LOGISTIC REGRESSION for UNGROUPED DATA

MODEL ASSUMPTIONS:

- $Y_i \sim Bern(\pi_i)$ indep. $i = 1, \dots, n$
- $\eta_i = \tilde{x}_i^T \beta$
- $g(\pi_i) = \log\left(\frac{\pi_i}{1-\pi_i}\right) = \eta_i$

$\downarrow$

if we invert the relationship between π_i and η_i we obtain

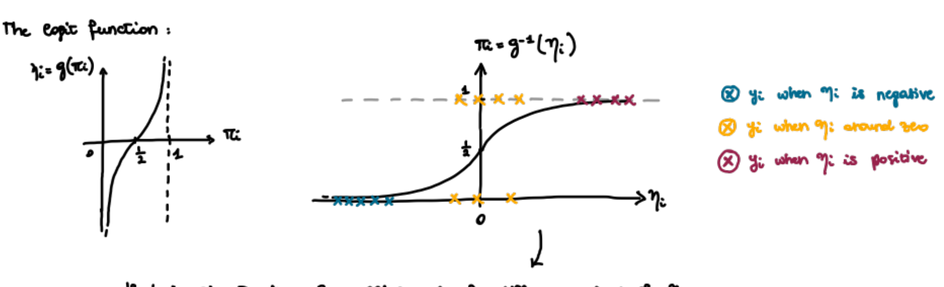
$$\pi_i = g^{-1}(\eta_i) = \frac{e^{\eta_i}}{1+e^{\eta_i}} \in (0,1)$$

Hence we can write the model as

$Y_i \sim Bernoulli(\pi_i)$ independent for $i = 1, \dots, n$

with $\pi_i = g^{-1}(\tilde{x}_i^T\beta) = \dfrac{e^{\tilde{x}_i^T\beta}}{1+e^{\tilde{x}_i^T\beta}} = \mathbb{E}[Y_i] = \mathbb{P}(Y_i = 1)$

The logit function:



If I imagine to draw Bernoulli samples for different values of η_i:

- if $\eta_i \ll 0 \Rightarrow \pi_i$ close to zero: I observe many failures
- if $\eta_i \approx 0 \Rightarrow \pi_i \approx 0.5$ similar number of failures and successes
- if $\eta_i \gg 0 \Rightarrow \pi_i$ close to 1: many successes

• INTERPRETATION of the COEFFICIENTS β_j

Logistic regression has a convenient interpretation of the parameters in terms of LOG-ODDS

Indeed, $\eta_i = \log \dfrac{\pi_i}{1-\pi_i}$

The ratio $\dfrac{\pi_i}{1-\pi_i} = \text{ODDS} = \dfrac{\text{prob. of success}}{\text{prob. of failure}}$

If I consider $(\text{odds} \cdot 100) = \frac{\pi_i}{1-\pi_i} \cdot 100$ = expected number of successes every 100 failures

eg. if odds = 2 in the beetles experiment $\Rightarrow$ success = dead, fail = alive.

"every 100 alive beetles, 200 are killed"

$\left(\text{odds} = 2 \Leftrightarrow \frac{\pi_i}{1-\pi_i} = 2 \Leftrightarrow \pi_i = 2/3\right)$

Since $\eta_i = \log\frac{\pi_i}{1-\pi_i} = \beta_1 x_{i1} + \dots + \beta_j x_{ij} + \dots + \beta_p x_{ip}$

$\rightarrow$ it is a linear model for the log-odds

the coefficient β_j is the change of the log-odds if x_{ij} is increased of 1 unit, while keeping the remaining covariates fixed.

