PROJECT 3: OTHER REGRESSIONS

MASM22/FMSN30: LINEAR AND LOGISTIC REGRESSION 7.5 HP, 2024

Oral presentation: **Wednesday 22 May – Friday 24 May**

Submit the slides **before** your presentation

More models for the PM_{10} emissions or the number of cars

You should choose **one** of the following three alternatives and give a 15 minute oral presentation of your finds. The data from both projects, together with the new variables `Cars_nbr` = the number of personal cars and `Population` = the number of inhabitants, is available in the file `kommunerProject3.xlsx`.

Alternative A. Poisson and Negative binomial regression

Continue Project 2 and model the **number of personal cars**, `Cars_nbr`, using Poisson and/or Negative binonial regression. Do **not** use `PM10` or `Cars` as explanatory variables but feel free to use any of the other variables from both Project 1 and 2. Remember to check for multicollinearity!

Assume that the number of cars is proportional to the number of inhabitants by using log-Population as an offset variable. Also test this assumption by instead using log-Population as an explanatory variable and investigate its β-parameter.

Determine whether you can use a Poisson regression or if a Negative binomial regression fits the data better and use the usual model selection and validation tools to find a suitable model and investigate its properties.

Alternative B/C. Ordinal and multinomial logistic regression

Continue either Project 1 or Project 2 and divide **PM_{10} per 1000 inhabitants**, `pm10` or the **number of cars per 1000 inhabitants**, `Cars` into three, or more, groups with increasing values.

If you want to divide `pm10` into g groups of equal width, use

```
kommuner |> mutate(pm10_cat = cut(pm10, breaks = g)) -> kommuner
```

If you instead want g groups with equal number of observations, use

```
kommuner |> mutate(
  pm10_cat = cut(pm10,
                 breaks = quantile(pm10, probs = seq(0, g)/g),
                 include.lowest = TRUE)) -> kommuner
```

Model the probabilities for the different categories using ordinal logistic regression. Feel free to use any of the variables from both Project 1 and 2 as explanatory variable but remember to check for multicollinearity! Use the usual model selection and validation tools to find a good model and investigate its properties.

Additionally, ignore the fact that the categories are ordered and model the probabilities using a multinomial logistic regression as well, and compare the results to the ones from the ordinal logistic regression.

End of Project 3