Introduction to statistical learning

3.3 Logistic regression

V. Lefieux

June 2018

Issue
Idea
Model
Estimation
Parameters interpretation
References

Table of contents

Issue

Idea

Model

Estimation

Parameters interpretation

Table of contents

Issue

Idea

Model

Estimation

Parameters interpretation

Issue

- Explain the presence/absence of cardiovascular disease (CHD) by age X of patients.
- Predict the status of a machine tool (on/off) as a function of its age, for predictive maintenance purposes, for example.
- Predict the probability of repayment of a loan (credit scoring).
- More generally, predict a qualitative variable (binary here) based on quantitative variables.

Issue
Idea
Model
Estimation
Parameters interpretation
References

Table of contents

Issue

Idea

Model

Estimation

Parameters interpretation

An example: cardiovascular disease

Y : with or without cardiovascular disease (CHD)

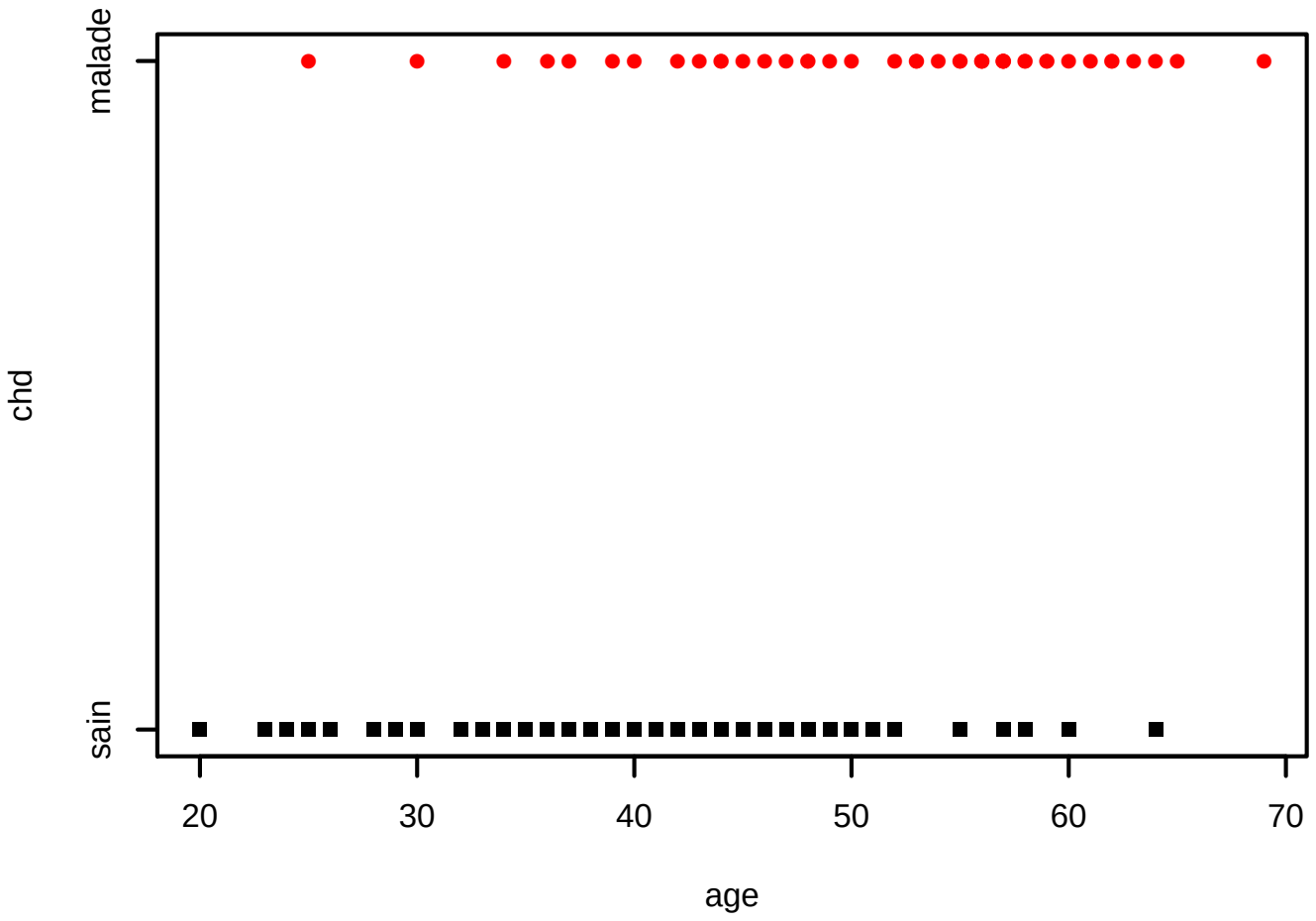
X : patient's age

$n = 100$ observations

Id	age	chd
1	20	healthy
2	23	healthy
3	24	healthy
⋮		⋮
98	64	diseased
99	65	diseased
100	69	diseased

Issue
Idea
Model
Estimation
Parameters interpretation
References

Graphical representation



Issue
Idea
Model
Estimation
Parameters interpretation
References

Y binary variable

Here the variable Y takes 2 values, we model it by:

$$Y\,/X = x \sim \mathcal{B}\left(p(x)\right) \ .$$

Here:

$$\mathbb{P}\left(Y = 1\,/X = x\right) = p(x) \ ,$$
$$\mathbb{P}\left(Y = 0\,/X = x\right) = 1 - p(x) \ .$$

So:

$$\mathbb{E}\left(Y\,/X = x\right) = p(x) \ ,$$
$$\mathrm{Var}\left(Y\,/X = x\right) = p(x)\left[1 - p(x)\right] \ .$$

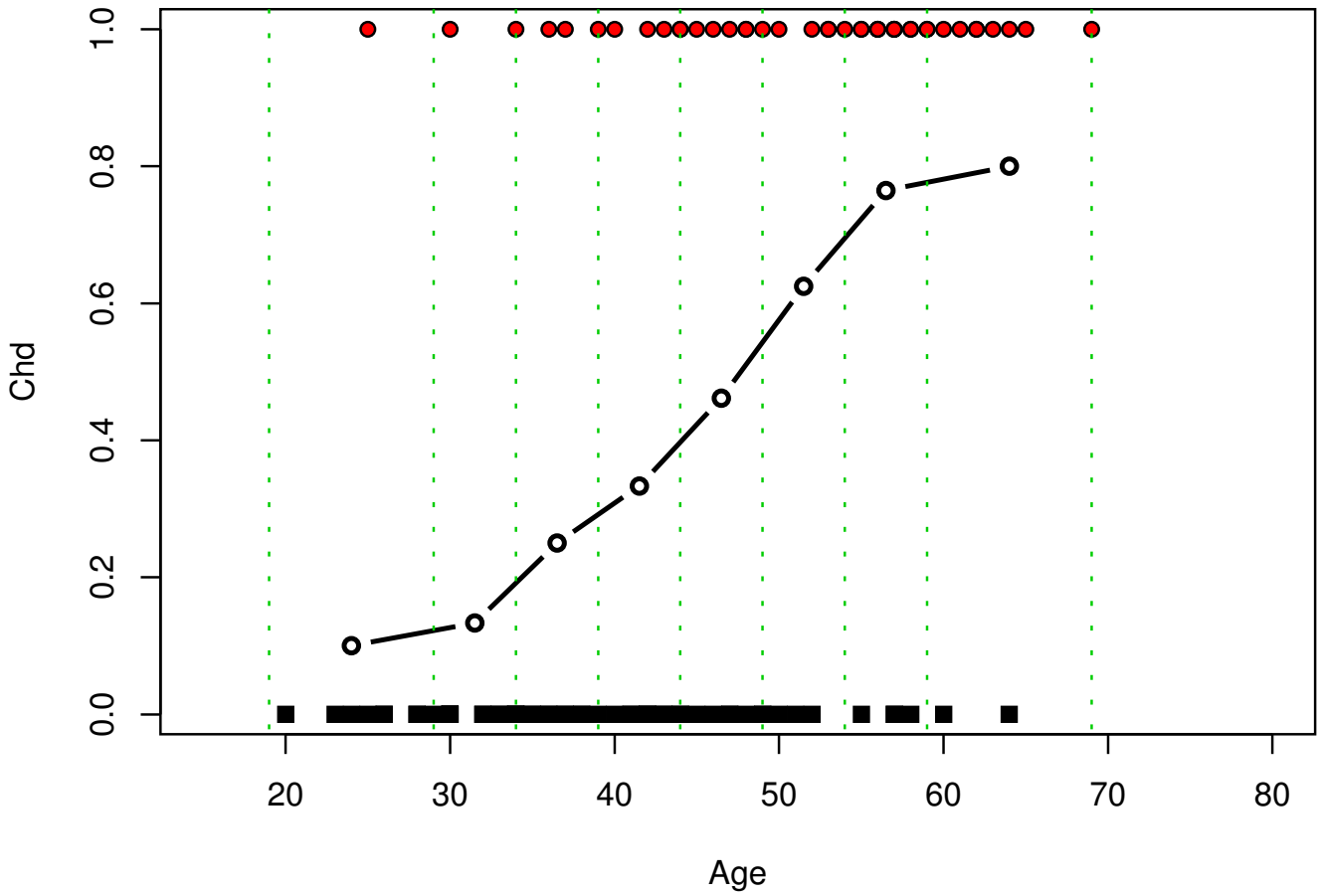
Issue
Idea
Model
Estimation
Parameters interpretation
References