Let's consider again the beetle data: $y_i = \underset{(1)}{\text{dead}} / \underset{(0)}{\text{alive}}$, $x_i = \log(\text{dose})$ (only one covariate)

$\mathbb{P}(Y_i = 1) = \pi_i = \dfrac{e^{\beta_1+\beta_2 x_i}}{1+e^{\beta_1+\beta_2 x_i}}$

$\log\frac{\pi_i}{1-\pi_i} = \beta_1 + \beta_2 x_i$

at a $\log(\text{dose}) = x_0$: $\log\frac{\pi_0}{1-\pi_0} = \beta_1 + \beta_2 x_0$

at $\log(\text{dose}) = x_1 = x_0 + 1$: $\log\frac{\pi_1}{1-\pi_1} = \beta_1 + \beta_2(x_0+1) = \beta_1 + \beta_2 x_0 + \beta_2$

hence $\log\frac{\pi_1}{1-\pi_1} - \log\frac{\pi_0}{1-\pi_0} = \beta_2$

$$\Rightarrow \log \frac{\left(\frac{\pi_1}{1-\pi_1}\right)}{\left(\frac{\pi_0}{1-\pi_0}\right)} = \beta_2 = \log\frac{(\text{odds} \mid x_0+1)}{(\text{odds} \mid x_0)} \Rightarrow \frac{\left(\frac{\pi_1}{1-\pi_1}\right)}{\left(\frac{\pi_0}{1-\pi_0}\right)} = e^{\beta_2}$$

Indeed, in general,

$$\frac{\left(\frac{\pi_i}{1-\pi_i} \mid x_{ij} = x_0+1\right)}{\left(\frac{\pi_i}{1-\pi_i} \mid x_{ij} = x_0\right)} = e^{\beta_j} \quad \text{"ODDS RATIO"}$$

$$\left(\frac{\pi_i}{1-\pi_i} \mid x_{ij} = x_0+1\right) = \left(\frac{\pi_i}{1-\pi_i} \mid x_{ij} = x_0\right) \cdot e^{\beta_j}$$

if I increase the j-th variable of 1 unit (additive variation in the covariate), the initial odds (at x_0) is multiplied of a coefficient e^{β_j} (multiplicative effect of the odds)

- INTERPRETATION with binary covariate (2×2 contingency table)

consider a logistic regression with only one binary covariate

e.g. in a study on the effect of a treatment

$Y_i = \begin{cases} 1 & \text{success} \\ 0 & \text{failure} \end{cases}$ $\quad z_i = \begin{cases} 1 & \text{treatment} \\ 0 & \text{placebo} \end{cases}$

	$z_i = 1$	$z_i = 0$
$y_i = 1$	•	•
$y_i = 0$	•	•

$Y_i \sim Bernoulli(\pi_i)$ $\quad \pi_i = \dfrac{e^{\beta_1+\beta_2 z_i}}{1+e^{\beta_1+\beta_2 z_i}}$

$(\pi_i \mid z_i = 1) = \mathbb{P}(Y_i = 1 \mid z_i = 1) = \dfrac{e^{\beta_1+\beta_2}}{1+e^{\beta_1+\beta_2}}$ and $(1-\pi_i \mid z_i = 1) = \mathbb{P}(Y_i = 0 \mid z_i = 1) = \dfrac{1}{1+e^{\beta_1+\beta_2}}$

$\Rightarrow \dfrac{\pi_i}{1-\pi_i} \mid z_i = 1 = e^{\beta_1+\beta_2}$ odds when individual i has the treatment

$(\pi_i \mid z_i = 0) = \mathbb{P}(Y_i = 1 \mid z_i = 0) = \dfrac{e^{\beta_1}}{1+e^{\beta_1}}$ and $(1-\pi_i \mid z_i = 0) = \mathbb{P}(Y_i = 0 \mid z_i = 0) = \dfrac{1}{1+e^{\beta_1}}$

$\Rightarrow \dfrac{\pi_i}{1-\pi_i} \mid z_i = 0 = e^{\beta_1}$ odds when individual i has the placebo

Hence $\dfrac{\left(\frac{\pi_i}{1-\pi_i} \mid z_i = 1\right)}{\left(\frac{\pi_i}{1-\pi_i} \mid z_i = 0\right)} = e^{\beta_2}$ odds ratio $\rightarrow$ The odds using a placebo are multiplied by a factor e^{β_2} to have the odds under the treatment

or, $\beta_2 = \log \left[\dfrac{\frac{\mathbb{P}(Y_i=1 \mid z_i=1)}{\mathbb{P}(Y_i=0 \mid z_i=1)}}{\frac{\mathbb{P}(Y_i=1 \mid z_i=0)}{\mathbb{P}(Y_i=0 \mid z_i=0)}} \right]$

