UNIVERSITY OF EDINBURGH

SCHOOL OF MATHEMATICS

Generalised Regression Models

GRM

Bruce Worton

Semester 1, 2022–2023

Multiple linear regression models and generalized linear models are widely used in statistical analysis. These models are appropriate for investigating how the distribution of a response variable is influenced by explanatory variables. Linear regression models under the assumption of Normality are shown to be special cases of a general Normal Linear Model, which is conveniently expressed in matrix notation, while generalized linear models provide a broad framework for statistical modelling of discrete or continuous data. Special cases of such models are considered, as well as models with random effects.

Course outline

1. Statistical modelling.
2. The exponential family of distributions and GLMs.
3. Multiple regression.
4. Inference for linear models.
5. Generalized linear models.
6. Using models with random effects.

Textbooks

1. Dobson, A.J. & Barnett, A., *An Introduction to Generalized Linear Models*, 2nd/3rd Edition, Chapman & Hall/CRC.
   Provides an introduction to linear models and generalized linear models.

2. McCullagh, P. & Nelder, J.A., *Generalized Linear Models*, 2nd Edition, Chapman & Hall.
   Provides a more advanced and detailed coverage than Dobson's book.

3. Venables, W.N. & Ripley, B.D., *Modern Applied Statistics with S*, 4th Edition, Springer.
   Comprehensive guide to S/R. Introduces the language, and then covers a wide range of statistical techniques, including linear models and generalized linear models.

4. Wood, S.N. *Generalized Additive Models: An Introduction with* R, Second Edition, CRC.
   Excellent coverage of the material. I would highly recommend!

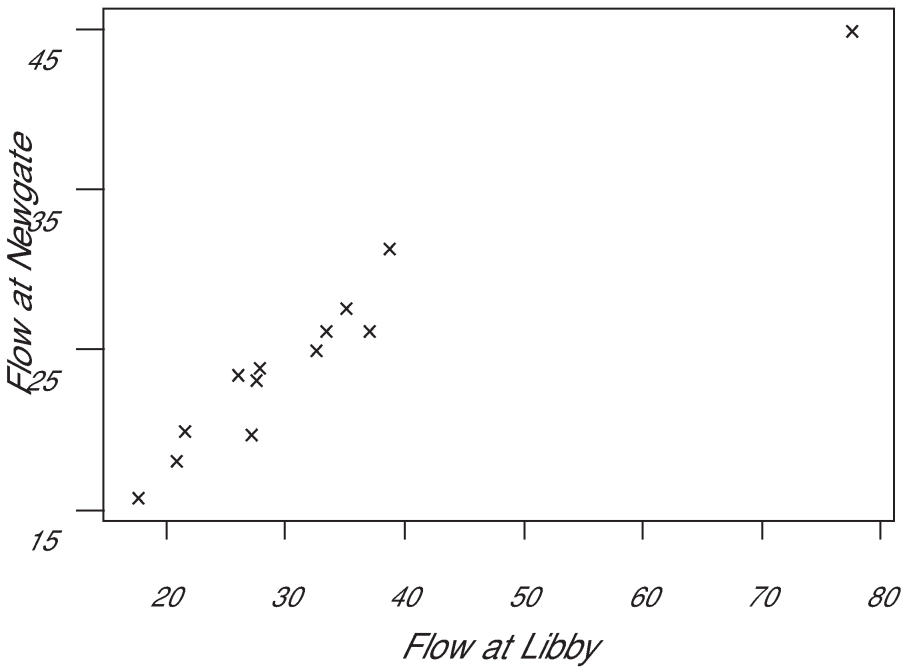
1 Statistical modelling: Relationships between variables

The following examples illustrate various aspects of regression analysis, as well as generalized linear modelling. All the sets of data described are available from Learn.

1.1 January flows on the Kootenai river

While planning a hydro-electric scheme on the Kootenai River in the USA, engineers examined records of the January river flow at Newgate, where a dam was to be built, and at Libby, a place about 50 miles downstream. The records for Libby went back 19 years but those for Newgate existed for only 13, so they wanted to use the earlier records at Libby to estimate the flow at Newgate for the six earlier years. The January flows at the two places are given below (in hundreds of cubic feet per second) and in the file Riverflow.txt: the plot shows the flows for the 13 later years.