A first idea: the division into age classes

Age	n	Healthy	Diseased	Average
$[19, 29[$	10	9	1	.10
$[29, 34[$	15	13	2	.13
$[34, 39[$	12	9	3	.25
$[39, 44[$	15	10	5	.33
$[44, 49[$	13	7	6	.46
$[49, 54[$	8	3	5	.625
$[54, 59[$	17	4	13	.76
$[59, 69[$	10	2	8	.8

Issue
Idea
Model
Estimation
Parameters interpretation
References

Graphical representation



Issue
Idea
Model
Estimation
Parameters interpretation
References

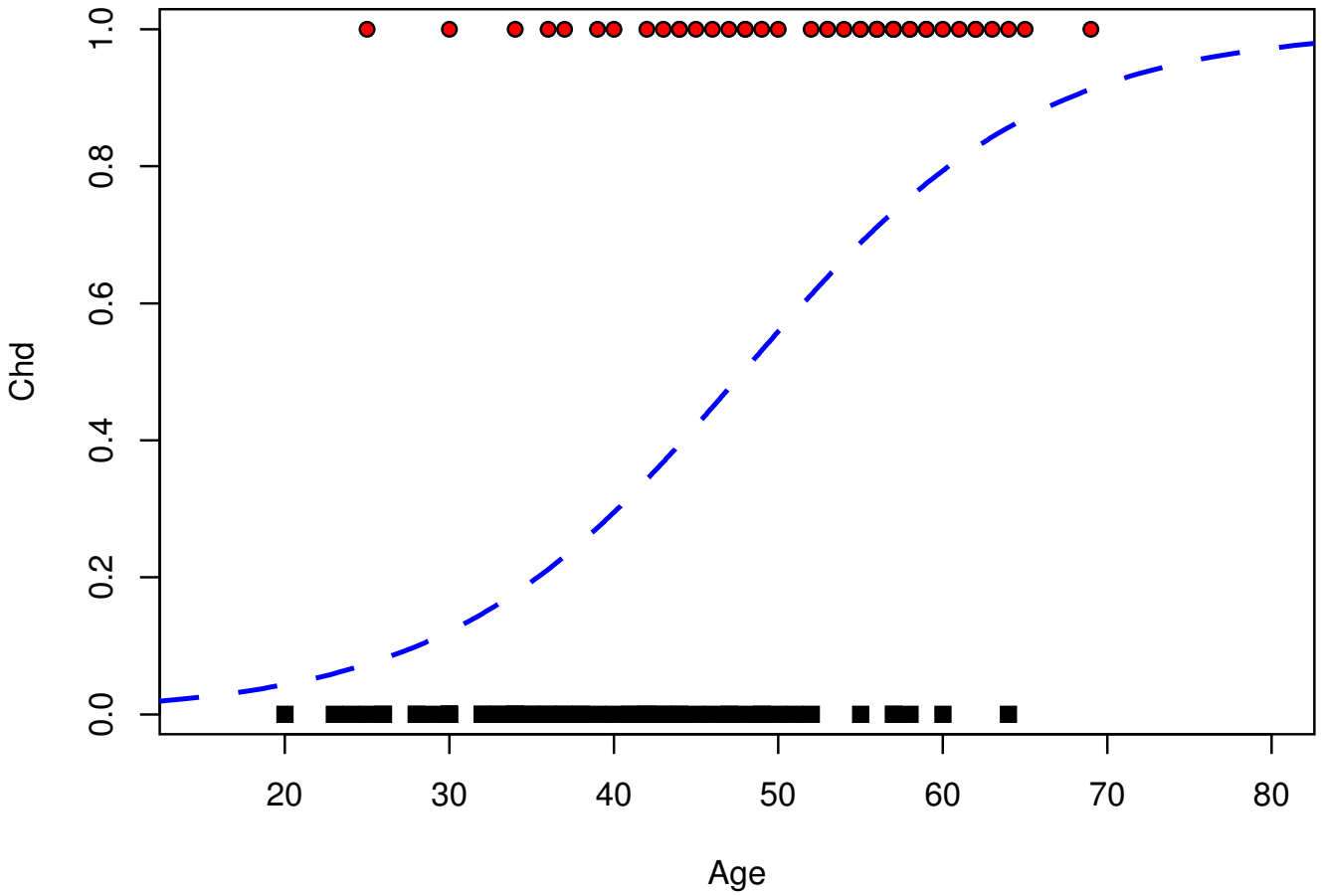
To go further

We would like to find a function:

- smoother,
- that uses all the data.

