Background

A Bachelor’s student, Martin, has performed an experiment measuring soil carbon (C) stocks in grasslands grazed by dairy cattle in France. He conducted his experiment across varying pastoral farms spread out across two provinces, Orleans (Central France) and Montpellier (South France).

Martin wants to identify how stocking rate (animal units per ha, AU ha-1) impacts soil C stocks in order to be able to recommend a level to reduce the climate impact of French pastoral dairy farms.

Based on previous research, he suspects that moderately grazed farms will have higher soil C than both heavily grazed and lightly grazed. But he is unsure how to define ‘light’ vs ‘heavy’, and how various factors affect his dependent variable of interest.

Based on the lecture slides provided (‘Fundamentals of data science’) and the R file ‘Dat.sci.funds.R’ complete the following exercise.

Exercise

1. Variable classification

What category of variable is stocking rate (‘stock\_rate\_value’)

What about the variable ‘stock\_rate\_group’?

Convert stock\_rate\_group’ to an ordinal variable by assigning factor levels to it.

1. Testing normality

Test whether the variable ‘stock\_rate\_value’ follows a normal distribution using the ‘Shapiro Wilks test’

Now perform the same test for rainfall.

1. Transformation

Create a new variable based on the standardisation of stocking density (the quantitative variable). For the central tendency use the mean, and use standard deviation as the ‘spread’ variable.

Check that the transformation was successful based on the mean and standard deviation of the newly standardised variable.

1. Correlation coefficients

As a first evaluation of the impact of stocking density on soil C, evaluate the correlation between the two variables.

Calculate Pearson’s correlation coefficient.

Now calculate Spearman’s correlation coefficient, comparing soil C to the factorized groups of stocking density you calculated in part 1. Note that for this, you will have to include stocking rate as a numeric value.

How do the correlation coefficients differ?

1. Hypothesis tests

Martin needs to make sure that the rainfall across the two blocks is representative of the national average rainfall for France, which 1,350 mm yr-1.

Perform a hypothesis test to determine if the mean rainfall in the experiment is significantly different from the national average. Can the hypothesis of equality of means be rejected? If yes, which of the two blocks differs most from the national average?

Finally, test the equality of variance between rainfall in the two blocks. Do the two blocks have equal or unequal variance?

1. ANOVA

Now let’s see how the experimental factors and site variability affects soil C stocks across the plots.

First, test the effect of site alone by conducting a one-way anova of ‘site’ on soil C

Second, add the main experimental factor, stocking density in a two-way ANOVA/

Third, add the remaining experimental factors.

Finally, include the remaining two climate variables.

How do the significance of treatments differ across the ANOVAs of varying complexity? Which factors are relevant and irrelevant with respect to soil C?