6 Further Aspects of Model Fitting

6.1 Diagnostics and Model Checking

Apart from considering the analysis of deviance to check the adequacy of the model, we also need to look at various other diagnostics and plots. We need to check that:

(a) the variance-mean relationship is correct;

(b) the effects of covariables are additive on the scale defined by the link function;

(c) the model 'fits' and is as parsimonious as possible.

The basic tools we use are analogous to those used for multiple regression models. However, since generalized linear models are much richer, there are many more checks and diagnostics. We shall consider some here. Consult McCullagh and Nelder (1989) for a more extended discussion.

Let $\mathbf{W}$ be a diagonal matrix of size n, where its i-th diagonal element

$$w_{ii} = \left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2 \frac{1}{\mathrm{V}(\mu_i)} = \frac{1}{[g'(\mu_i)]^2} \frac{1}{\mathrm{V}(\mu_i)}$$

with $g(\mu_i) = \eta_i$, $i = 1, \ldots, n$. Recall that $\mathrm{var}(Y_i) = \phi\, b''(\theta_i) = \phi\, \mathrm{V}(\mu_i)$. [Note that the w_{ii} should not be confused with the w_i, the weights sometimes used with the scale parameter.]

The w_{ii} may be thought of as *intrinsic weights* associated with the observations y_i and arising via the variance-mean relationship and the link function.

Also, define

$$\mathbf{H} = \mathbf{W}^{1/2}\mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}^{1/2}$$

where $\mathbf{X}$ is the design matrix for our model. This is a weighted version of the hat matrix in multiple regression which has $\mathbf{W} = \mathbf{I}$.

Now let

- h_i be the i-th diagonal element of $\mathbf{H}$, using the estimated parameter vector $\hat{\boldsymbol{\beta}}$
- $\widehat{\mu}_i$ be the fitted value for y_i, corresponding to $\hat{\boldsymbol{\beta}}$.

6.1.1 Residuals

We consider a range of different types of residual. It should be noted that the names given to the different residuals are not consistent across textbooks and statistical packages, and it is important to check the definitions.

Pearson Residuals

$$r_i^P \;=\; \frac{y_i - \widehat{\mu}_i}{\sqrt{\mathrm{V}(\widehat{\mu}_i)}} \qquad \left[\text{ or } \frac{y_i - \widehat{\mu}_i}{\sqrt{\mathrm{V}(\widehat{\mu}_i)/w_i}} \text{ if weights are used.} \right]$$

If we have Poisson data then

$$r_i^P \;=\; \frac{y_i - \widehat{\mu}_i}{\sqrt{\widehat{\mu}_i}}.$$

If we have Normal data then

$$r_i^P \;=\; y_i - \widehat{\mu}_i.$$

Raw Residuals

$$r_i \;=\; y_i - \widehat{\mu}_i$$

Deviance Residuals

$$r_i^D \;=\; \mathrm{sgn}(y_i - \widehat{\mu}_i)\sqrt{d_i},$$

where the deviance D is given by

$$D \;=\; \sum_{i=1}^{n} d_i$$

and $\mathrm{sgn}(x)$, for $x \in \mathbb{R}$, is just the sign of x.

r_i^D identifies whether the i-th observation makes a disproportionate contribution to the deviance. The deviance residuals are the residuals that generally behave most like residuals from a Normal model with identity link.

Standardized Residuals

There are various standardized versions of the residuals so far considered. We consider:

$$r_i^{PS} \;=\; \frac{r_i^P}{\sqrt{1-h_i}}\,, \qquad r_i^{DS} \;=\; \frac{r_i^D}{\sqrt{1-h_i}}$$

$$r_i^{PS'} \;=\; \frac{r_i^P}{\sqrt{\hat{\phi}(1-h_i)}}\,, \qquad r_i^{DS'} \;=\; \frac{r_i^D}{\sqrt{\hat{\phi}(1-h_i)}}$$

Deletion Residuals

The exact deletion residuals are

$$y_i - \hat{\mu}_i^{(i)}$$

where

$$\hat{\mu}_i^{(i)} \;=\; g^{-1}\left(\mathbf{x}_i'\,\hat{\boldsymbol{\beta}}^{(i)}\right)$$

and $\hat{\boldsymbol{\beta}}^{(i)}$ is the estimated parameter vector if observation i is omitted.