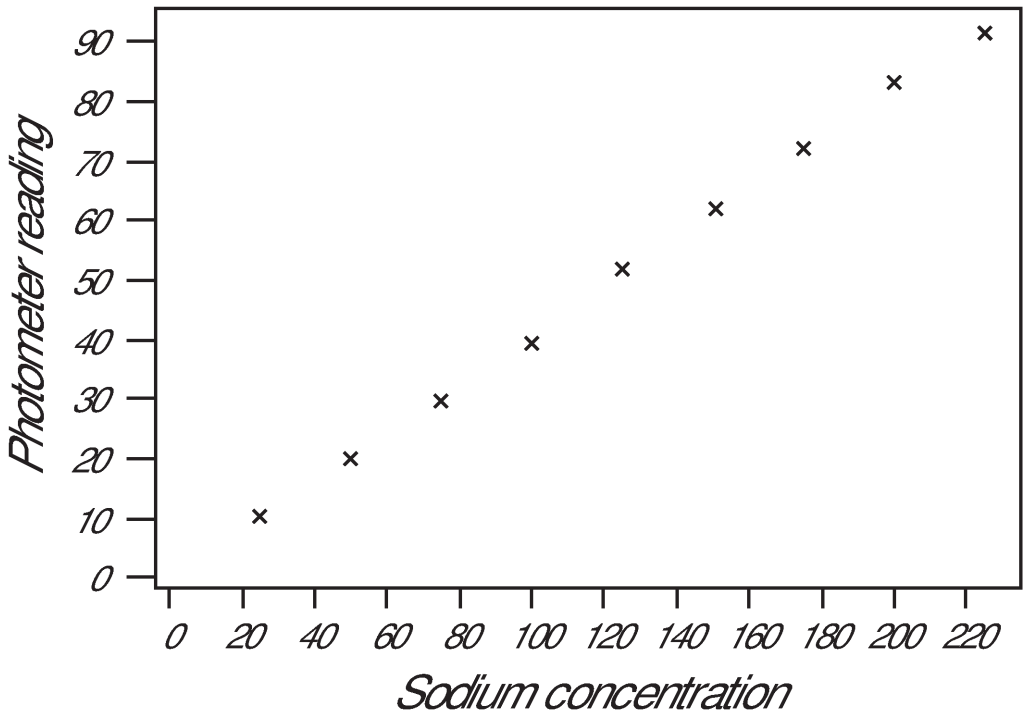
Newgate	–	–	–	–	–	–	19.7	18.0	26.1	44.9
Libby	42.0	24.0	38.0	49.4	24.6	24.2	27.1	20.9	33.4	77.6
Newgate	26.1	19.9	15.7	27.6	24.9	23.4	23.1	31.3	23.8	
Libby	37.0	21.6	17.6	35.1	32.6	26.0	27.6	38.7	27.8	



1.2 Calibration of a flame photometer

A flame photometer is an instrument that can be used to measure the concentration of sodium in chemical samples. It needs to be calibrated before analysing unknown concentrations, and so standard samples are tested first to provide a 'calibration curve' for converting further measurements into estimates of concentration. The results of such a calibration trial are given below and in Flame.txt.

