Introduction to General and Generalized Linear Models

Generalized Linear Models - part II

Henrik Madsen
Poul Thyregod
Jan Kloppenborg Møller

Informatics and Mathematical Modelling
Technical University of Denmark
DK-2800 Kgs. Lyngby

March 9, 2012

Today

- The generalized linear model
    - Link function
    - Estimation
    - Fitted values
    - Residuals
- Likelihood ratio test

The Generalized Linear Model

Definition (The generalized linear model)

Assume that $Y_1, Y_2, \ldots, Y_n$ are mutually independent, and the density can be described by an exponential dispersion model with the same variance function $V(\mu)$.

A *generalized linear model* for $Y_1, Y_2, \ldots, Y_n$ describes an affine hypothesis for $\eta_1, \eta_2, \ldots, \eta_n$, where

$$\eta_i = g(\mu_i)$$

is a transformation of the mean values $\mu_1, \mu_2, \ldots, \mu_n$.

The hypothesis is of the form

$$\mathcal{H}_0 : \boldsymbol{\eta} - \boldsymbol{\eta}_0 \in L,$$

where L is a linear subspace $\mathbb{R}^n$ of dimension k, and where $\boldsymbol{\eta}_0$ denotes a vector of *known off-set values*.

GLM vs GLM

General linear models	Generalized linear models
Normal distribution	Exponential dispersion family
Mean value linear	Function of mean value linear
Independent observations	Independent observations
Same variance	Variance function of mean
Easy to apply	Almost as easy to apply
Exact results	Approximate results

Dimension and design matrix

Definition (Dimension of the generalized linear model)

The dimension k of the subspace L for the generalized linear model is the *dimension of the model*

Definition (Design matrix for the generalized linear model)

Consider the linear subspace $L = \text{span}\{x_1, \ldots, x_k\}$, i.e. the subspace is spanned by k vectors $(k < n)$, such that the hypothesis can be written

$$\boldsymbol{\eta} - \boldsymbol{\eta}_0 = \boldsymbol{X}\boldsymbol{\beta} \text{ with } \boldsymbol{\beta} \in \mathbb{R}^k,$$

where $\boldsymbol{X}$ has full rank. The $n \times k$ matrix $\boldsymbol{X}$ is called the *design matrix*. The i^{th} row of the design matrix is given by the *model vector*

$$\boldsymbol{x}_i = \begin{pmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{ik} \end{pmatrix},$$

for the i^{th} observation.

The link function

Definition (The link function)

The *link function*, $g(\cdot)$ describes the relation between the linear predictor η_i and the mean value parameter $\mu_i = \mathrm{E}[Y_i]$. The relation is

$$\eta_i = g(\mu_i)$$

The inverse mapping $g^{-1}(\cdot)$ thus expresses the mean value μ as a function of the linear predictor η:

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

that is

$$\mu_i = g^{-1}(\boldsymbol{x}_i{}^T\boldsymbol{\beta}) = g^{-1}\left(\sum_j x_{ij}\beta_j\right)$$

