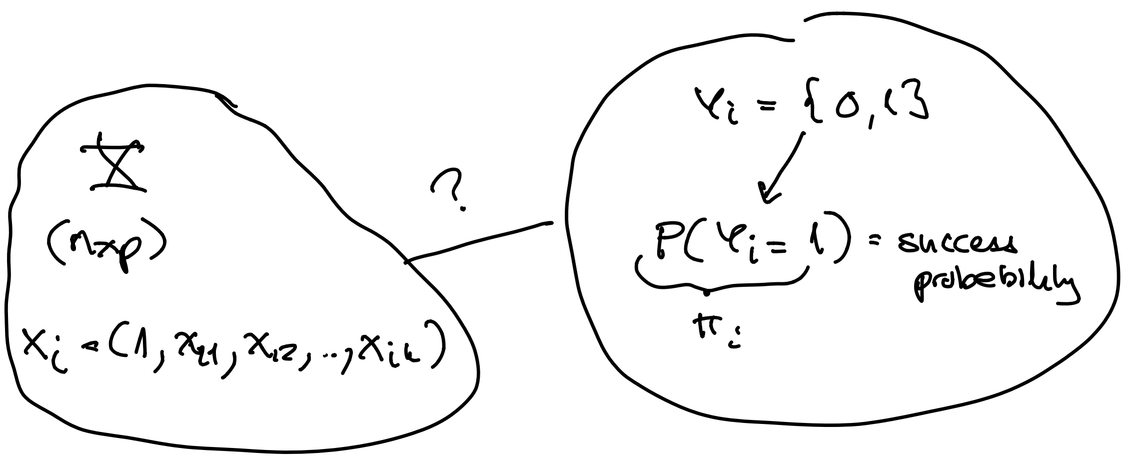
11.09.2017 TMA4315 GLM

Module 3: Binary regression (week 1)



Beetle ex: (x_i, Y_i) $i = 1, .., n$ individual data

Normal model	Binary model
1. $Y_i \sim N(\mu_i, \sigma^2)$	1. $Y_i \sim bin(n_i = 1, \pi_i)$
$E(Y_i) = \mu_i$	$E(Y_i) = \pi_i$
$Var(Y_i) = \sigma^2$	$Var(Y_i) = \sigma^2$

exponential family distribution

2. linear predictor $\eta = x_i^T \beta$

3. $\mu_i = \eta_i$

3. $\underset{\pi_i}{\mu_i} = h(\eta_i)$ = response function

$\eta_i = h^{-1}(\mu_i) = g(\mu_i)$ - link function

The logit model

$Y_i \sim \text{bin}(n_i = 1, \pi_i)$

$E(Y_i) = \pi_i \;, \; Var(Y_i) = \pi_i(1-\pi_i)$

$E(Y_i) = \mu_i$

$$\underset{\pi_i}{\mu_i} = h(\eta_i) = \frac{\exp(\eta_i)}{1+\exp(\eta_i)} \quad \left(= \frac{1}{1+\exp(-\eta_i)} \right)$$

$\mu_i = \pi_i$

response function: $h(\eta_i)$

logistic function: $\frac{\exp(\eta_i)}{1+\exp(\eta_i)}$

Inverse of response function = link function

$\eta_i = g(\mu_i) = h^{-1}(\mu_i)$: What is this for our model?

$$\mu = \frac{e^\eta}{1+e^\eta} \iff (1+e^\eta)\mu = e^\eta$$

response

$$\mu + \mu e^\eta = e^\eta$$

$$\mu = e^\eta - \mu e^\eta = (1-\mu)e^\eta$$

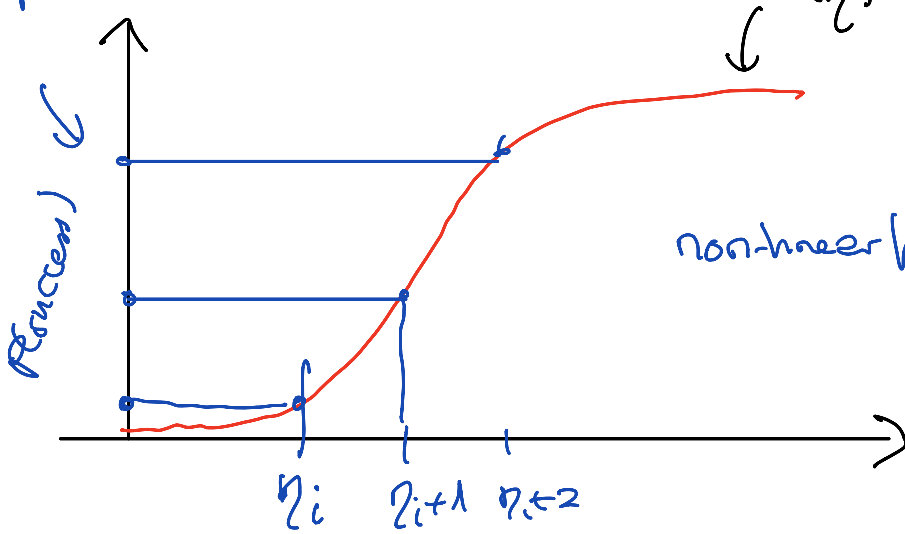
$$e^\eta = \frac{\mu}{1-\mu}$$

logit link function

$$\eta = \ln\left(\frac{\mu}{1-\mu}\right) \Rightarrow \eta_i = \ln\left(\frac{\mu_i}{1-\mu_i}\right)$$