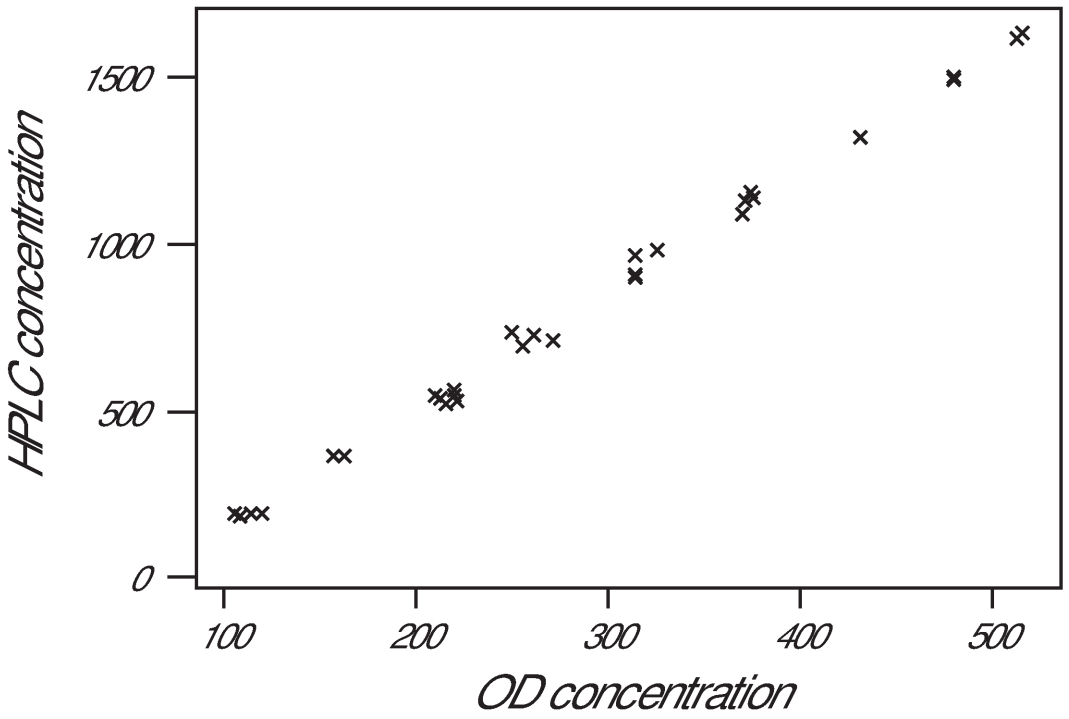
Sodium concentration	25	50	75	100	125	151	175	200	225
Photometer reading	10	20	29.5	39.5	52	62	72	83.5	91.5



1.3 Comparison of two methods for measuring the bitterness of beer

The bitterness of a beer is a function of the concentration of *iso-alpha acids*; the standard method of measuring this concentration is the optical density (OD) method. The research department of a brewery conducted an experiment to compare this method with a new and more convenient one based on high-pressure liquid chromatography (HPLC). Fifteen samples of beer were prepared with different degrees of bitterness intended to cover the range of interest, and each was split into two sub-samples. The concentration of iso-alpha acids in each sub-sample was measured using both methods: the results are given in Beer.txt and shown below. [For the sake of commercial confidentiality, linear transformations have been applied to the measurements.]

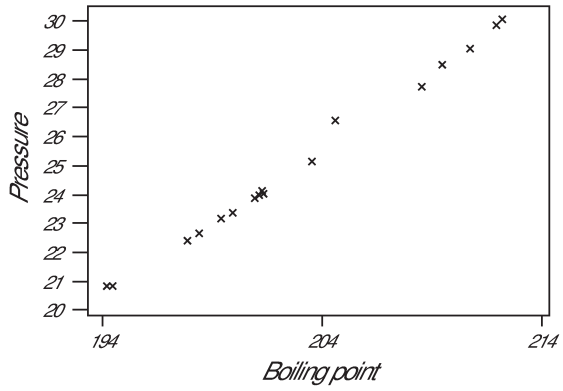
Sample number	1	1	2	2	3	3	4	4	5	5
HPLC reading	193.1	190.4	538.5	544.5	707.6	723.3	911.6	903.7	1153.5	1144.5
OD reading	115	121	214	210	272	262	314	314	374	376
Sample number	6	6	7	7	8	8	9	9	10	10
HPLC reading	1321.4	1319.0	1495.4	1500.4	1621.2	1634.2	181.9	191.4	366.1	360.8
OD reading	432	432	480	480	512	516	109	106	158	163
Sample number	11	11	12	12	13	13	14	14	15	15
HPLC reading	531.2	522.3	548.0	560.3	732.1	693.9	963.2	986.9	1135.0	1089.8
OD reading	222	216	220	220	250	256	314	326	372	370