• ESTIMATION

likelihood

$$L(\underline{\beta}) \propto \prod_{i=1}^n p(y_i \mid \underline{\beta}) = \prod_{i=1}^n \pi_i^{y_i}(1-\pi_i)^{1-y_i} = \prod_{i=1}^n \left(\frac{e^{\tilde{x}_i^T\beta}}{1+e^{\tilde{x}_i^T\beta}}\right)^{y_i} \left(\frac{1}{1+e^{\tilde{x}_i^T\beta}}\right)^{1-y_i}$$

$$\ell(\underline{\beta}) = \sum_{i=1}^n \{y_i \log \pi_i + (1-y_i)\log(1-\pi_i)\}$$

$\log \pi_i = \log \dfrac{e^{\tilde{x}_i^T\beta}}{1+e^{\tilde{x}_i^T\beta}} = \tilde{x}_i^T\beta - \log(1+e^{\tilde{x}_i^T\beta})$

$\log(1-\pi_i) = \log \dfrac{1}{1+e^{\tilde{x}_i^T\beta}} = -\log(1+e^{\tilde{x}_i^T\beta})$

$$\ell(\underline{\beta}) = \sum_{i=1}^n \left\{ y_i \tilde{x}_i^T\beta - \cancel{y_i \log(1+e^{\tilde{x}_i^T\beta})} - \log(1+e^{\tilde{x}_i^T\beta}) + \cancel{y_i \log(1+e^{\tilde{x}_i^T\beta})} \right\}$$
$$= \sum_{i=1}^n \left\{ y_i \tilde{x}_i^T\beta - \log(1+e^{\tilde{x}_i^T\beta}) \right\}$$

$$\frac{\partial \ell(\beta)}{\partial \beta_r} = \sum_{i=1}^n \left\{ y_i x_{ir} - \frac{1}{1+e^{\tilde{x}_i^T\beta}} \cdot e^{\tilde{x}_i^T\beta} \cdot x_{ir} \right\}$$

$$\ell_*(\underline{\beta}) = \frac{\partial}{\partial \underline{\beta}} \ell(\underline{\beta}) = \sum_{i=1}^n \left\{ y_i \tilde{x}_i - \frac{e^{\tilde{x}_i^T\beta}}{1+e^{\tilde{x}_i^T\beta}} \cdot \tilde{x}_i \right\}$$
$$= \sum_{i=1}^n \tilde{x}_i \left(y_i - \frac{e^{\tilde{x}_i^T\beta}}{1+e^{\tilde{x}_i^T\beta}} \right) = \sum_{i=1}^n \tilde{x}_i (y_i - \pi_i) = X^T(\underline{y} - \underline{\pi})$$

likelihood equation: $\ell_*(\underline{\beta}) = 0 \Rightarrow X^T(\underline{y} - \underline{\pi}) = 0$ again they resemble the normal equations but they are not linear in β

Similarly to the Poisson case, we need to solve the equation numerically and we do not have a closed-form expression of the MLE $\hat{\underline{\beta}}$.

Finally, the 2nd derivative is

$$\frac{\partial^2 \ell(\beta)}{\partial \beta_r \partial \beta_s} = \frac{\partial}{\partial \beta_s}\left(-\sum_{i=1}^n \frac{e^{\tilde{x}_i^T\beta}}{1+e^{\tilde{x}_i^T\beta}} x_{ir}\right) = -\sum_{i=1}^n \frac{e^{\tilde{x}_i^T\beta} \cdot x_{is} \cdot x_{ir}(1+e^{\tilde{x}_i^T\beta}) - e^{\tilde{x}_i^T\beta} x_{ir} \cdot e^{\tilde{x}_i^T\beta} x_{is}}{(1+e^{\tilde{x}_i^T\beta})^2}$$
$$= -\sum_{i=1}^n \frac{e^{\tilde{x}_i^T\beta} x_{is} x_{ir}}{(1+e^{\tilde{x}_i^T\beta})^2} = -\sum_{i=1}^n x_{ir} x_{is} \pi_i (1-\pi_i)$$

