SMiB
Homework/Labwork 2

Please, choose one problem. The first one relies on a classical linear model, requires a model selection (preferable for those who are new to this type of modeling). The second problem requires the application of a mixed model.

Problem 1. Due March 11th, 2024 EOD
The file "no2.txt" contains data on the air pollution (specifically NO_2 levels) at a traffic intersection. Data description:

1. **NO2 particles - the response variable**

2. number of cars per hour

3. temperature 2 meters above the ground (Celsius)

4. wind speed (meters/second)

5. temperature difference between 25m and 2m above the ground

6. wind direction (degrees)

7. hour of day

8. day number since October 1.

Find a suitable regression model to explain the NO_2 levels. Write a brief report summarizing your work and interpreting your results. Provide meaningful plots.

Problem 2. Due March 17th, 2024 EOD
The file "crop_yields.csv" contains data on crop yield in various countries over the years.
Data description:

1. Area (the name of the country)

2. Crop (the name of the crop, wheat, rice, maize, etc)

3. Year (the years differ somewhat between countries, it is ok to not treat subsequent years as a time series, it is enough to assume correlation between all yields in one country and one crop)

4. **yield [hg/ha] - the response variable**

5. average_rain_fall_mm_per_year

6. pesticides_tonnes

7. avg_temp

8. surface_area_sq-km_2001 (country's area, not the crop area, added by the CS students)

9. gdp_percap_currUSD_2001 (added by the CS students, could be correlated with the intensive/extensive agriculture)

   You can focus your analysis on a single crop of your choice. Please decide on a type of mixed effects that need to be applied. Try to analyze the random effects (use ranef call in lme4 library to extract the random effects from the fit model), see if they effects cluster in any way, that is interpretable. Provide meaningful plots.