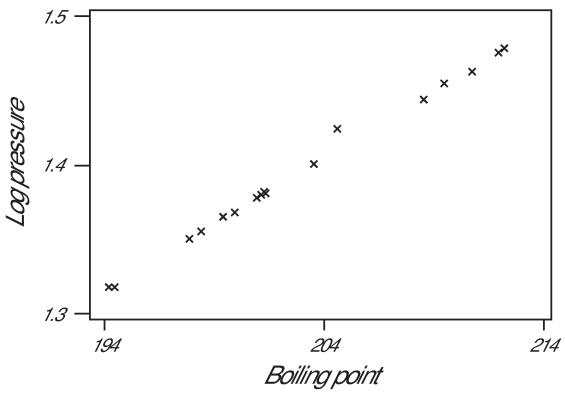


1.4 Forbes' data: Atmospheric pressure and the boiling point of water

In 1857 James Forbes, Professor of Natural Philosophy at the University of Edinburgh, published a paper entitled 'Further experiments and remarks on the measurement of heights by the boiling point of water' (*Transactions of the Royal Society of Edinburgh*, 21, 135–143). It was known that altitude could be estimated from atmospheric pressure, measured with a barometer, but barometers in those days were fragile and clumsy; he wanted to show that atmospheric pressure could be more conveniently measured by determining the boiling point of water and then using the relationship between boiling point and barometric pressure. The following values (also given in Forbes.txt) are the barometric pressures (in inches of mercury and adjusted for ambient air temperature) and the corresponding boiling points (in °F) recorded at 17 locations in Edinburgh and in the Alps. In the graphs which follow, the pressures and their logarithms (to base 10) are plotted against the boiling points: Forbes' theory suggested that the *logarithm* of pressure should be linearly related to boiling point.

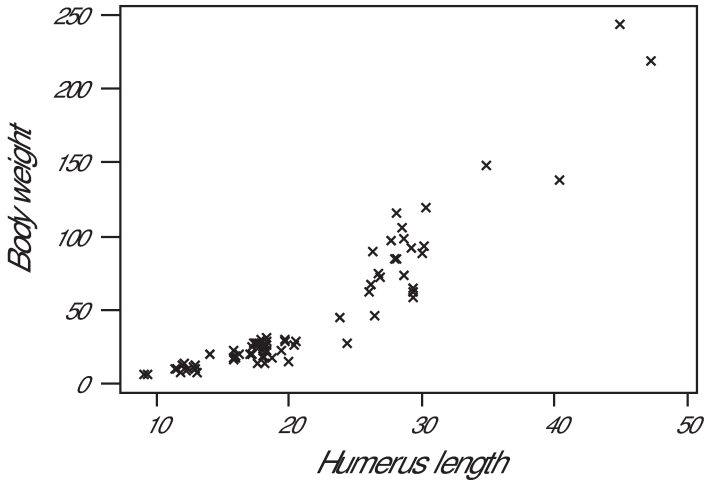
Pressure	20.79	20.79	22.40	22.67	23.15	23.35	23.89	23.99	24.02
Boiling point	194.50	194.25	197.90	198.43	199.45	199.95	200.93	201.15	201.35
Pressure	24.10	25.14	26.57	27.76	28.49	29.04	29.88	30.06	
Boiling point	201.30	203.55	204.60	208.57	209.47	210.72	211.95	212.18	

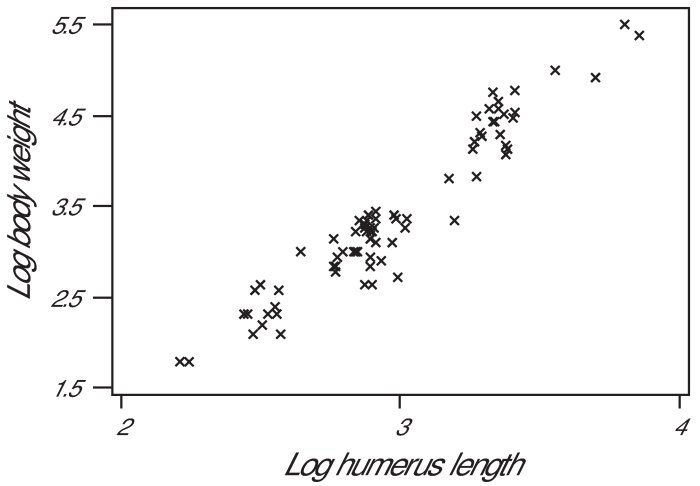




1.5 Estimating the weights of birds from the lengths of wing bones

A zoologist studying the diets of owls wanted to estimate the weight of birds caught and eaten by owls using measurements on birds' bones found in owl pellets. He collected and measured fresh specimens of 78 birds belonging to species he thought were typical of the prey of owls, ranging in size from goldcrests at about 5 grams to a magpie at 244 grams: the data are in BirdWt.txt. The first plot shows the fresh body weight (g) of each bird and length of its humerus bone (mm). The second plot is of the (natural) logarithms of these two measurements.



