Generalised Regression Models

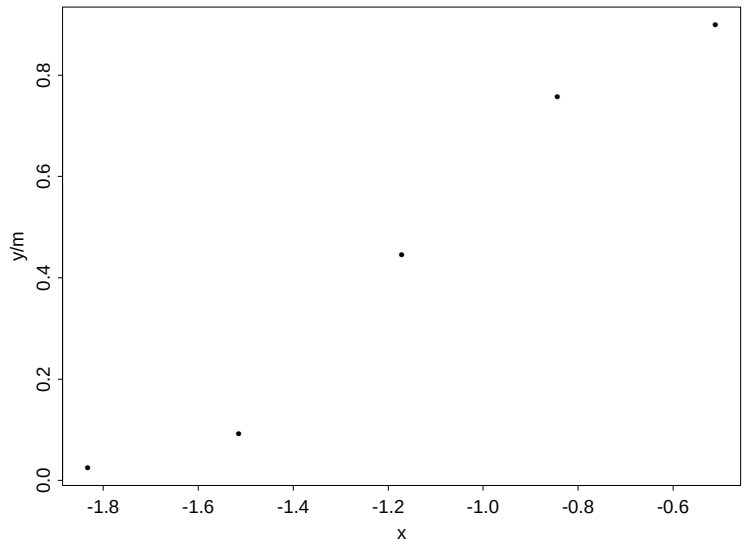
GRM: Example — Logistic Regression in R/S-PLUS **Semester 1, 2022–2023**

1 Weevil data set: binomial responses

Five doses of an insecticide (*Malathion*) were applied to granary weevils. For each dose (d_i), the number of insects (m_i) receiving that level of dose and the number killed (y_i) were recorded. Assume the data set is available in a file called `weevil.dat` (the file may be downloaded from the course webpage).

```
weevil.df <- read.table('weevil.dat',
                 col.names=c('d','m','y'))
weevil.df
     d   m   y
1 0.16 120   3
2 0.22 120  11
3 0.31 119  53
4 0.43 120  91
5 0.60 119 107

attach(weevil.df)

x <- log(d)
plot(x,y/m)
```

A plot of proportion killed y/m against the explanatory variable $x = \log(d)$ suggests a sigmoid shape.



2 Fitting the logistic regression model

To fit a logistic regression model the `glm` function can be used with `family=binomial(link=logit)`.

```
> weevil.glm <- glm( cbind(y,m-y) ~ x, family=binomial(logit))

> weevil.glm
Call:
glm(formula = cbind(y, m - y) ~ x, family = binomial(logit))

Coefficients:
 (Intercept)        x
    4.889407 4.538052

Degrees of Freedom: 5 Total; 3 Residual
Residual Deviance: 4.061521
```

Note that the formula is of the form `cbind(y, m - y) ~ x`, i.e., the 'response' is a two-column matrix with the number of successes in the first column and the number of failures in the second column.