$\Rightarrow \dfrac{\partial \ell(\beta)}{\partial \beta \partial \beta^T} = -X^T U X$ with $U = diag\{\pi_1(1-\pi_1), \dots, \pi_n(1-\pi_n)\} = U(\beta)$

$j(\underline{\beta}) = -\ell_{**}(\beta) = X^T U X$ and $j(\hat{\underline{\beta}}) = X^T U(\hat{\beta}) X$

• INFERENCE

DISTRIBUTION of the ESTIMATOR of the REGRESSION PARAMETERS (approximate distribution)

$\hat{\underline{\beta}} \dot{\sim} N_p(\underline{\beta}, j(\hat{\underline{\beta}})^{-1})$

and the marginal is $\hat{\beta}_j \dot{\sim} N(\beta_j, [j(\hat{\underline{\beta}})^{-1}]_{jj})$ $\quad j = 1, \dots, p$

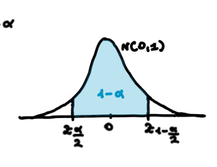
Confidence intervals and tests are obtained as usual:

- $(1-\alpha)$ confidence interval $\hat{B}_j$:

$\mathbb{P}(\beta_j \in \hat{B}_j) = 1-\alpha \quad \forall \beta_j \in \mathbb{R} \Rightarrow \mathbb{P}\left(z_{\frac{\alpha}{2}} \le \underbrace{\frac{\hat{\beta}_j - \beta_j}{\sqrt{[j(\hat{\beta})^{-1}]_{jj}}}}_{\dot{\sim} N(0,1)} \le z_{1-\frac{\alpha}{2}} \right) = 1-\alpha$, $\quad z_{\frac{\alpha}{2}} = -z_{1-\frac{\alpha}{2}}$

hence a CI is

$\hat{B}_j = \hat{\beta}_j \pm z_{1-\frac{\alpha}{2}} \sqrt{[j(\hat{\beta})^{-1}]_{jj}}$

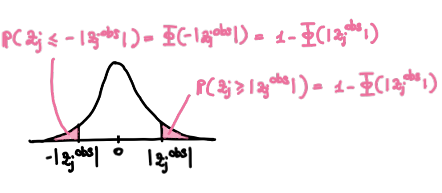


- test $H_0: \beta_j = b_j$ vs $H_1: \beta_j \ne b_j$

under H_0 we obtain the distribution $\hat{\beta}_j \dot{\sim} N(b_j, [j(\hat{\beta})^{-1}]_{jj})$

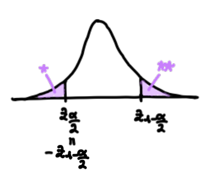
Hence $z_j = \dfrac{\hat{\beta}_j - b_j}{\sqrt{[j(\hat{\beta})^{-1}]_{jj}}} \dot{\sim} N(0,1)$ under H_0

- p-value is $\alpha^{obs} = \mathbb{P}_{H_0}(|z_j| \ge |z^{obs}|) = 2(1 - \Phi(|z_j^{obs}|))$



- rejection region of a test with significance level α is:

$R = (-\infty, z_{\frac{\alpha}{2}}) \cup (z_{1-\frac{\alpha}{2}}, +\infty)$
$= (-\infty, -z_{1-\frac{\alpha}{2}}) \cup (z_{1-\frac{\alpha}{2}}, +\infty)$
$= (*) \cup (**)$