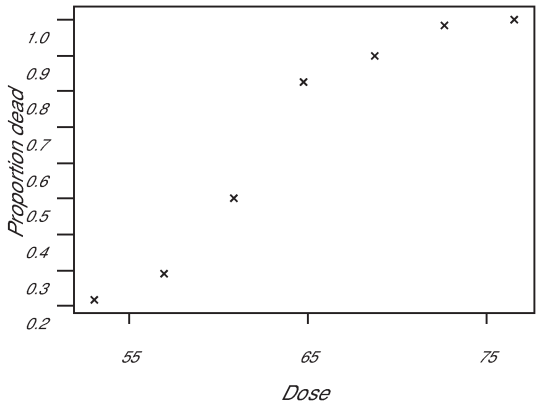


1.6 Numbers of beetles killed by different doses of insecticide

Seven doses of carbon disulphide, an insecticide, were applied to groups of beetles. The doses (in mg/l), the numbers of beetles receiving these doses, and the numbers found to be dead after 5 hours are given below and in Beetles.txt. The first plot, of the proportion dead against dose, shows that its slope increases and then decreases with increasing dose. The second plot, of the logit of this proportion against log dose, is closer to linearity.

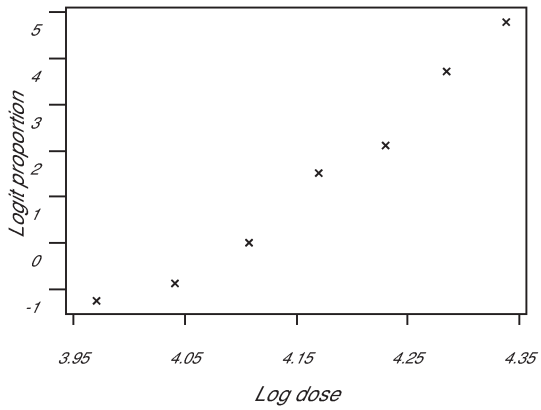
Dose	53.00	56.91	60.84	64.76	68.69	72.61	76.54
Number of beetles	60	62	56	63	59	62	60
Number dead	13	18	28	52	53	61	60

$$\texttt{Proportion dead} = \frac{\texttt{Number dead}}{\texttt{Number of beetles}}$$




```
logit = loge((No.  dead + 0.5)/(No.  in group - No.  dead + 0.5))
```

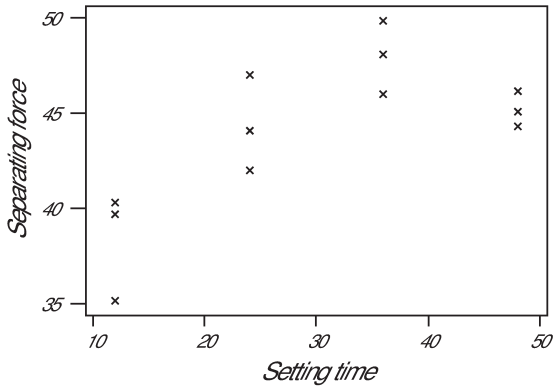
Plot of logit against log dose.



1.7 Effect of setting time on the strength of an adhesive

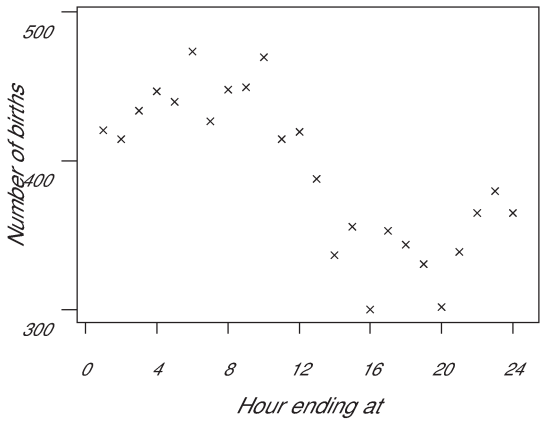
In a study of the setting strength of a woodwork adhesive, the following procedure was carried out. Adhesive was applied to two strips of wood. After 15 minutes the two strips were clamped together at right angles for a fixed time while the adhesive set. Then the force (in kg) required to separate the two strips was measured. Three pairs of pieces were used at each of four setting times with the results shown below (and given in Adhesive.txt).