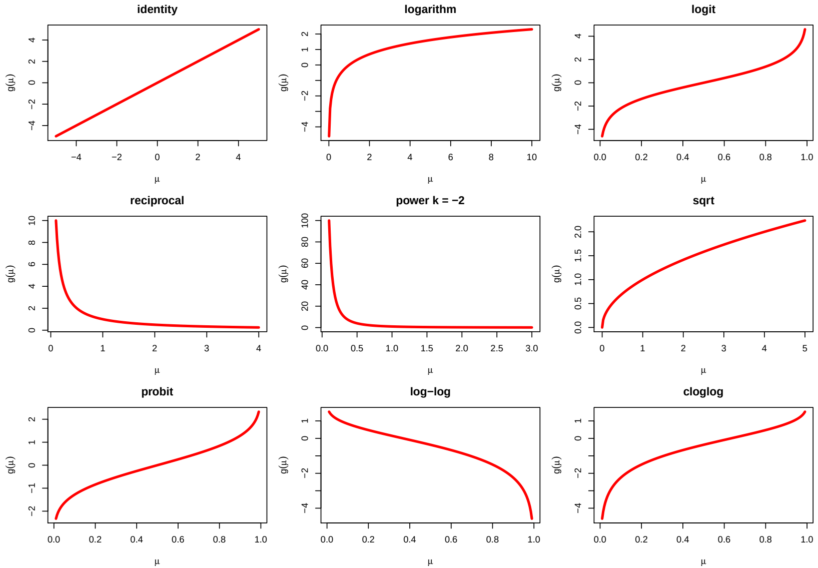
Link functions

The most commonly used link functions, $\eta = g(\mu)$, are :

Name	Link function $\eta = g(\mu)$	$\mu = g^{-1}(\eta)$
identity	μ	η
log	$\log(\mu)$	$\exp(\eta)$
logit	$\log(\mu/(1-\mu))$	$\exp(\eta)/[1+\exp(\eta)]$
inverse	$1/\mu$	$1/\eta$
power	μ^k	$\eta^{1/k}$
sqrt	$\sqrt{\mu}$	η^2
probit	$\Phi^{-1}(\mu)$	$\Phi(\eta)$
log-log	$\log(-\log(\mu))$	$\exp(-\exp(\eta))$
cloglog	$\log(-\log(1-\mu))$	$1-\exp(-\exp(\eta))$

Table: Commonly used link function.

Link functions



The canonical link

The canonical link is the function which transforms the mean to the canonical location parameter of the exponential dispersion family, i.e. it is the function for which $g(\mu) = \theta$. The canonical link function for the most widely considered densities are

Density	Link: $\eta = g(\mu)$	Name
Normal	$\eta = \mu$	identity
Poisson	$\eta = \log(\mu)$	log
Binomial	$\eta = \log[\mu/(1-\mu)]$	logit
Gamma	$\eta = 1/\mu$	inverse
Inverse Gauss	$\eta = 1/\mu^2$	`1/mu^2`

Table: Canonical link functions for some widely used densities.

Specification of a generalized linear model

a) Distribution / Variance function:
Specification of the distribution – or the *variance function* $V(\mu)$.

b) Link function:
Specification of the *link function* $g(\cdot)$, which describes a function of the mean value which can be described linearly by the explanatory variables.

c) Linear predictor:
Specification of the linear dependency

$$g(\mu_i) = \eta_i = (\boldsymbol{x}_i)^T \boldsymbol{\beta}.$$

d) Precision (optional):
If needed the precision is formulated as *known individual weights*, $\lambda_i = w_i$, or as a *common dispersion parameter*, $\lambda = 1/\sigma^2$, or a *combination* $\lambda_i = w_i/\sigma^2$.

Specification of a generalized linear model in R

```
> mice.glm <- glm(formula = resp ~ conc,
+                 family = binomial(link = logit),
+                 weights = NULL,
+                 data = mice
+                )
```

- formula; as in general linear models
- family
  - binomial(link = logit | probit | cauchit | log | cloglog)
  - gaussian(link = identity | log | inverse)
  - Gamma(link = inverse | identity | log)
  - inverse.gaussian(link = 1/mu^2 | inverse | identity | log)
  - poisson(link = log | identity | sqrt)
  - quasi(link = ... , variance = ...))
  - quasibinomial(link = logit | probit | cauchit | log | cloglog)
  - quasipoisson(link = log | identity | sqrt)

Log-likelihood function

The log likelihood function w.r.t. the canonical parameter

$$\ell_\theta(\boldsymbol{\theta}; \boldsymbol{y}) = \sum_{i=1}^{n} w_i(\theta_i y_i - \kappa(\theta_i))$$