Issue
Idea
Model
Estimation
Parameters interpretation
References

Expected function

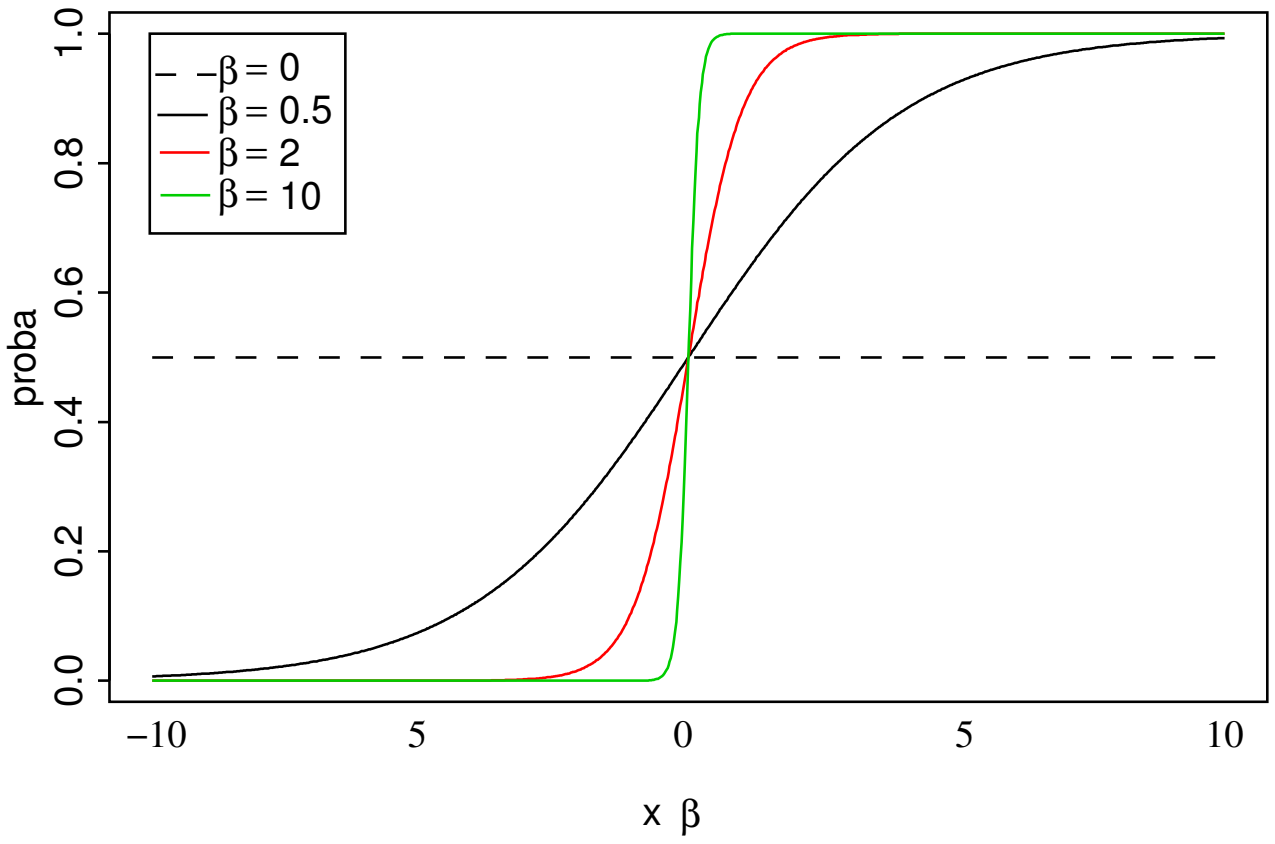


Issue
Idea
Model
Estimation
Parameters interpretation
References

Equation of an S curve

A first way to get a S curve is to consider the function :

$$x \mapsto \frac{\exp\left(x^\top \beta\right)}{1 + \exp\left(x^\top \beta\right)} \ .$$



Issue
Idea
Model
Estimation
Parameters interpretation
References

Comparison with the linear model

In the linear model:

$$\mathbb{E}\left(Y / X = x\right) = x^\top \beta \ .$$

When Y is binary:

$$\mathbb{E}\left(Y / X = x\right) = p(x) \in [0, 1]$$

but it is possible to consider g link functions such as:

$$g\left(p(x)\right) = x^\top \beta \ .$$

Issue
Idea
Model
Estimation
Parameters interpretation
References

Table of contents

Issue

Idea

Model

Estimation

Parameters interpretation

The logit function

Let's consider:

$$\mathbb{E}\left(Y\,/X=x\right)=p(x)=\frac{\exp\left(x^\top\beta\right)}{1+\exp\left(x^\top\beta\right)}\ .$$

The following function, named logit:

