1. Suppose that students answer questions on a test and that a specific student has an aptitude $T$. A particular question might have difficulty $d_i$ and the student will get the answer correct only if $T > d_i$. Now if we consider $d_i$ fixed and $T \sim N(\mu, \sigma^2)$, then the probability that a randomly selected student will get the answer wrong is $p_i = P(T < d_i)$.

Show how you might model this situation using a probit regression model.

2. The dataset discoveries lists the number of great scientific discoveries for the years 1860 to 1959, as chosen by “The World Almanac and Book of Facts”, 1975 Edition. Has the discovery rate remained constant over time?

To answer this question, fit a poisson regression model with a log link, and use the deviance to compare a null model with models including the year and year squared as predictors.

3. The ships dataset from the MASS package gives the number of damage incidents and aggregate months of service for different types of ships broken down by year of construction and period of operation. Load the dataset using the commands library(MASS) then data(ships).

Develop a model for the rate of incidents (i.e. a poisson regression model with log link), describing the effect of the important predictors.

4. The infert dataset from the survival package presents data from a study of infertility after spontaneous and induced abortion. Using a logistic regression model, analyse and report on the factors related to infertility based on this data. (Don’t use the factor stratum, as it is confounded with the other predictors.)

5. The dataset africa from the faraway package gives information about the number of military coups in sub-saharan Africa and various political and geographical information.

Use the AIC to choose a parsimonious generalised linear model for the number of coups. Give an interpretation of the effect on the response of the variables you include in your model.