The *score function* w.r.t. the canonical parameter, θ:

$$\frac{\partial}{\partial \theta_i} \ell_\theta(\boldsymbol{\theta}; \boldsymbol{y}) = w_i(y_i - \tau(\theta_i))$$

or in matrix form

$$\frac{\partial}{\partial \boldsymbol{\theta}} \ell_\theta(\boldsymbol{\theta}; \boldsymbol{y}) = \mathrm{diag}(\boldsymbol{w})(\boldsymbol{y} - \boldsymbol{\tau}(\boldsymbol{\theta}))$$

$\mathrm{diag}(\boldsymbol{w})$ denotes a diagonal matrix where i^{th} element is w_i, and

$$\boldsymbol{\tau}(\boldsymbol{\theta}) = \begin{Bmatrix} \tau(\theta_1) \\ \tau(\theta_2) \\ \vdots \\ \tau(\theta_n) \end{Bmatrix}$$

The observed information w.r.t. the canonical parameter

The observed information wrt. to $\boldsymbol{\theta}$ is

$$\begin{aligned} \boldsymbol{j}(\boldsymbol{\theta};\boldsymbol{y}) &= \operatorname{diag}\{\boldsymbol{w}\}\frac{\partial}{\partial\boldsymbol{\theta}}\boldsymbol{\tau}(\boldsymbol{\theta}) \\ &= \operatorname{diag}\{w_i\kappa''(\theta_i)\} \\ &= \operatorname{diag}\{w_i\ V(\tau(\theta_i))\} \end{aligned}$$

where $V(\tau(\theta))$ denotes the value of the variance function for $\mu = \tau(\theta)$.

Note that, since the observed information wrt. the canocical parameter θ does not depend on the observation $\boldsymbol{y}$, the expected information is the same as the observed information.

Note that the Hessian of the likelihood function depends on θ. In the normal case the Hessian was constant.

The score function w.r.t. the mean value parameter

The log likelihood function w.r.t. the mean value parameter

$$\ell_\mu(\boldsymbol{\mu}; \boldsymbol{y}) = -\frac{1}{2} \sum_{i=1}^{n} w_i d(y_i; \mu_i)$$

The score function w.r.t. the mean value parameter is:

$$l'_\mu(\boldsymbol{\mu}; \boldsymbol{y}) = \text{diag}\left\{ \frac{w_i}{V(\mu_i)} \right\} (\boldsymbol{y} - \boldsymbol{\mu})$$

which shows that the ML-estimate for $\boldsymbol{\mu}$ is $\widehat{\boldsymbol{\mu}} = \boldsymbol{y}$.

For a single coordinate we have the log-likelihood:
$\ell(\mu_i; y_i) = -\frac{1}{2} w_i 2 \int_{\mu_i}^{y_i} \frac{y_i - u}{V(u)} du = w_i \int_{y_i}^{\mu_i} \frac{y_i - u}{V(u)} du$
So we get the score:
$\frac{\partial}{\partial \mu_i} \ell(\mu_i; y_i) = w_i \int_{y_i}^{\mu_i} f(u) du = w_i \frac{\partial}{\partial \mu_i} (F(\mu_i) - F(y_i)) = w_i f(\mu_i) = w_i \frac{y_i - \mu_i}{V(\mu_i)}$

The observed information w.r.t. the mean value parameter

The observed information is

$$\boldsymbol{j}(\boldsymbol{\mu};\boldsymbol{y}) = \text{diag}\left\{ w_i \left[\frac{1}{V(\mu_i)} + (y_i - \mu_i)\frac{V'(\mu_i)}{V(\mu_i)^2} \right] \right\},$$

and then the expected information corresponding to the set $\boldsymbol{\mu}$ is

$$\boldsymbol{i}(\boldsymbol{\mu}) = \text{diag}\left\{ \frac{w_i}{V(\mu_i)} \right\},$$

Note that the observed information w.r.t. the mean value parameter $\boldsymbol{\mu}$ depends on the observation, $\boldsymbol{y}$.