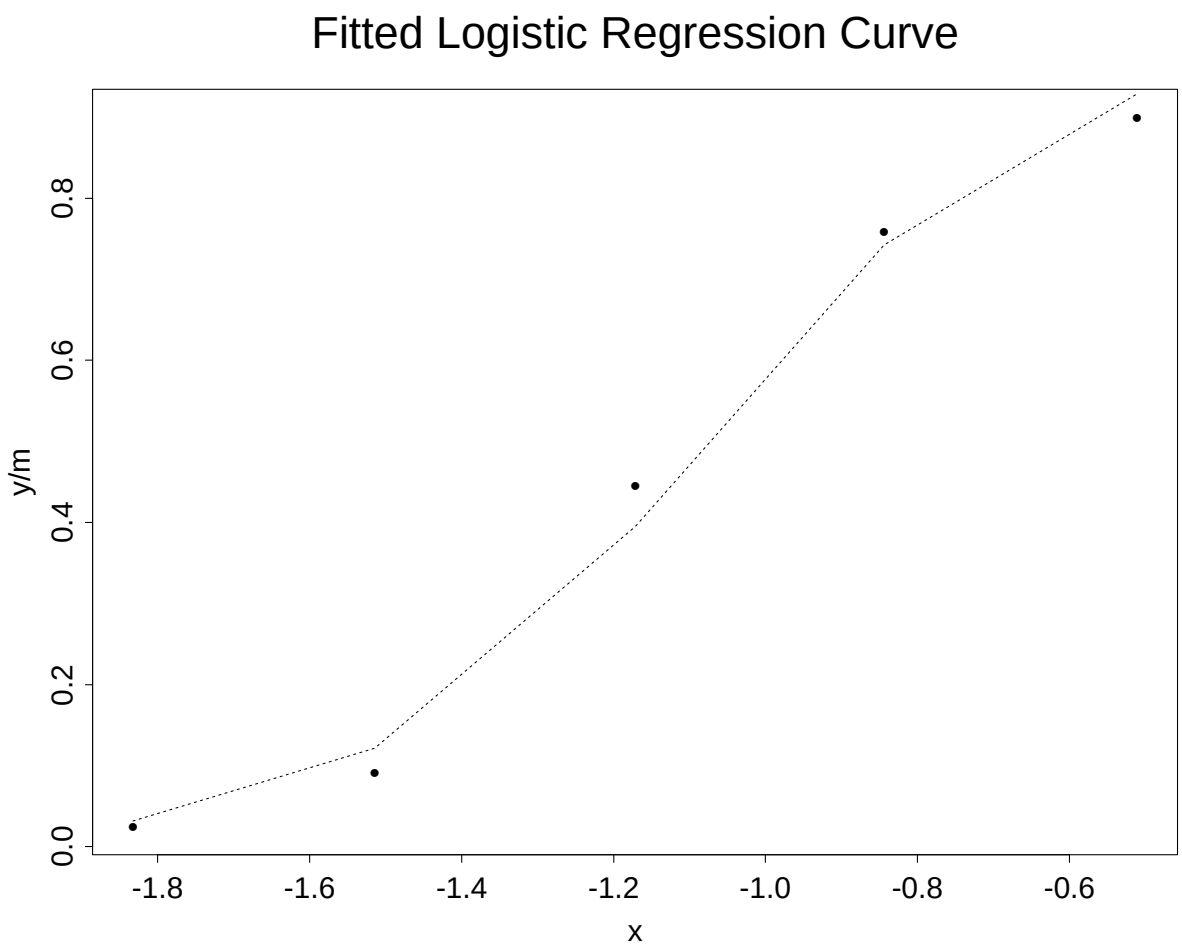
3 Extracting information from the `glm` object

We extract information about a `glm` object with generic functions such as `summary`.

```
> summary(weevil.glm)
Call: glm(formula = cbind(y, m - y) ~ x, family = binomial(logit))
Deviance Residuals:
          1         2        3        4         5
 -0.4203754 -1.028024 1.112377 0.397395 -1.196941

Coefficients:
               Value Std. Error  t value
(Intercept) 4.889407  0.3916909 12.48282
          x 4.538052  0.3424748 13.25076

(Dispersion Parameter for Binomial family taken to be 1 )

    Null Deviance: 345.594 on 4 degrees of freedom

Residual Deviance: 4.061521 on 3 degrees of freedom

Number of Fisher Scoring Iterations: 3

Correlation of Coefficients:
  (Intercept)
x 0.955965
```

4 Plotting the fitted values

```
fitted(weevil.glm)                          # fitted values
         1         2         3         4         5
 0.03146399 0.1211289 0.3952058 0.7425811 0.9289885


plot(x,y/m)

lines(x,fitted(weevil.glm),lty=2)
title('Fitted Logistic Regression Curve')
```



5 Residuals

The Pearson and deviance residuals may be obtained with the `residuals` function.

```
> residuals(weevil.glm,type="pearson")      # resids

         1          2         3         4         5
 -0.4055089 -0.9890373 1.119489 0.3946554 -1.266625



> residuals(weevil.glm,type="deviance")
         1         2         3         4         5
 -0.4203754 -1.028024 1.112377 0.397395 -1.196941

par(mfrow=c(2,3))
plot(weevil.glm)                            # diagnostic plots
```

6 Analysis of deviance

```
> anova(weevil.glm)                          # analysis of deviance
Analysis of Deviance Table

Binomial model

Response: cbind(y, m - y)

Terms added sequentially (first to last)
     Df Deviance Resid. Df Resid. Dev
NULL                     4   345.5940
   x  1 341.5325         3     4.0615
```