- TEST for COMPARING NESTED MODELS

(test about a subset of the parameters)

we have the proposed "full" model

$Y_i \sim Bernoulli(\pi_i)$ indep. for $i = 1, \dots, n$

$\pi_i = \dfrac{e^{\tilde{x}_i^T\beta}}{1+e^{\tilde{x}_i^T\beta}}$ $\quad \tilde{x}_i^T\beta = \beta_1 + \beta_2 x_{i2} + \dots + \beta_{p_0} x_{ip_0} + \underbrace{\boxed{\beta_{p_0+1} x_{i\,p_0+1} + \dots + \beta_p x_{ip}}}_{=0 \text{ under } H_0}$

we want to test

$\begin{cases} H_0: \beta_{p_0+1} = \dots = \beta_p = 0 \\ H_1: \overline{H_0} \end{cases}$

as usual, we partition $\underline{\beta} = \begin{bmatrix} \underline{\beta}^{(0)} \\ \underline{\beta}^{(1)} \end{bmatrix}$ $\quad \underline{\beta}^{(0)} \in \mathbb{R}^{p_0}$, $\underline{\beta}^{(1)} \in \mathbb{R}^{p-p_0}$

so the test can be reformulated as

$\begin{cases} H_0: \underline{\beta}^{(1)} = \underline{0} \\ H_1: \underline{\beta}^{(1)} \ne \underline{0} \end{cases}$

Similarly to what we have seen with the Poisson regression, to perform this test we use the LIKELIHOOD RATIO TEST:

$$W = 2 \log \frac{\hat{L}(\text{model})}{\tilde{L}(\text{restricted})} = 2\{\hat{\ell}(\text{model}) - \tilde{\ell}(\text{restricted})\} \dot{\sim} \chi^2_{\underbrace{p-p_0}_{\text{number of parameters we are testing}}} \text{ under } H_0$$

We estimate the full model (H_1) and obtain $(\hat{\underline{\beta}}^{(0)}, \hat{\underline{\beta}}^{(1)}) = \hat{\underline{\beta}}$

$\Rightarrow \ell(\hat{\underline{\beta}}^{(0)}, \hat{\underline{\beta}}^{(1)}) = \sum_{i=1}^n y_i \log \hat{\pi}_i + \sum_{i=1}^n (1-y_i)\log(1-\hat{\pi}_i)$

we estimate the restricted model (H_0) and obtain $(\tilde{\underline{\beta}}^{(0)}, \underline{0}) = \tilde{\underline{\beta}}$

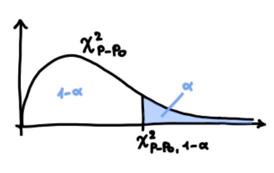
$\Rightarrow \ell(\tilde{\underline{\beta}}^{(0)}, \underline{0}) = \sum_{i=1}^n y_i \log \tilde{\pi}_i + \sum_{i=1}^n (1-y_i)\log(1-\tilde{\pi}_i)$

The observed value of the test is

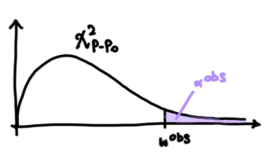
$$w^{obs} = 2\{\ell(\hat{\underline{\beta}}^{(0)}, \hat{\underline{\beta}}^{(1)}) - \ell(\tilde{\underline{\beta}}^{(0)}, \underline{0})\}$$
$$= 2\left\{\sum_{i=1}^n y_i \log\hat{\pi}_i + \sum_{i=1}^n (1-y_i)\log(1-\hat{\pi}_i) - \sum_{i=1}^n y_i \log\tilde{\pi}_i - \sum_{i=1}^n (1-y_i)\log(1-\tilde{\pi}_i)\right\}$$
$$= 2\left\{\sum_{i=1}^n \left[y_i \log\hat{\pi}_i - y_i\log\tilde{\pi}_i + (1-y_i)\log(1-\hat{\pi}_i) - (1-y_i)\log(1-\tilde{\pi}_i)\right]\right\}$$
$$= 2\left\{\sum_{i=1}^n \left[y_i \log\frac{\hat{\pi}_i}{\tilde{\pi}_i} + (1-y_i)\log\frac{(1-\hat{\pi}_i)}{(1-\tilde{\pi}_i)}\right]\right\}$$