For a single coordinate we get:

$$-\frac{\partial}{\partial \mu_i} w_i(y_i - \mu_i)\frac{1}{V(\mu_i)} = -w_i\left(-\frac{1}{V(\mu_i)} + (y_i - \mu_i)\frac{-1}{(V(\mu_i))^2}V'(\mu_i)\right)$$

Maximum likelihood estimation

Theorem (Estimation in generalized linear models)

Consider the generalized linear model as defined on slide 3 for the observations $Y_1, \dots Y_n$ and assume that $Y_1, \dots Y_n$ are mutually independent with densities, which can be described by an exponential dispersion model with the variance function $V(\cdot)$, dispersion parameter σ^2, and optionally the weights w_i.

Assume that the linear predictor is parameterized with $\boldsymbol{\beta}$ corresponding to the design matrix $\boldsymbol{X}$, then the maximum likelihood estimate $\widehat{\boldsymbol{\beta}}$ for $\boldsymbol{\beta}$ is found as the solution to

$$[\boldsymbol{X}(\boldsymbol{\beta})]^T \boldsymbol{i}_\mu(\boldsymbol{\mu})(\boldsymbol{y} - \boldsymbol{\mu}) = \boldsymbol{0} \, ,$$

where $\boldsymbol{X}(\boldsymbol{\beta})$ denotes the local design matrix and $\boldsymbol{\mu} = \boldsymbol{\mu}(\boldsymbol{\beta})$ given by

$$\mu_i(\boldsymbol{\beta}) = g^{-1}(\boldsymbol{x}_i{}^T \boldsymbol{\beta}) \, ,$$

denotes the fitted mean values corresponding to the parameters $\boldsymbol{\beta}$, and $\boldsymbol{i}_\mu(\boldsymbol{\mu})$ is the expected information with respect to $\boldsymbol{\mu}$.

The estimates must be found by an iterative procedure.

Proof

$$\begin{aligned}\ell'_\beta(\boldsymbol{\beta}, \boldsymbol{y}) &= \left[\frac{\partial \mu}{\partial \beta}\right]^T \ell'_\mu(\boldsymbol{\mu}(\boldsymbol{\beta}), \boldsymbol{y}) \\ &= \left[\frac{\partial \mu}{\partial \beta}\right]^T \text{diag}\left\{\frac{w_i}{V(\mu_i)}\right\}(\boldsymbol{y} - \boldsymbol{\mu}) \\ &= \left[\frac{\partial \mu}{\partial \beta}\right]^T \boldsymbol{i}(\boldsymbol{\mu})(\boldsymbol{y} - \boldsymbol{\mu})\end{aligned}$$

And $\frac{\partial \mu}{\partial \beta} = \left[\frac{\partial \mu}{\partial \eta}\right]^T \frac{\partial \eta}{\partial \beta}$ is called the local design matrix.

Properties of the ML estimator

Theorem (Asymptotic distribution of the ML estimator)

Under the hypothesis $\boldsymbol{\eta} = \boldsymbol{X}\boldsymbol{\beta}$ *we have asymptotically*

$$\frac{\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}}{\sqrt{\sigma^2}} \in N_k(\mathbf{0}, \boldsymbol{\Sigma}),$$

where the dispersion matrix $\boldsymbol{\Sigma}$ *for* $\widehat{\boldsymbol{\beta}}$ *is*

$$\mathrm{D}[\widehat{\boldsymbol{\beta}}] = \boldsymbol{\Sigma} = [\boldsymbol{X}^T \boldsymbol{W}(\boldsymbol{\beta}) \boldsymbol{X}]^{-1}$$

with

$$\boldsymbol{W}(\boldsymbol{\beta}) = \mathrm{diag}\left\{ \frac{w_i}{[g'(\mu_i)]^2 V(\mu_i)} \right\},$$

Linear prediction for the generalized linear model

Definition (Linear prediction for the generalized linear model)