Interpret response function:

$\mu = \pi$



$h(\eta) = \dfrac{e^{\eta}}{1+e^{\eta}}$ = response function

non-linear function

$\eta = \beta_0 + \beta_1 x$

linear predictn

response function: $\dfrac{e^{\eta_i}}{1+e^{\eta_i}}$ ← plot this

$$\eta_i \rightleftarrows \mu_i \; (= \pi_i)$$

$(-\infty, \infty)$ $\in [0,1]$

link function: $\ln\left(\dfrac{\pi_i}{1-\pi_i}\right)$

Now – use link function (logit) to aid in interpreting β's, but first: What is odds?

$$\text{Odds} = \frac{P(Y_i = 1)}{P(Y_i = 0)} = \frac{\pi_i}{1 - \pi_i}$$

$\pi_i = \frac{1}{5} \Rightarrow \text{odds} = \frac{\frac{1}{5}}{\frac{4}{5}} = \frac{1}{4}$, etc.

How to interpret a logit model?

$$\overset{\mu_i}{\overset{\parallel}{\pi_i}} = \frac{\exp(\eta_i)}{1+\exp(\eta_i)} \iff \eta_i = \ln\left(\frac{\pi_i}{1-\pi_i}\right)$$

$$\eta_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik}$$

$$\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik} = \ln\left(\frac{\pi_i}{1-\pi_i}\right)$$

(take exp)

$$(*) \quad \exp(\beta_0)\cdot\exp(\beta_1 x_{i1})\cdots\exp(\beta_k x_{ik}) = \underbrace{\frac{\pi_i}{1-\pi_i}}_{}$$

odds

multiplicative model

So, what if I increase x_{i1} to $(x_{i1}+1)$ $\Rightarrow$ what happen with $(*)$?

$$\exp(\beta_0)\cdot\underbrace{\exp(\beta_1 x_{i1})\cdot\exp(\beta_1)}_{\exp(\beta_1(x_{i1}+1))}\cdot\exp(\beta_2 x_{i2})\cdots = \underset{\frac{\pi_i}{1-\pi_i}}{\bullet}\cdot\exp(\beta_1)$$

$\Rightarrow$ odds multiplied by $\underline{\underline{e^{\beta_1}}}$.

$\beta_1 = 0 \Rightarrow$ no change $e^0 = 1$

$\beta_1 < 0 \Rightarrow e^{neg}$ decrease in odds

$\beta_1 > 0 \Rightarrow e^{pos}$ increase in odds

Ex: infant respiratory disease (see example on Module page)

Now: how to interpret the $\hat{\beta}$'s in this example

1) sex Girl: odds boy $\cdot$ $e^{\hat{\beta}}$ = odds girl

getting disease

$\hat{\beta} = -0.3126$

$\exp(\hat{\beta}) = 0.73$

0.73 $\hat{\beta} < 0 \Rightarrow$ decrease.

decrease in odds from boy to girl

food Breast : reference category is bottle

$\hat{\beta} = -0.6693$

$\exp(\hat{\beta}) = 0.51$

odds bottle $\cdot$ $e^{\hat{\beta}}$ = odds breast

0.51

$\Rightarrow$ decrease in odds from bottle to breast

food Suppl

$\hat{\beta} = -0.1725$: odds bottle $\cdot$ $e^{\hat{\beta}}$ = odds Suppl.

$\exp(\hat{\beta}) = 0.84$ 0.84

Now - turn to prediction & interpretation

Six possible covariate patterns:

1: boy & bottle
2: boy & suppl
3: boy & breast
4: girl & bottle
5: girl & suppl
6: girl & breast

Prediction: $\hat{\pi} = \dfrac{\exp(\hat\beta_0 + \hat\beta_1 x_1 + \hat\beta_2 x_2 + \hat\beta_3 x_3)}{1 + \exp(\hat\beta_0 + \hat\beta_1 x_1 + \hat\beta_2 x_2 + \hat\beta_3 x_2)}$

x_1: 0 = boy, 1 = girl

x_2: 0 = bottle 1 = suppl

x_3: 0 = bottle 1 = breast

Least favourable : $\hat\pi = 0.166$ for boys & bottle

Most favourable : $\hat\pi = 0.069$ girls & breast

* Mountains ← bin, logit

fit, eval., test, goodness of fit

predict new x 2

• Football Poisson w/ pzbn each MLR

↑

optim

which & my val

$\boxed{\begin{array}{l} H - A \\ \text{\# win} \quad 2 \times K \text{ table} \end{array}}$