The deletion residuals can be calculated exactly using a weights vector consisting of ones, and changing the w_i from one to zero to omit cases in turn. This is computationally demanding and there are approximations available.

6.1.2 Residual Plots

It is generally recommended that the r_i^D or r_i^{DS} are used in residual plots. We are usually interested in pattern rather than size in the plots, so the estimated scale parameter $\hat{\phi}$ may be omitted. Plots may include the following.

(a) The r_i^D or r_i^{DS} can be plotted against some function of the fitted values such as $g(\hat{\mu}_i)$, or the variance-stabilizing transform of $\hat{\mu}_i$.

$$\begin{array}{cc} \hat{\mu}_i & \text{for Normal errors} \\ \sqrt{\hat{\mu}_i} & \text{for Poisson errors} \\ \sin^{-1}\sqrt{\hat{\mu}_i} & \text{for Binomial errors} \\ \log\hat{\mu}_i & \text{for Gamma errors} \end{array}$$

If all is well we expect the residuals to have zero mean, to be randomly positive or negative, and to have constant range as the function of $\hat{\mu}_i$ varies.

Systematic deviations may take various forms

(i) There may appear to be a curved relationship, perhaps due to the wrong link function, the omission of an interaction term or essential covariate, the need to transform a covariate etc.

(ii) There may be a systematic change in the range, which indicates an incorrect variance-mean relationship or distribution.

(b) r_i^D or r_i^{DS} can be plotted against covariables present in the linear predictor. The null pattern is the same as for (a). Once again a systematic trend may indicate the wrong link function, a missing interaction or power, or the need to transform the covariable.

(c) r_i^D or r_i^{DS} can be plotted against covariables not already in the model.

(d) Normal plots may be produced for the r_i^{DS}.

6.1.3 Influence and Leverage

Leverage
Recall that h_i is a measure of the distance between $\mathbf{x}_i$ and $\overline{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i$ and is used as

a measure of leverage. We have $h_i \in (0,1)$, and any value greater than $2p/n$ should be investigated. Another measure of leverage, v_i, based on h_i, is sometimes used.

$$v_i = \frac{h_i}{1-h_i}, \quad \text{which} \in (0,\infty)$$

Influence
Cook's distance is a measure of the distance between $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}^{(i)}$, and is therefore a measure of influence. It may be expressed as

$$C_i = \frac{\left(r_i^{PS'}\right)^2}{p} \frac{h_i}{1-h_i} \quad i = 1, \ldots, n$$

and so depends on both the standardized Pearson residual and on the leverage. Values greater than 1 should be investigated.

6.2 Example: Toxoplasmosis Data

Let us return to the toxoplasmosis data of chapter 3. Although we concluded that the third-order polynomial provided a significantly better fit than having just the constant in the linear predictor, its (scaled) deviance was very significant on the χ^2_{30} distribution, and so not a good fit to the data.

```
> summary(toxoplas.glm)

Call:
glm(formula = prop ~ rain.cor + I(rain.cor^2) + I(rain.cor^3),
    family = binomial, data = toxoplas, weights = no.tested)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.7620  -1.2166  -0.5079   0.3538   2.6204

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    0.09939    0.10197   0.975 0.329678
rain.cor      -2.55187    0.88276  -2.891 0.003843 **
I(rain.cor^2) -6.06369    2.96348  -2.046 0.040743 *
I(rain.cor^3) 39.32248   11.73606   3.351 0.000806 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 74.212  on 33  degrees of freedom
Residual deviance: 62.635  on 30  degrees of freedom
```

```
AIC: 161.33

Number of Fisher Scoring iterations: 3

> pchisq(62.635, 30, lower.tail=F)
[1] 0.0004365815
```