Setting time	12	12	12	24	24	24	36	36	36	48	48	48
Separating force	35.1	39.7	40.3	42.0	44.1	47.0	46.0	48.1	49.9	44.3	45.1	46.2



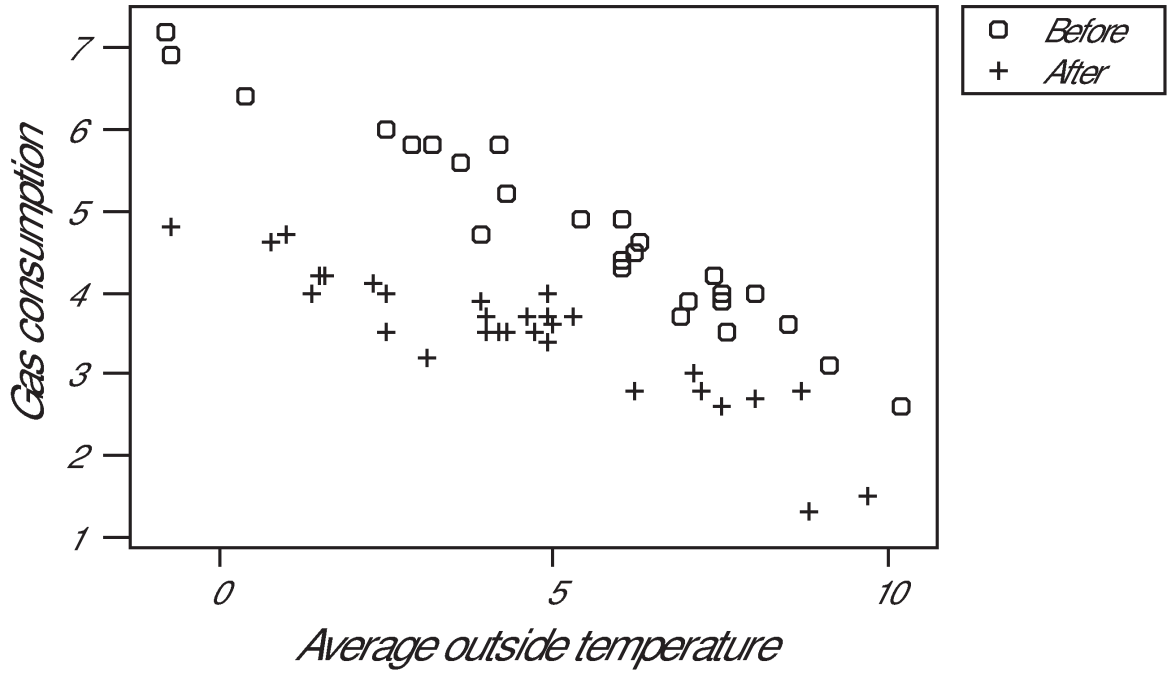
1.8 Numbers of hospital births and time of day

The numbers of normal human births in a hospital were recorded for each hour of the day over several years. The hours and numbers per hour are plotted below and given in the file Births.txt.



1.9 Domestic gas consumption with and without cavity-wall insulation

The average outside temperature (in °C) and the weekly gas consumption (in thousands of cubic feet) were recorded at a house in south-east England which had gas-fired central heating for 26 weeks before and 30 weeks after cavity-wall insulation was installed. The house thermostat was set to 20°C throughout the 56 weeks. The 56 pairs of values are given in Insulate.txt along with a column indicating which of them were recorded before and after installation.



1.10 Transformations

As we can see from some of the previous examples, transformations of the data are often necessary to obtain a linear relationship (or, at least, to improve linearity) between a response variable (y) and an explanatory variable (x).

As a further example, compare the following (logarithmic, to base 10) plot with the plot shown in Section 1.1.