting a multiple regression problem.

**5. Design (10 pts)**

This quarter you have learned a number of statistical techniques centered around the general regression framework. Build a flowchart with the techniques we covered. Explain when/why you would use each technique. There are many examples out there, but your task is to design a decision framework with the tools you have learned this quarter (not all of the possible statistical techniques available on the internet). You can do this by hand and upload an image (make sure it is clear) or use a graphical software like PowerPoint or other

**6. Do (20 pts)**

Suppose you’ve been busily working in the laboratory sectioning abalone shells for the purpose of age determination. In addition to age information, you’ve also noted the sex of each specimen and you have taken several shell measurements (length, diameter, and height) and weight measurements (whole, shucked, viscera, and shell). Ultimately, you would like to develop the best predictive relationship of age as a function of the variables you’ve measured. Because sectioning abalone is tedious and you’d rather be skiing.

Preliminary analysis has shown that **shell diameter**, **shell weight,** and **shucked weight** are promising indicators. Formulate a candidate set of linear model parameterizations (designed to predict abalone age) involving combinations of the three variables; also include the **categorical sex variable**, as gonad may make a difference in the relationship between weight and age. Do not fit any interactions, even though you might suspect sex and shucked weight have a potential interaction based on the previous info (here, main effects only).

Assess the input data and assumptions of your modeling framework with standard techniques.

Write out the candidate models, fit them to the data, and use model selection to discriminate among the parameterizations.

Report your AICs and associated results in table form to demonstrate how you selected a best fitting model.

Assess your best fitting model for violation of assumptions and report the parameter estimates, standard errors, and relevant statistics associated with the best fitting model(s).

Make a concluding statement about your results.

Data file: *abalone2.csv*

+5

Make a plot showing your model predictions with a caption that describes what your plot depicts. Predictions must be estimated using the predict() function. Credit will not be given for using geom\_smooth() in ggplot() unless you have modified this function to plot your model fit (e.g., no “method=lm” or the default loess, unless you can clearly justify why this would be a true representation of your model).

**Assessment (1 pt each)**

How long did it take you to do this exam?

How fair do you think this exam was (1-5, 5 being fair)?

Looking back over the course, what do you wish we had covered in class better?

What is the most important thing you learned this quarter?