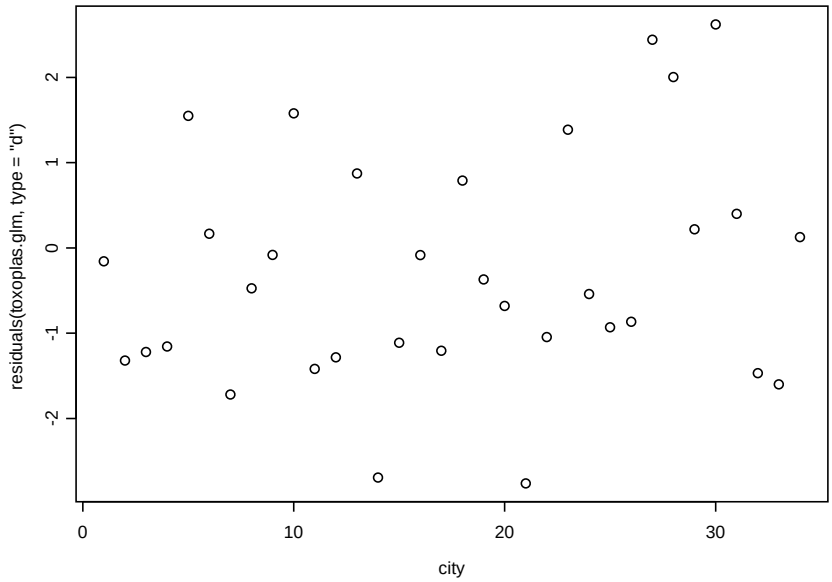
It may be worth computing some of the residuals, identifying any 'problematic' observations, removing them, and then re-fitting the model.

We've rounded the 3 sets of residuals to 3 decimal places. Note that for this Binomial error structure, the dependent variables are the proportions in each city testing positive, and so the fitted values (which can be obtained using `fitted`) are the estimated probabilities of testing positive in each city.

Cities 14, 21, 27, and 30 all appear to have (relatively) high deviance and Pearson residuals. These can clearly be seen on an index plot of (deviance) residuals vs city number.

```
> round(cbind(resid(toxoplas.glm, type = "response"),
+ resid(toxoplas.glm, type = "pearson"),
+ resid(toxoplas.glm, type = "deviance")), 3)

     [,1]   [,2]   [,3]
1  -0.039 -0.156 -0.156
2  -0.206 -1.302 -1.320
3  -0.261 -1.171 -1.220
4  -0.180 -1.137 -1.156
5   0.451  1.282  1.549
6   0.037  0.165  0.166
7  -0.299 -1.699 -1.719
8  -0.053 -0.471 -0.474
9  -0.017 -0.083 -0.083
10  0.237  1.509  1.577
11 -0.141 -1.390 -1.419
12 -0.560 -1.129 -1.282
13  0.079  0.880  0.874
14 -0.273 -2.567 -2.694
15 -0.461 -0.925 -1.112
16 -0.013 -0.084 -0.084
17 -0.517 -1.034 -1.206
18  0.053  0.786  0.790
19 -0.061 -0.369 -0.369
20 -0.075 -0.667 -0.680
21 -0.385 -2.685 -2.762
22 -0.421 -0.852 -1.045
23  0.204  1.357  1.387
24 -0.031 -0.543 -0.541
25 -0.065 -0.931 -0.930
26 -0.108 -0.869 -0.867
27  0.134  2.460  2.442
28  0.275  2.015  2.004
29  0.017  0.219  0.219
30  0.147  2.568  2.620
31  0.055  0.397  0.399
32 -0.230 -1.455 -1.469
33 -0.311 -1.524 -1.601
34  0.010  0.127  0.127

sum(residuals(toxoplas.glm, type = "deviance")^2)
[1] 62.6346
```

```
> plot(city, residuals(toxoplas.glm, type = "d"))
```