$$g(p)=\ln\left(\frac{p}{1-p}\right)$$

is bijective (derivable) and:

$$g\left(p(x)\right)=\ln\left(\frac{p(x)}{1-p(x)}\right)=x^\top\beta\ .$$

This function is used in logistic regression.
The outstanding issue is the estimation of the *beta* parameter.

Issue
Idea
Model
Estimation
Parameters interpretation
References

Table of contents

Issue

Idea

Model

Estimation

Parameters interpretation

Maximum likelihood estimate I

Let's consider the likelihood:

$$p\left(y_1, \ldots, y_n; \beta\right) = \prod_{i=1}^{n} \mathbb{P}\left(Y = y_i / X = x_i\right) \ .$$

So:

$$p\left(y_1, \ldots, y_n; \beta\right) = \prod_{i=1}^{n} p\left(x_i\right)^{y_i} \left(1 - p\left(x_i\right)\right)^{1-y_i} \ .$$

The log-vraisemblance is:

$$\ell\left(y_1, \ldots, y_n; \beta\right) = \sum_{i=1}^{n} \left\{y_i \ln\left[p\left(x_i\right)\right] + \left(1 - y_i\right) \ln\left[1 - p\left(x_i\right)\right]\right\} \ .$$

We obtain:

$$\ell\left(y_1, \ldots, y_n; \beta\right) = \sum_{i=1}^{n} \left\{y_i x_i^\top \beta - \ln\left[1 + \exp\left(x_i^\top \beta\right)\right]\right\} \ .$$

