The Scaled Deviance

It's a fact that for any model M

$$\ell_M \leq \ell_S$$

Definition

The scaled deviance of model M is

$$SD_M = 2(\ell_S - \ell_M)$$

Scaled is sometimes just called deviance

We use scaled to compare two models using a χ^2 test.

Example

$Y_i \sim N(\mu_i, \sigma^2)$

$$\ell_M(\mu_i, \sigma^2; \underline{y}) = -\frac{n}{2}\ln(2\pi\sigma^2) - \sum_{i=1}^{n} \frac{(y_i - \hat{\mu}_i)^2}{2\sigma^2}$$

For the saturated model,

$$\hat{\mu}_i = y_i$$

$$l_S(\mu_i, \sigma^2; \underline{y}) = -\frac{n}{2}\ln(2\pi\sigma^2)$$

$$SD_M = \sum_{i=1}^{n} \frac{(y_i - \hat{\mu}_i)}{\sigma^2}$$

Under the null model, which has one parameter β_0, $\hat{\mu}_i = \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ for each i

and so

$$SD_0 = \chi^2_{n-1}$$

null model

In general $SD_M \sim \chi^2_{n-p}$

where n is the number of data

and p is the number of parameters in M.

$n-p$ is called the degrees of freedom.

Using Scaled Deviance to compare between two models

You could say that the smaller the deviance, the better is the fit. However, the fit can be too good, as for the saturated model. In that case the model loses its predictive value.

~~To check, we~~

To check whether the deviance is significantly smaller, we use the χ^2 test.

We can carry out this test in R by

anova(model1, model2, test="Chi")

You should only do this with two nested models, where the linear predictor ~~extends~~ of model 2 extends the one for model 1, and the distributions and links are the same.

$\beta_0 + \beta_1 x$ and $\beta_0 + \beta_1 x + \beta_2 x^2$ are nested

$\beta_0 + \beta_1 x + \beta_2 x^2$ and $\beta_0 + \beta_1 x + \beta_2 \ln x$

are not nested.

A second way of comparing two models is using Akaikes Information Criterion (AIC).

It is defined by

$$AIC = -\ell_m + 2p$$

where p = number of parameters.

The smaller the AIC is, the better is the model. Checking the AIC is equivalent to the χ^2 test when $5 \leq p \leq 15$.

In R, the AIC is displayed using

Summary(model)

Processes for choosing the best model

1. Forward

Start with null model, $\eta = \beta_0$, $\gamma \sim 1$

Add the covariate (to first power) which either reduces the AIC most or causes the most significant decrease in deviance. Continue trying to add covariates to the first power. If they don't reduce AIC or pass the χ^2 test, then we should not add them.

Then we can try adding interaction terms (e.g. $x_1 x_2$) or quadratic terms (x_i^2).

We should add linear terms before interactions.

Do this until adding more terms increases AIC or does not decrease deviance significantly.

In the end, all the coefficients should be significant.

Backwards Method

Start with all available covariates and their interaction. Remove interactions, removing the ones that decrease AIC most or decrease deviance most significantly. Continue in this way going on to covariates to the first power. We stop when there is no way to remove an interaction or covariate and decrease AIC or decrease deviance significantly.

All coefficients should be significant.

Residuals

We should check the residuals of our model.

The basic residuals are

$$y_i - \hat{\mu}_i$$

In R we can find the $\hat{\mu}_i$ by

fitted(model).

We wont be using basic residuals.

The pearson ~~RES~~ residuals are defined to be

$$\frac{y_i - \hat{\mu}_i}{\sqrt{Var(\hat{\mu}_i)}}$$

$$\frac{y_i - \mu_i}{\sqrt{Var(\mu_i)}}$$

Pearson residuals are used for normal data, but may be skewed for nonnormal data. The command for finding them in R is

residuals(model, type="pearson")

Deviance Residuals

We use these residuals. Deviance residuals are defined by $\text{sign}(y_i - \hat{\mu}_i)d_i$

where scaled deviance equals

$$SD_m = \sum_{i=1}^{n} d_i^2,$$

where $d_i \geq 0$ depends only on y_i and μ_i

and $\text{sign}(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ -1 & \text{if } x < 0 \end{cases}$

Example

$Y_i \sim N(\mu_i, \sigma^2)$ σ known

Pearson residuals are $\dfrac{y_i - \hat{\mu}_i}{\sigma}$

The scaled deviance is

$$\sum_{i=1}^{n} \left(\frac{y_i - \hat{\mu}_i}{\sigma} \right)^2 \quad \text{so} \quad d_i = \left| \frac{y_i - \hat{\mu}_i}{\sigma} \right|$$

The deviance residuals are

$$\text{sign}(y_i - \hat{\mu}_i) \left| \frac{y_i - \hat{\mu}_i}{\sigma} \right|$$

$$= \frac{y_i - \mu_i}{\sigma}$$

same as Pearson residuals.

Using the residuals to check the fit

the deviance residuals should show no pattern and not depend on the fitted values.

The R command to display residuals is

plot(model)

or plot(model, 1).

Estimate the Response Variable

Given new data, we can calculate the value of the linear predictor η and then find the estimated response

$$\mu = g^{-1}(\eta)$$

where g is the link.

You can do this manually, or type

Predict(model, newdata, type="response")

where newdata is a data frame containing the new values of the variables.

If type="response" is omitted, then the fitted values to the original data is displayed.