We can re-run the analysis with the aforementioned cities removed by using the `weights` options.

```
> wts <- no.tested
> wts[14] <- 0
> wts[21] <- 0
> wts[27] <- 0
> wts[30] <- 0

> toxoplasdelete.glm <- update(toxoplas.glm, weights = wts)
> summary(toxoplasdelete.glm)

Call:
glm(formula = prop ~ rain.cor + I(rain.cor^2) + I(rain.cor^3),
    family = binomial, data = toxoplas, weights = wts)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.7100  -0.9053  -0.1943   0.2379   2.5025

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    -0.1211     0.1211  -1.001 0.316983
rain.cor       -3.7197     1.2213  -3.046 0.002322 **
I(rain.cor^2)  -3.3925     3.1574  -1.074 0.282626
I(rain.cor^3)  48.7321    14.4588   3.370 0.000751 ***
---
```

```
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 44.616  on 29  degrees of freedom
Residual deviance: 30.909  on 26  degrees of freedom
AIC: 114.66

Number of Fisher Scoring iterations: 3
```

This time we end up with a deviance of 30.909 on 26 degrees of freedom, which is adequate, ($p = 0.232$).

6.3 Example: Dose-Response Data

Recall, from Chapter 2, the data we analysed showing the numbers of insects that have died after several hours exposure to a certain pesticide.

Dose x_i $\log_{10} CS_2$ mgl^{-1}	Number of insects, n_i	Number killed, y_i
1.6907	59	6
1.7242	60	13
1.7552	62	18
1.7842	56	28
1.8113	63	52
1.8369	59	53
1.8610	62	61
1.8839	60	60

Independence between observations, Binomial error structure, logit link, and linear predictor given by

$$\eta_i = \beta_1 + \beta_2 x_i$$

was assumed and the model fitted in R; an abbreviated version of the output is presented below. (Recall that `dose` was given a corrected value by subtracting 1.78.)

Logit link: $\log\left(\frac{\pi_i}{1-\pi_1}\right) = \beta_1 + \beta_2 x_i$

```
> pesticide.glm <- glm(prop ~ dose, family = binomial(link = "logit"),
    weights = number, data = pesticide)
> summary(pesticide.glm)

Call:
glm(formula = prop ~ dose, family = binomial(link = "logit"),
    data = pesticide, weights = number)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
```

```
-1.5941  -0.3944   0.8329   1.2592   1.5940

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   0.2837     0.1308   2.168   0.0301 *
dose         34.2703     2.9121  11.768   <2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 284.202  on 7  degrees of freedom
Residual deviance:  11.232  on 6  degrees of freedom
AIC: 41.43

Number of Fisher Scoring iterations: 4
```

This has a (scaled) deviance of 11.23 on 6 degrees of freedom, which is just about acceptable, ($p = 0.082$).

Let us look at some residual plots and different link functions though. We shall try the probit link and the complementary log-log link, and plot the deviance residuals against $\sin^{-1}\sqrt{\hat{\mu}_i}$ for each model.

First, we must fit the *probit* and *complementary log-log* models.

Probit link: $\Phi^{-1}(\pi_i) = \beta_1 + \beta_2 x_i$

```
> pesticide2.glm <- glm(prop ~ dose,
                        family = binomial(link = probit),
                        weights = number, data = pesticide)
> summary(pesticide2.glm)

Call:
glm(formula = prop ~ dose, family = binomial(link = probit),
    data = pesticide, weights = number)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.5714  -0.4703   0.7501   1.0632   1.3449

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.18046    0.07386   2.443   0.0146 *
dose        19.72794    1.48721  13.265   <2e-16 ***
---
```

```
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 284.20  on 7  degrees of freedom
Residual deviance:  10.12  on 6  degrees of freedom
AIC: 40.318

Number of Fisher Scoring iterations: 4
```

