02424 Assignment 2, 2021

This is the second of three mandatory assignments for the course 02424. It must handed in using the Learn (time and date is given thee and in the general course plan in Learn). The submissions most contain one collected attached file in Portable Document Format (PDF), other document formats will not be accepted.

The report of each assignment must be prepared in groups of 3, and the final grading wil be based on the reports and the (individual) oral exam.

When writing the report please explain carefully what you did in each step, back up your statements with quantitative measures when possible, explicitly write down all models used in mathematical notation, and last but not least keep it short and concise.

Ozone model

In this part you will model ozone concentration in Los Angeles, the data is oploaded to Learn along with this asignment, but is also included in the package `gclus`, and more information on the data can be obtained from the there, e.g.

```
library(gclus)
data(ozone)
head(ozone)

##   Ozone Temp InvHt Pres Vis  Hgt Hum InvTmp Wind
## 1     3   40  2693  -25 250 5710  28  47.66    4
## 2     5   45   590  -24 100 5700  37  55.04    3
## 3     5   54  1450   25  60 5760  51  57.02    3
## 4     6   35  1568   15  60 5720  69  53.78    4
## 5     4   45  2631  -33 100 5790  19  54.14    6
## 6     4   55   554  -28 250 5790  25  64.76    3
```

Part 1

You should only consider additive and linear effects

1. Make a short presentation of the data
2. Fit a general linear model, and perform a residual analysis (you may consider transformations of the dependent variable)
3. Fit at least two different (sensible) generalized linear models to the data (you do not have report residual plots of all the models here), and compare these model by a quantitative numbers (you can play around with the distribution assumption and the link function).
4. Compare the model under question 2 and the model chosen from question 3, which one would you prefer (if you choose a quantitative measure you may need to take the transformation into account)?

5. Reduce the model you have chosen under the previous part.

6. Present the final model.

Clothing insulation level: Count data

In this part you should analyze the dataset `clo.count` the data set is constructed using the data you used in the first assignment, but `clo` now contain the number of times that each subject change clothing insulation level during a day, total time time of observation (`time`), number of observations during the day (`nobs`), the sex of the subject (`sex`), and average outdoor and indoor operation temperature (`tOut`, and `tInOp`), during the day.

1. Devellop a generalized linear model, based on the Binomial distribution, when ignoring `subjId` and `day`

2. Devellop a generalized linear model, based on the Poisson distribution, when ignoring `subjId` and `day` (You should consider including an offset in your model).

3. Discuss the interpretation of the two models you fitted above.

4. Write a small conclusion of your findings.

Fan Speed:

In the part you should analyze the dataset `CeilingFan.csv`. The dataset consist of a thermal sensation vote (`TSV`) on a 3 level scale, the speed of a fan `fanspeed` (also 3 levels), type of fan `fanType`, and a subjects identifier `subjId` (which you should ignore here).

- Using the usual cotengency table test if independence between TSV and fanSpeed, can be assmued.

- Using the package `ordinal` costruct 2 models to test the same hypothesis as in 1 (but with the method `anova`)

- Fit and devellop a model for TSV as a function of fanSpeed and fanType (hint you may consult Example 4.12 of the textbook).

- Present the result, e.g. interpretaion of the parameters and some visual presentation.

References

[1] Fanger, P.O. (1970). *Thermal Comfort Analysis and Applications in Environmental Engineering*. McGraw-Hill, New York.

[2] Schweiker, M. and Wagner, A. (2015). *A framework for an adaptive thermal heat balance model (ATHB)*. Building and Environment (94), Elsevier Ltd.