MAST30027: Modern Applied Statistics

Week 3 Lab

1. The binomial random variable $Y \sim \text{bin}(m, p)$ for known m (not a parameter) has mass function:

$$f(y; p) = \binom{m}{y} p^y (1-p)^{m-y} \text{ for } y = 0, 1, \ldots, m.$$

   Show that the binomial distribution is an exponential family.

2. The `infert` dataset from the `survival` package presents data from a study of infertility after spontaneous and induced abortion. You can load the dataset using the following command.

```
library(survival)
data(infert)
?infert
str(infert)
```

   The response is `case`, with 1 indicating infertility and 0 fertility. The data comes from a case-control study, the aim of which was to estimate the effect of the number of prior induced and spontaneous abortions on the probability of becoming infertile. We will consider education, age and parity (something numeric, whatever it is) as other predictors.

   Fit a binomial regression model with `case` as a response variable and induced (the number of prior induced abortions), spontaneous (the number of prior spontaneous abortions), education, age and parity as predictors. Test whether the education is important to predict the probability of becoming infertile when all other predictors are in the model.

3. 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.

   Load the dataset using the command data(discoveries).