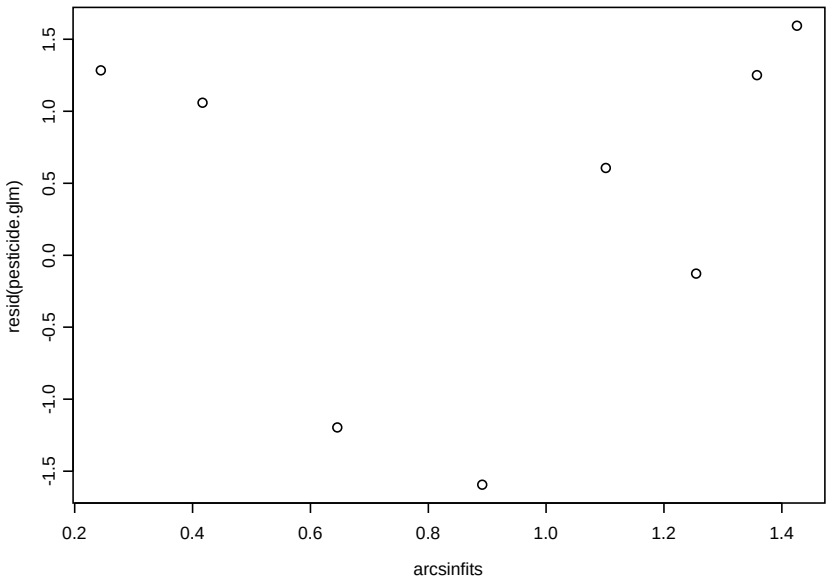
Complementary log-log link: $\log[-\log(1-\pi_i)] = \beta_1 + \beta_2 x_i$

```
> pesticide3.glm <- glm(prop ~ dose, family = binomial(link = cloglog),
    weights = number, data = pesticide)
> summary(pesticide3.glm)

Call:
glm(formula = prop ~ dose, family = binomial(link = cloglog),
    data = pesticide, weights = number)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-0.80329  -0.55135   0.03089   0.38315   1.28883

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.33903    0.08735  -3.881 0.000104 ***
dose        22.04117    1.79936  12.249  < 2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 284.2024  on 7  degrees of freedom
Residual deviance:   3.4464  on 6  degrees of freedom
AIC: 33.644

Number of Fisher Scoring iterations: 4
```

Logit link: $\log\left(\frac{\pi_i}{1-\pi_1}\right) = \beta_1 + \beta_2 x_i$

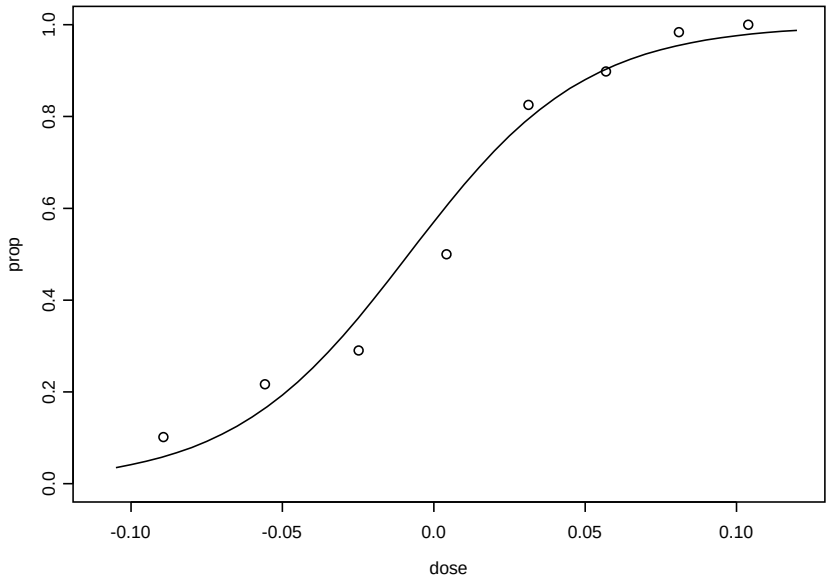
(Residual plot)

```
> arcsinfits <- asin(sqrt(fitted(pesticide.glm)))
> plot(arcsinfits, resid(pesticide.glm))
```