If the null hypothesis is true, $\hat{\ell}(\text{model}) \approx \tilde{\ell}(\text{restricted}) \Rightarrow W \approx 0$

If the null hypothesis is not true, $\hat{\ell}(\text{model}) > \tilde{\ell}(\text{restricted}) \Rightarrow W > 0$ $\rightarrow$ reject for large values



Rejection / critical region with significance level α: $R = (\underbrace{\chi^2_{p-p_0, 1-\alpha}}_{\text{quantile } (1-\alpha) \text{ of a } \chi^2_{p-p_0} \text{ distribution}}; +\infty)$



p-value: $\alpha^{obs} = \mathbb{P}_{H_0}(W \ge w^{obs})$

- TEST about the OVERALL MODEL

We compare the proposed model with the "null" model (i.e. a model with only the intercept)

$\begin{cases} H_0: \beta_2 = \beta_3 = \dots = \beta_p = 0 \\ H_1: \overline{H_0} \end{cases}$

We use the previous test with $p_0 = 1$

$\underline{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}$ $\quad \beta_1 = \beta^{(0)} \in \mathbb{R}$, $\quad \underline{\beta}^{(1)} \in \mathbb{R}^{p-1}$

We need to compute the maximum of the log-likelihood under the null model

under H_0 $\quad Y_i \sim Bern(\pi_i)$

$\pi_i = \dfrac{e^{\beta_1}}{1+e^{\beta_1}} = \pi$ $\quad$ one-to-one correspondence: I can compute the estimate of $\tilde{\pi}$ and obtain $\tilde{\beta}_1 = \log\frac{\tilde{\pi}}{1-\tilde{\pi}}$

$L(\pi) = \prod_{i=1}^n \pi^{y_i}(1-\pi)^{1-y_i}$

$\ell(\pi) = \sum_{i=1}^n \{y_i \log\pi + (1-y_i)\log(1-\pi)\}$

$\ell_*(\pi) = \sum_{i=1}^n \left\{\frac{y_i}{\pi} - \frac{1-y_i}{1-\pi}\right\} = \sum_{i=1}^n \left\{\frac{y_i - \cancel{\pi y_i} - \pi + \cancel{\pi y_i}}{\pi(1-\pi)}\right\}$

$\ell_*(\pi) = 0 \Rightarrow \sum_{i=1}^n y_i - n\pi = 0 \Rightarrow \tilde{\pi} = \bar{y}$ $\quad$ MLE under the null model is the sample mean

$\ell_{**}(\pi) = \sum_{i=1}^n \left\{-\frac{y_i}{\pi^2} - \frac{(1-y_i)}{(1-\pi)^2}\right\} = -\frac{n\bar{y}}{\pi^2} - \frac{n - n\bar{y}}{(1-\pi)^2}$

$\ell_{**}(\tilde{\pi}) = -\frac{n\cancel{\bar{y}}}{\bar{y}^{\cancel{2}}} - \frac{n\cancel{(1-\bar{y})}}{(1-\bar{y})^{\cancel{2}}} = -\frac{n}{\bar{y}} - \frac{n}{(1-\bar{y})} < 0$

Hence $\tilde{\ell}(\text{null}) = \ell(\tilde{\pi}) = \ell(\tilde{\beta}_1)$

$= \sum_{i=1}^n \{y_i \log\tilde{\pi} + (1-y_i)\log(1-\tilde{\pi})\} = \sum_{i=1}^n \{y_i \log\bar{y} + (1-y_i)\log(1-\bar{y})\}$

$= n\bar{y}\log\bar{y} + n(1-\bar{y})\log(1-\bar{y})$

Under H_1 we do not have a closed-form expression for $\hat{\underline{\beta}} = (\hat{\underline{\beta}}^{(0)}, \hat{\underline{\beta}}^{(1)})$

$\Rightarrow \ell(\hat{\underline{\beta}}^{(0)}, \hat{\underline{\beta}}^{(1)}) = \sum_{i=1}^n y_i \log\hat{\pi}_i + \sum_{i=1}^n (1-y_i)\log(1-\hat{\pi}_i) = \hat{\ell}(\text{model})$

