02424 Assignment 2

This is the second of three mandatory assignments for the course 02424. It must handed in using the Campusnet (time and date is given at Campusnet). 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 keep it short and concise.

Ozone model

In this part you should model ozone concentration in Los Angeles, the data is oploaded to campusnet 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

In the fist part 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
3. The analysis above should suggest a transformation. Use a simple transformation on the dependent variable and perform the residual analysis again
4. 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).
5. Compare the model under question 3 and the model chosen from question 4, which one would you prefer (if you choose a quantitative measure you will need to take the transformation into account)?

6. For the chosen generalized linear model write down explicitly the diagonal elements of the weight lomatrix ($\boldsymbol{W}$) as a function of μ_i, check your calculation by comparing the dispersion matrix of the parameters from the R function (`summary(fit)$cov.scaled`) with your own calculation.

Part 2:

1. Devellop the model you have chosen under the previous part, you might consider both higher order polynomials and interaction terms.
2. 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. Fit the two models above but including `subjId` rather than `sex`. Present the parameters graphically and discuss the interpretation.
5. Write a small conclusion of your findings.

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.