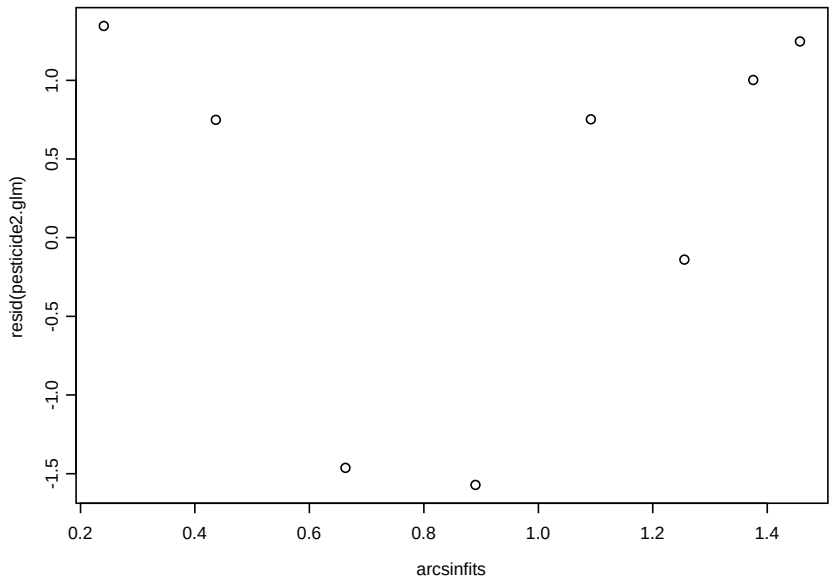
(Fitted model)

```
> x <- seq(-0.105, 0.12, 0.005)
> plot(dose, prop, xlim = c(-0.11, 0.12), ylim = c(0, 1))
lines(x,predict(pesticide.glm,data.frame(dose=x),type="response"))
```

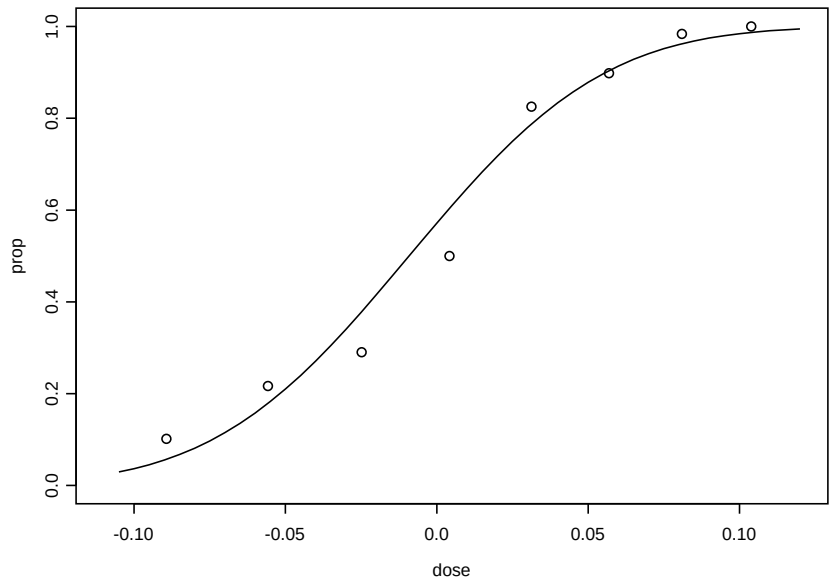


Probit link: $\Phi^{-1}(\pi_i) = \beta_1 + \beta_2 x_i$

```
> arcsinfits <- asin(sqrt(fitted(pesticide2.glm)))
> plot(arcsinfits, resid(pesticide2.glm))
```