The linear prediction $\widehat{\boldsymbol{\eta}}$ is defined as the values

$$\widehat{\boldsymbol{\eta}} = \boldsymbol{X}\widehat{\boldsymbol{\beta}}$$

with the linear prediction corresponding to the i'th observation is

$$\widehat{\eta}_i = \sum_{j=1}^{k} x_{ij}\widehat{\beta}_j = (\boldsymbol{x}_i)^T\widehat{\boldsymbol{\beta}}.$$

The linear predictions $\widehat{\boldsymbol{\eta}}$ are approximately normally distributed with

$$\mathrm{D}[\widehat{\boldsymbol{\eta}}] \approx \widehat{\sigma}^2\boldsymbol{X}\boldsymbol{\Sigma}\boldsymbol{X}^T$$

where $\boldsymbol{\Sigma}$ is the dispersion matrix for $\widehat{\boldsymbol{\beta}}$.

Fitted values for the generalized linear model

Definition (Fitted values for the generalized linear model)

The fitted values are defined as the values

$$\widehat{\boldsymbol{\mu}} = \boldsymbol{\mu}(\boldsymbol{X}\widehat{\boldsymbol{\beta}}) \ ,$$

where the i^{th} value is given as

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

with the fitted value $\widehat{\eta}_i$ of the linear prediction.

The fitted values $\widehat{\boldsymbol{\mu}}$ are approximately normally distributed with

$$\mathrm{D}[\widehat{\boldsymbol{\mu}}] \approx \widehat{\sigma}^2 \left[\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\eta}}\right]^2 \boldsymbol{X}\boldsymbol{\Sigma}\boldsymbol{X}^T$$

where $\boldsymbol{\Sigma}$ is the dispersion matrix for $\widehat{\boldsymbol{\beta}}$.

Residual deviance

Definition (Residual deviance)

Consider the generalized linear model defined on slide 3. The *residual deviance* corresponding to this model is

$$D(\boldsymbol{y}; \boldsymbol{\mu}(\widehat{\boldsymbol{\beta}})) = \sum_{i=1}^{n} w_i d(y_i; \widehat{\mu}_i)$$

with $d(y_i; \widehat{\mu}_i)$ denoting the unit deviance corresponding the observation y_i and the fitted value $\widehat{\mu}_i$ and where w_i denotes the weights (if present).
If the model includes a dispersion parameter σ^2, the *scaled* residual deviance is

$$D^*(\boldsymbol{y}; \boldsymbol{\mu}(\widehat{\boldsymbol{\beta}})) = \frac{D(\boldsymbol{y}; \boldsymbol{\mu}(\widehat{\boldsymbol{\beta}}))}{\sigma^2}.$$

Residuals

Residuals represents the difference between the data and the model. In the classical GLM the residuals are $r_i = y_i - \widehat{\mu}_i$. These are called response residuals for GLM's. Since the variance of the response is not constant for most GLM's we need some modification. We will look at:

- Deviance residuals
- Pearson residuals

Residuals

Definition (Deviance residual)

Consider the generalized linear model from for the observations $Y_1, \dots Y_n$.

The *deviance residual* for the i'th observation is defined as

$$r_i^D = r_D(y_i; \widehat{\mu}_i) = \mathrm{sign}(y_i - \widehat{\mu}_i)\sqrt{w_i d(y_i, \widehat{\mu}_i)}$$

where $\mathrm{sign}(x)$ denotes the *sign function* $\mathrm{sign}(x) = 1$ for $x > 0$ og $\mathrm{sign}(x) = -1$ for $x < 0$, and with w_i denoting the weight (if relevant), $d(y; \mu)$ denoting the unit deviance and $\widehat{\mu}_i$ denoting the fitted value corresponding to the i'th observation.

Assessments of the deviance residuals is in good agreement with the likelihood approach as the deviance residuals simply express differences in log-likelihood.

Residuals

Definition (Pearson residual)

Consider again the generalized linear model from for the observations $Y_1, \ldots Y_n$.

