Midterm

Sarah Roberts

2023-02-23

Answer the following questions in 1 knitted PDF document. Your answers do not need to be in paragraphs (they can be bullets). Also turn in your .Rmd file.

#Problem 1 - 30 Evaluate the effect of salt on plant biomass growth in 24 experimental vegetation plots (download salt.csv). The assigned treatment (one of 6 levels of salt addition: 10, 15, 20, 25, 30, 35 g m-2) is applied to the soil of a plot, and at the end of the experiment the biomass of plants in each plot is measured. The experimental units are grouped into four blocks of six plots each, based on geographic proximity, and the treatments are assigned completely at random within each block. Thus, each treatment occurs exactly once in each block.

Include: a) Null and alternative hypotheses of your tests (2.5) b) Justification for your choice of test (2.5) c) A description of how you checked the assumptions of your statistical test (10) d) Results of your statistical test, interpreting your test in 2-3 sentences that include the appropriate reporting of the statistics as well as an appropriate figure (10) e) An interpretation of any necessary post-hoc tests (5)

#Problem 2 - 30 Pangolins are scaly anteaters that inhabit the tropical forests of Asia and Africa, and hunted for their meat and their scales, which are made of hair. A researcher wants to evaluate the effect of diet on the thickness of pangolin scales, with the idea that thicker scales would better protect pangolins from predators. She rears pangolins and provides them with identical diets, but different doses (0.5, 1, 2 milligrams) of supplements (Vitamin B and Zinc). Download the ScaleThickness.csv file and analyze the data to determine the potential effects of doses and supplements and whether there is an interaction between the two of them.

Include: a) Null and alternative hypotheses of your tests (2.5) b) Justification for your choice of test (2.5) c) A description of how you checked the assumptions of your statistical test (10) d) Results of your statistical test, interpreting your test in 2-3 sentences that include the appropriate reporting of the statistics as well as an appropriate figure (10) e) An interpretation of any necessary post-hoc tests (5)

#Problem 3 - 35 A company is looking for a faster internet service. To decide between two different internet providers (Turbo Net and Speed Web), the company performs an experiment in which it collects data on website loading times from each of the providers (download Internet.cvs). Is there a significant difference in the website loading times of the two providers? If so, which provider should the company choose? Make sure to do the following: (a) specify your hypotheses (5) (b) check that the data meet the assumptions of the statistical test you plan to use (10) (c) Test your hypotheses. If your data do not meet the assumptions, test your hypotheses both by transforming the data and by using a nonparametric approach (10) (d) Explain the results of your statistical tests. (10)

#Problem 4 - 50 Use data (yaleEPI2018.csv) from Yale's Environmental Performance Index (EPI) to evaluate if GDP per capita (GDPpc) influences the Yale Environmnetal Performance Index (yaleEPI2018.csv). The units of GPDpc are USD per person (2018 USD). The Environmental Performance Index (EPI2018Score) is an index for each nation on its environmental performance. It is an index combined from numerous weighted measures. (a) Hypothesize a relationship between your variables. Include the null and alternative hypothesis. (5) (b) Calculate the correlation coefficient between the two variables. Interpret. (5) (c) Develop and discuss the scatterplots between the two variables. Does the relationship look linear or linear-log (or something else)? Should you transform one or both variables? Include your scatterplots (original, transformed) and

discussion.(10) (d) Perform a simple linear regression analysis on the two variables (after transformation, if needed). Make sure that your x and y variables are correct. Write the model equation and interpret the coefficient for the slope (10) (e) Interpret p-value of the slope coefficient in terms of hypothesis above. (5) (f) Interpret the R2 value (5)

#Problem 5 - 55 The dataset ozone.data.csv consists of data on air quality (in terms of ozone concentration) and several weather variables, including temperature, solar radiation, and wind speed (don't worry about units here). Model the effect of the weather variables on ozone concentration, making sure to consider informative interactions between the independent variables. Once you have determined the best model to predict ozone concentration. In a short write-up, provide the following: (a) Null and alternative hypotheses of your model (5) (b) Results of your statistical test, interpreting the fit of the selected model in 2-3 sentences that include the appropriate reporting of the statistics in a table (20) (c) An interpretation of the regression model coefficients (i.e., what do each of the main effect(s) and interaction effect(s) mean (10) (d) A description of how you checked the assumptions of your statistical test (10) (e) Graphs that depict (i) the relationship between the independent variable and dependent variable when a variable is only included as a main effect, (ii) the relationship between the interacting independent variables with the dependent variable (10)

#Problem 6 - 50 Read the Fish paper and answer the following questions: a) What is the research question underpinning the study (10 points). b) What type of model did the authors choose to use and why? (5 points) c) Describe the data that the authors used. What is the response variable? Explanatory variables? (10 points) d) Interpret the equation and results from table 4 for Caesio lunaris. (10 points).
e) How did the authors compare models? (5 points) f) Discuss two of the limitations the authors identify in their analysis? Discuss one potential limitation they do not bring up. (10 points).