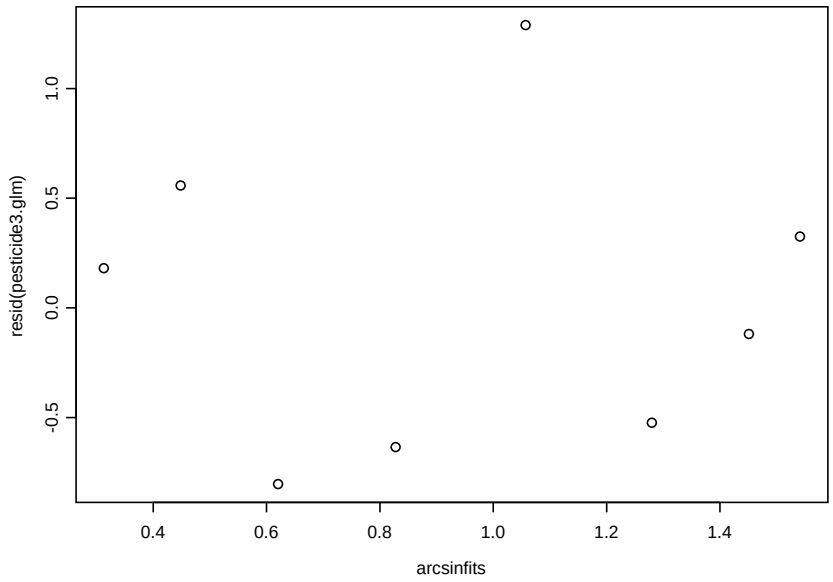



```
plot(dose,prop,xlim=c(-0.11, 0.12),ylim=c(0,1))
lines(x,predict(pesticide2.glm,data.frame(dose=x),type="response"))
```

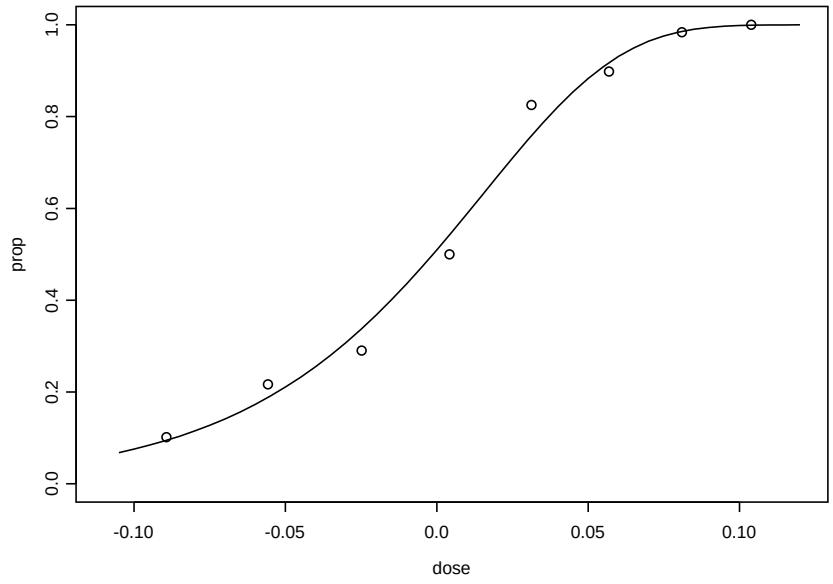


Complementary log-log link: $\log[-\log(1-\pi_i)] = \beta_1 + \beta_2 x_i$

```
> arcsinfits <- asin(sqrt(fitted(pesticide3.glm)))
> plot(arcsinfits, resid(pesticide3.glm))
```




```
plot(dose,prop,xlim=c(-0.11, 0.12),ylim=c(0,1))
lines(x,predict(pesticide3.glm,data.frame(dose=x),type="response"))
```



Clearly the complementary log-log function gives a much more satisfactory residual deviance. The plots show obvious curvature for both the first two link functions, and a satisfactory 'null' plot for the third.