**Data sets available for lack of independence:**

*ipomopsis.txt*

The example concerns an experiment on the impact of grazing on the seed production of a biennial plant. Forty plants were allocated to two treatments, grazed and ungrazed, and the grazed plants were exposed to rabbits during the first two weeks of stem elongation. They were then protected from subsequent grazing by the erection of a fence and allowed to regrow. Because initial plant size was thought likely to influence fruit production, the diameter of the top of the rootstock was measured before each plant was potted up. At the end of the growing season, the fruit production (dry weight in milligrams) was recorded on each of the 40 plants, and this forms the response variable in the analysis.

*Clams.txt*

A text file containing biomass (labelled as AFD which stands for ash free dry weight) of 398 wedge clams (*Donax hanleyanus*) and the length of the clams collect for six different months (Ieno, published data). The data were measured on a beach in Argentina in 1997. An initial scatter plot of the data showed a clear non-linear relationship, and therefore, both AFD and length were log-transformed to linearise the relationship to get the variables LNLENGTH and LNAFD. Due to different stages of the life cycle of wedge clams, the biomass-length relationship may change between months, especially before and after the spawning period (Spawning refers to the process of releasing the eggs and sperm into water by aquatic animals) in September–October and February–March.

*fertilizer.txt*

A common cause of temporal pseudoreplication in growth experiments with fixed effects is when each individual is measured several times as it grows during the course of an experiment. This data set as has a single fixed effect (a two-level categorical variable, with fertilizer added or not) and six replicate plants in each treatment, with each plant measured on five occasions (after 2, 4, 6, 8 or 10 weeks of growth). The response variable is root length, measured non-destructively through a glass panel, which is opened to the light only when the root length measurements are being taken.

*rats.txt*

Three experimental treatments were administered to rats, and the glycogen content of the rats’ livers was analysed as the response variable. There were two rats per treatment, so the total sample was n = 3 × 2 =6. After each rat was killed, its liver was cut up into three pieces: a left-hand bit, a central bit and a right-hand bit. So now there are six rats each producing three bits of liver, for a total of 6 × 3 = 18 numbers. Finally, two separate preparations were made from each macerated bit of liver, to assess the measurement error associated with the analytical machinery. At this point there are 2 × 18 = 36 numbers in the data frame as a whole.

*RIKZ.txt*

The data was collected by the Dutch institute RIKZ in the summer of 2002. In each inter-tidal area (denoted by ‘beach’), five samples were taken, and the macro-fauna and abiotic variables were measured. Zuur et al. (2007). Species richness (the number of different species) and NAP (the height of a sampling station compared to mean tidal level) and Exposure ( an index composed of the following elements: wave action, length of the surf zone, slope, grain size, and the depth of the anaerobic layer) were measured. The underlying question for these data is whether there is a relationship between species richness, exposure, and NAP.