The LR test in this case is $W = 2(\hat{\ell}(\text{model}) - \tilde{\ell}(\text{null})) \dot{\sim} \chi^2_{p-1}$

$$w^{obs} = 2\left\{\sum_{i=1}^n \left[y_i\log\hat{\pi}_i + (1-y_i)\log(1-\hat{\pi}_i) - y_i\log\bar{y} - (1-y_i)\log(1-\bar{y})\right]\right\}$$
$$= 2\left\{\sum_{i=1}^n \left[y_i \log\frac{\hat{\pi}_i}{\bar{y}} + (1-y_i)\log\frac{(1-\hat{\pi}_i)}{(1-\bar{y})}\right]\right\}$$

- DEVIANCE for Bernoulli regression

We defined the deviance as the LR statistic to compare the saturated model with a model with $p < n$ parameters.

(Saturated model: a model with n parameters, one for each observation. We obtain a model with a perfect fit, since we are interpolating the n points).

What happens to the Bernoulli log-likelihood when we compute it for the saturated model?

We have $Y_i \sim Bernoulli(\pi_i)$ with a separate π_i $\forall i$

$L(\pi_i) = \pi_i^{y_i}(1-\pi_i)^{1-y_i}$

$\ell(\pi_i) = y_i\log\pi_i + (1-y_i)\log(1-\pi_i) = \begin{cases} \log\pi_i & \text{if } y_i = 1 \\ \log(1-\pi_i) & \text{if } y_i = 0 \end{cases}$

$\ell_*(\pi_i) = \dfrac{y_i}{\pi_i} - \dfrac{(1-y_i)}{1-\pi_i}$

$= \dfrac{(1-\pi_i)y_i - \pi_i(1-y_i)}{\pi_i(1-\pi_i)}$

$\ell_*(\pi_i) = 0 \Rightarrow y_i - \cancel{\pi_i y_i} - \pi_i + \cancel{\pi_i y_i} = 0$

$\tilde{\pi}_i^S = y_i \in \{0, 1\}$ $\quad$ I estimate a probability that is exactly equal to 0 if $y_i = 0$, and equal to 1 if $y_i = 1$.

if $y_i = 1 \Rightarrow \tilde{\pi}_i^S = 1 \Rightarrow \ell(\tilde{\pi}_i^S) = \log 1 = 0$

if $y_i = 0 \Rightarrow \tilde{\pi}_i^S = 0 \Rightarrow \ell(\tilde{\pi}_i^S) = \log(1-0) = 0$

the loglikelihood for the saturated model is always equal to 0.

Hence $D = \text{deviance}(\text{model}) = 2\{\tilde{\ell}(\text{saturated}) - \hat{\ell}(\text{model})\} = -2\hat{\ell}(\text{model})$

$$D = -2\left(\sum_{i=1}^n y_i\log\hat{\pi}_i + \sum_{i=1}^n (1-y_i)\log(1-\hat{\pi}_i)\right)$$
$$= \sum_{i=1}^n \underbrace{-2(y_i\log\hat{\pi}_i + (1-y_i)\log(1-\hat{\pi}_i))}_{\text{individual contribution } D_i} = \sum_{i=1}^n D_i$$

if $y_i = 1$ $\quad D_i = -2\log\hat{\pi}_i$

if $y_i = 0$ $\quad D_i = -2\log(1-\hat{\pi}_i)$

When we assume a Bernoulli distribution for Y_i, the deviance is not useful to evaluate the goodness of fit of the model.

However, we can still derive the test about the overall model as

$D(\text{null}) - D(\text{model}) = -2\tilde{\ell}(\text{null}) - (-2\hat{\ell}(\text{model}))$
$= 2\{\hat{\ell}(\text{model}) - \tilde{\ell}(\text{null})\}$ = LR test between null and proposed model

"null deviance" - "residual deviance"

Also the analysis of the residuals in this setting is not useful.