**Binomial GLM (Logistic Regression): Malformation example**

In a small study data was collected on the presence of a sex organ malformation on a infant to study how the mother's alcohol consumption affected it. The information collected on each subject is

$$y = \begin{cases} 1 & \text{present} \\ 0 & \text{absent} \end{cases} \qquad x = \text{ mother's average drinks per day}$$

The data is tabulated as follows

| Consumption | Consumption score ($x$) | Absent | Present |
|---|---|---|---|
| 0 | 0.0 | 17066 | 48 |
| < 1 | 0.5 | 14464 | 38 |
| 1-2 | 1.5 | 788 | 5 |
| 3-5 | 4.0 | 126 | 1 |
| $\geq 6$ | 7 | 37 | 1 |

```
> Alcohol=c(0,0.5,1.5,4,7)
> Absent=c(17066,14464,788,126,37)
> Present=c(48,38,5,1,1)
> cbind(Alcohol,Absent,Present)

     Alcohol Absent Present
[1,]     0.0  17066      48
[2,]     0.5  14464      38
[3,]     1.5    788       5
[4,]     4.0    126       1
[5,]     7.0     37       1

> malform.logit=glm(cbind(Present,Absent)~Alcohol,family=binomial(link=logit))
> summary(malform.logit)

Call:
glm(formula = cbind(Present, Absent) ~ Alcohol, family = binomial(link = logit))

Deviance Residuals:
      1        2        3        4        5
 0.5921  -0.8801   0.8865  -0.1449   0.1291

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -5.9605     0.1154 -51.637   <2e-16 ***
Alcohol       0.3166     0.1254   2.523   0.0116 *
---
Signif. codes:  0 Ś***Š 0.001 Ś**Š 0.01 Ś*Š 0.05 Ś.Š 0.1 Ś Š 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 6.2020  on 4  degrees of freedom
Residual deviance: 1.9487  on 3  degrees of freedom
AIC: 24.576

Number of Fisher Scoring iterations: 4
```

The estimated regression is

$$\log\left(\frac{\hat{\pi}(x)}{1 - \hat{\pi}(x)}\right) = -5.961 + 0.3166x$$

with the $\beta_1$ being significant and positive, implying that as alcohol consumption increases so does the probability of malformation. As $x \uparrow 1$ unit the odds of malformation (to no malformation) increase multiplicative by $e^{0.3166} = 1.3725$. That is, for a mother consuming alcohol, there is about a 37% increase in a chance of malformation compared to a mother consuming 1 unit less (within the range of the data).

A 50-50 probability of malformation occurs at $-(-5.961)/0.3166 = 18.82817$ drinks which is outside the range of our data.

To test the significance of multiple $\beta$'s (however we only have one $\beta$ corresponding to one predictor) we take the difference of the deviance of the reduced to the full model

```
> ts=malform.logit$null.deviance-malform.logit$deviance;ts
```

```
[1] 4.253277
```

```
> df=malform.logit$df.null-malform.logit$df.residual;df
```

```
[1] 1
```

```
> 1-pchisq(ts,df)
```

```
[1] 0.03917467
```

The observed and fitted probabilities are:

```
> tab=round(cbind(fitted.values(malform.logit),Present/(Present+Absent)),3)
> dimnames(tab)=list(c("0","0.5","1.5","4","7"),c("Fitted prob.","Observed prob."))
> tab
```

|     | Fitted prob. | Observed prob. |
|-----|--------------|----------------|
| 0   | 0.003        | 0.003          |
| 0.5 | 0.003        | 0.003          |
| 1.5 | 0.004        | 0.006          |
| 4   | 0.009        | 0.008          |
| 7   | 0.023        | 0.026          |

**Model for Binomial data**