

Data exploration, regression, GLM and GAM course

Highland Statistics Ltd www.highstat.comn

Exercise 15: Ragweed data

Data description

The variable of interest is ragweed pollen. To avoid allergic reactions in humans due to pollen, it is important to have a model that can predict daily pollen levels. And it is also important to know which environmental variables can cause high pollen levels. The data were sampled in Michigan, US in 1991-1994.

The variables are given in Table 1. The data are in the file ragweed.txt. Each row in the text file represents a day.

Table 1. Variables in the file ragweed.txt.

Variable	Description	Type
Ragweed	Ragweed level for that day (grains / m ³)	Count, response variable
Temperature	Temperature	Continuous explanatory variable
Rain	Indicator for rainfall (1 is > 3 hours of steady or brief rain; 0 is otherwise)	Categorical explanatory variable
Windspeed	Windspeed (knots)	Continuous explanatory variable
DayinSeason	Day number in year	Continuous explanatory variable
Year	Year	Continuous explanatory variable

References

These data were originally analysed in Stark et al. (1993) and in Ruppert, Wand, Carroll (2003).

- Stark, P. C., Ryan, L. M., McDonald, J. L. and Burge, H. A. (1997). Using meteorological data to model and predict daily ragweed pollen levels. *Aerobiologia*, 13, 177-184.
- Ruppert, D., Wand, M.P. and Carroll, R.J. (2003) *Semiparametric Regression* Cambridge University Press

Underlying question and task

The task is to model Ragweed as a function of the covariates. Besides knowing which covariates are important we would also like to know whether the seasonal effect is changing over time.