STAT210/410

Assessment 2: Multiple Linear regression

DUE: 23rd March 2024, 11:59pm AEDT

Litter decomposition plays an important role in carbon and nutrient cycling. One aspect which impacts variation in decomposition rates is degradation by UV light. Researchers[1] investigated the relationship between litter quality and their decomposition rates in a desert ecosystem. Leaf litter samples from 17 desert species were sealed into litterbags and placed at three different positions; on the soil surface under strong solar exposure (light), on the soil surface under shade conditions (shade), or buried in the soil at 10 cm depth (buried), for a whole year.

The data set, `decomp.csv` contains 51 observations with 10 variables:

- **Species:** Species of plant the leaves came from.
- **Treatment:** one of the three positions of the litterbags; light, buried, shade.
- **Litt.DMC**: dry matter content of the leaves (mg g^{-1}).
- **Litt.SLA:** Specific leaf area (mm2 mg-1).
- **Litt.C**: Percentage concentration of carbon (%).
- **Litt.N**: Percentage concentration of Nitrogen (%).
- **Litt.CN**: Carbon to Nitrogen ratio.
- **Litt.P**: Percentage concentration of Phosphorous (%).
- **Litt.lignin**: percentage concentration of lignin (%).
- **K:** Decomposition rate

NOTE: As part of your coding, you might want to select a different baseline treatment. I would suggest "light".

Question 1 **15 marks**

Perform some exploratory analysis. Use the `ggpairs()` function to plot the data and make one other plot or table with summary statistics that you think provides useful information about this study.

In a few sentences, summarise the key correlations between the predictors and the response variable and any correlations between the predictors. Also include useful insights about the study from your second plot or table.

NOTE: `ggpairs()` is a function within the `GGally` library, so don't forget to load this library before you attempt to produce the plot.

[1]Data sourced from: Liu, Guofang et al. (2017), Specific leaf area predicts dryland litter decomposition via two mechanisms, Journal of Ecology, Article-journal, https://doi.org/10.1111/1365-2745.12868

Question 2 **15 marks**

Fit a main effects model to the form:

$$\widehat{K} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k$$

Do not include Species in your model. Present the table of regression coefficients in your report and write down the least squares regression equation from the model.

NOTE: Rember for groups, the intercept is your baseline and then you have k-1 dummy variables.

Question 3 **10 marks**

Write down and test appropriate hypotheses for all terms in the model. Which terms are significant in predicting decomposition (K) at a 5% level?

Now drop the non-significant terms and refit the model using only the explanatory variables that are significant. NOTE: This is not the best way to select variables, but as a way to practice fitting models. We will explore better variable selection processes in the weeks to come.

Question 4 **5 marks**

Present the table of regression coefficients for the refitted model in your report and write down the least squares equation.

Question 5 **10 marks**

Test a model with the main effects of the refitted model from Question 4 and include the predictors; litt.C & litt.CN. Would you consider this a better model than the simplified model you fitted in Question 4? Why or why not?

Hint: look at the adjusted R^2 value.

Based on your answer to Question 5, either drop or retain these predictors in the model. This will be your final model.

Question 6: **10 marks**

Produce diagnostic plots for the final model and explain what can be understood from these plots. Do the conditions of a linear regression appear to be met?

Question 7 **20 marks**

Using your final model, estimate mean decomposition (K) and a 95% confidence interval for the estimate, as well as the prediction interval when:

- Treatment = shade, litt.DMC = 300, litt.SLA = 15, litt.C = 56, litt.N = 1.4, litt.CN = 30, litt.P. = 0.05, litt.lignin = 45
- Treatment = buried, litt.DMC = 350, litt.SLA = 20, litt.C = 70, litt.N = 1.2, litt.CN = 20, litt.P. = 0.03, litt.lignin = 55

NOTE: You may not need to include values for all 8 explanatory variables, depending on your final model.

Interpret these values and comment on the reliability of these predictions. As part of your comment on reliability make a plot indicating where the these prediction points fall in relation to the data you collected.

Question 8 **10 marks**

Write a concise, informative conclusion based on your analysis and results. You might want to include:

- The final model equation
- Comment on the adjusted R2 value
- Comment on the conditions of the model
- Comment on which predictors were important
- Interpret any relevant confidence intervals

Presentation **5 marks**

Marks will be awarded for notation and presentation:

- Clear expression, correct use of terminology and mathematical notation. Use the equation editor in word to write your equations.
- Presentation of figures and tables: ensure that you include all relevant R output as figures (plots) and tables (numerical output). No screenshots of output/plots.
- Clearly and concisely annotated R code: include your R script file as an appendix. No code should be in the written part of your assignment. All code should be in the appendix and there should be no output in your appendix.