The *Pearson residuals* are defined as the values

$$r_i^P = r_P(y_i; \widehat{\mu}_i) = \frac{y_i - \widehat{\mu}_i}{\sqrt{V(\widehat{\mu}_i)/w_i}}$$

The Pearson residual is thus obtained by scaling the response residual with $\sqrt{\mathrm{Var}[Y_i]}$. Hence, the Pearson residual is the response residual normalized with the estimated standard deviation for the observation.

Likelihood ratio tests

- The approximative normal distribution of the ML-estimator implies that many distributional results from the classical GLM-theory are carried over to generalized linear models as approximative (asymptotic) results.
- An example of this is the likelihood ratio test.
- In the classical GLM case it was possible to derive the exact distribution of the likelihood ratio test statistic (the F-distribution).
- For generalized linear models, this is not possible, and hence we shall use the asymptotic results for the logarithm of the likelihood ratio.

Likelihood ratio test

Theorem (Likelihood ratio test)

Consider the generalized linear model. Assume that the model

$$\mathcal{H}_1 : \boldsymbol{\eta} \in L \subset \mathbb{R}^k$$

holds with L parameterized as $\boldsymbol{\eta} = \boldsymbol{X}_1\boldsymbol{\beta}$, and consider the hypotheses

$$\mathcal{H}_0 : \boldsymbol{\eta} \in L_0 \subset \mathbb{R}^m$$

where $\boldsymbol{\eta} = \boldsymbol{X}_0\boldsymbol{\alpha}$ and $m < k$, and with the alternative $\mathcal{H}_1 : \boldsymbol{\eta} \in L \backslash L_0$. Then the likelihood ratio test for $\mathcal{H}_0$ has the test statistic

$$-2 \log \lambda(\boldsymbol{y}) = \mathrm{D}\left(\boldsymbol{y}; \boldsymbol{\mu}(\widehat{\boldsymbol{\alpha}})\right) - \mathrm{D}\left(\boldsymbol{y}; \boldsymbol{\mu}(\widehat{\boldsymbol{\beta}})\right)$$

When $\mathcal{H}_0$ is true, the test statistic will asymptotically follow a $\chi^2(k-m)$ distribution.

If the model includes a dispersion parameter, σ^2, then $\mathrm{D}\left(\boldsymbol{\mu}(\widehat{\boldsymbol{\beta}}); \boldsymbol{\mu}(\widehat{\boldsymbol{\alpha}})\right)$ will asymptotically follow a $\sigma^2\chi^2(k-m)$ distribution.

Test for model 'sufficiency'

- In analogy with classical GLM's one often starts with formulating a rather comprehensive model, and then reduces the model by successive tests.
- In contrast to classical GLM's we may however test the goodness of fit of the initial model.
- The test is a special case of the likelihood ratio test.

Test for model 'sufficiency'

Test for model 'sufficiency'

Consider the generalized linear model, and assume that the dispersion $\sigma^2 = 1$.

Let $\mathcal{H}_{full}$ denote the *full*, or *saturated* model, i.e. $\mathcal{H}_{full} : \boldsymbol{\mu} \in \mathbb{R}^n$ and consider the hypotheses

$$\mathcal{H}_0 : \boldsymbol{\eta} \in L \subset \mathbb{R}^k$$

with L parameterized as $\boldsymbol{\eta} = \boldsymbol{X}_0\boldsymbol{\beta}$.

Then, as the residual deviance under $\mathcal{H}_{full}$ is 0, the test statistic is the residual deviance $\mathrm{D}\left(\boldsymbol{\mu}(\widehat{\boldsymbol{\beta}})\right)$. When $\mathcal{H}_0$ is true, the test statistic is distributed as $\chi^2(n-k)$. The test rejects for large values of $\mathrm{D}\left(\boldsymbol{\mu}(\widehat{\boldsymbol{\beta}})\right)$.

Residual deviance measures goodness of fit

- The residual deviance $\mathrm{D}\left(\boldsymbol{y};\boldsymbol{\mu}(\widehat{\boldsymbol{\beta}})\right)$ is a reasonable measure of the goodness of fit of a model $\mathcal{H}_0$.
- When referring to a hypothesized model $\mathcal{H}_0$, we shall sometimes use the symbol $G^2(\mathcal{H}_0)$ to denote the residual deviance $\mathrm{D}\left(\boldsymbol{y};\boldsymbol{\mu}(\widehat{\boldsymbol{\beta}})\right)$.
- Using that convention, the partitioning of residual deviance may be formulated as

$$G^2(\mathcal{H}_0|\mathcal{H}_1) = G^2(\mathcal{H}_0) - G^2(\mathcal{H}_1)$$

with $G^2(\mathcal{H}_0|\mathcal{H}_1)$ interpreted as the goodness fit test statistic for $\mathcal{H}_0$ conditioned on $\mathcal{H}_1$ being true, and $G^2(\mathcal{H}_0)$ and $G^2(\mathcal{H}_1)$, denoting the unconditional goodness of fit statistics for $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively.

Analysis of deviance table

- The initial test for goodness of fit of the initial model is often represented in an *analysis of deviance table* in analogy with the ANOVA table for classical GLM's.
- In the table the goodness of fit test statistic corresponding to the initial model $G^2(\mathcal{H}_1) = \mathrm{D}\left(\boldsymbol{y}; \boldsymbol{\mu}(\widehat{\boldsymbol{\beta}})\right)$ is shown in the line labelled "Error".
- The statistic should be compared to percentiles in the $\chi^2(n-k)$ distribution.
- The table also shows the test statistic for $\mathcal{H}_{null}$ under the assumption that $\mathcal{H}_1$ is true.
- The test investigates whether the model is necessary at all, i.e. whether at least some of the coefficients differ significantly from zero.

Analysis of deviance table

- Note, that in the case of a generalized linear model, we can start the analysis by using the residual (error) deviance to test whether the model may be maintained, at all.
- This is in contrast to the classical GLM's where the residual sum of squares around the initial model $\mathcal{H}_1$ served to estimate σ^2, and therefore we had no reference value to compare with the residual sum of squares.
- In the generalized linear models the variance is a known function of the mean, and therefore in general there is no need to estimate a separate variance.

Analysis of deviance table

Source	f	Deviance	Mean deviance	Goodness of fit interpretation
Model $\mathcal{H}_{null}$	$k-1$	$\mathrm{D}\left(\boldsymbol{\mu}(\widehat{\boldsymbol{\beta}}); \widehat{\boldsymbol{\mu}}_{null}\right)$	$\dfrac{\mathrm{D}\left(\boldsymbol{\mu}(\widehat{\boldsymbol{\beta}}); \widehat{\boldsymbol{\mu}}_{null}\right)}{k-1}$	$G^2(\mathcal{H}_{null} \mid \mathcal{H}_1)$
Residual (Error)	$n-k$	$\mathrm{D}\left(\boldsymbol{y}; \boldsymbol{\mu}(\widehat{\boldsymbol{\beta}})\right)$	$\dfrac{\mathrm{D}\left(\boldsymbol{y}; \boldsymbol{\mu}(\widehat{\boldsymbol{\beta}})\right)}{n-k}$	$G^2(\mathcal{H}_1)$
Corrected total	$n-1$	$\mathrm{D}\left(\boldsymbol{y}; \widehat{\boldsymbol{\mu}}_{null}\right)$		$G^2(\mathcal{H}_{null})$

Table: Initial assessment of goodness of fit of a model $\mathcal{H}_0$. $\mathcal{H}_{null}$ and $\widehat{\boldsymbol{\mu}}_{null}$ refer to the *minimal model*, i.e. a model with all observations having the same mean value.

Odds Ratio

If an event occurs with probability p, then the odds in favor of the event is

$$\text{Odds} = \frac{p}{1-p}$$

A comparison between two events can be made by computing the odds ratio:

$$\text{OR} = \frac{p_1/(1-p_1)}{p_2/(1-p_2)}$$

An odds ratio larger than 1 is an indication the event is more likely in the first group than in the second group.