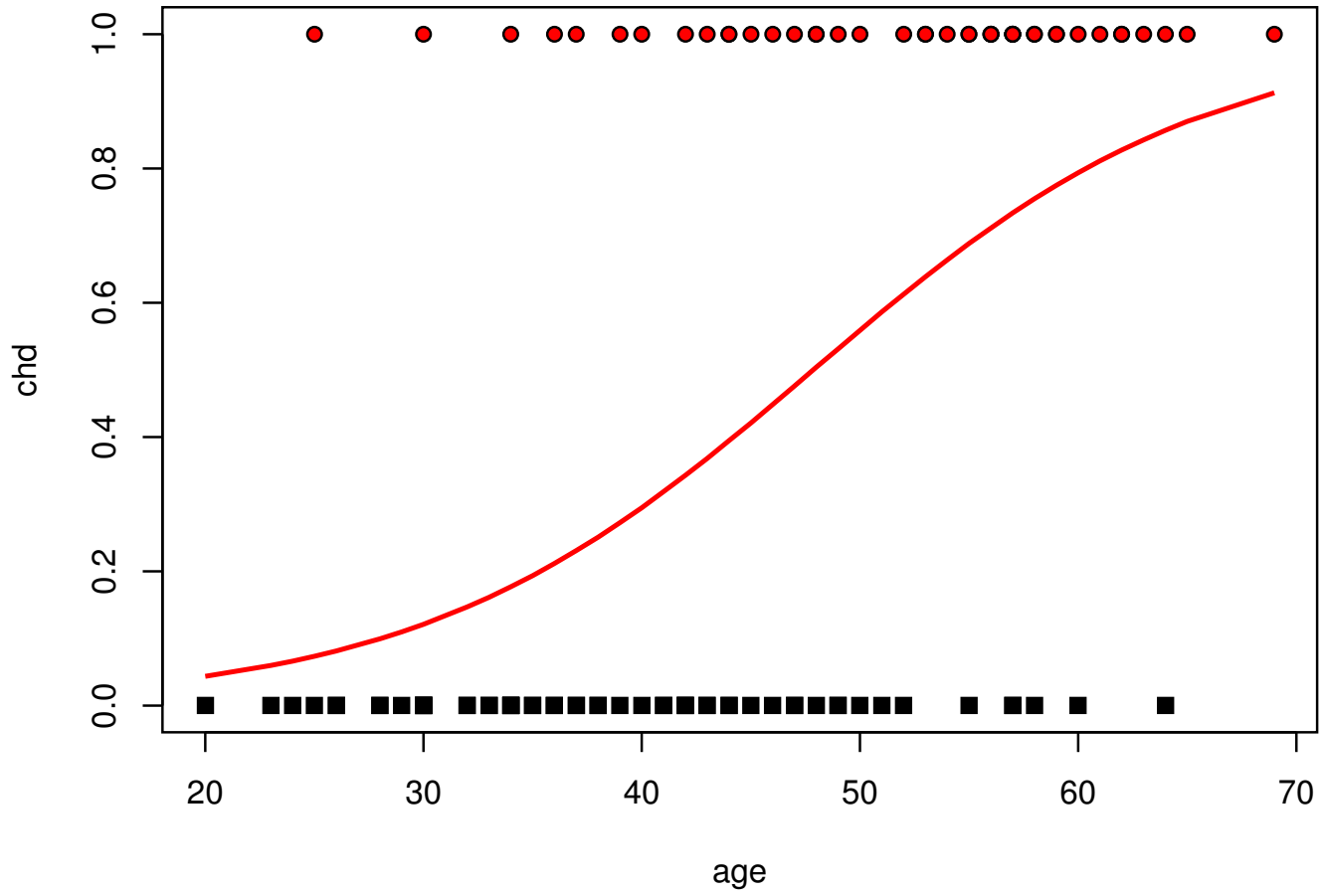
Maximum likelihood estimate II

- Partial derivatives are calculated and cancelled to obtain the normal equations.
- Unfortunately, there are no explicit solutions for $\widehat{\beta}$.
- Likelihood has (usually) a single maximum, and there are iterative numerical algorithms to obtain this maximum, including :
  - Newton's algorithm,
  - Fisher's score algorithm.

Issue
Idea
Model
Estimation
Parameters interpretation
References

Back to introductory example

$$\widehat{\mathbb{P}}(CHD = 1 \,/\, age) = \frac{\exp(-5.30945 + 0.11092 \times age)}{1 + \exp(-5.30945 + 0.11092 \times age)} \; .$$



Issue
Idea
Model
Estimation
Parameters interpretation
References

Table of contents

Issue

Idea

Model

Estimation

Parameters interpretation

Direct interpretation

When β_j associated with the variable X_j is:

- Positive : if X_j increases then p increases.
- Negative : if X_j increases then p decreases.

In the example $\widehat{\beta}_{age} = 0.11$, so the probability of CHD increases with age.

Issue
Idea
Model
Estimation
Parameters interpretation
References

Interpretation by odd-ratios I

The odds (of an event) is the ratio of the probability of an event to its complementary.

Let's imagine a game whose probability of winning is $p = \frac{1}{20}$. The odd is worth:

$$\frac{p}{1-p} = \frac{\frac{1}{20}}{1-\frac{1}{20}} = \frac{1}{19} \ .$$

On average, there is 1 winner for 19 losers.

Issue
Idea
Model
Estimation
Parameters interpretation
References

Interpretation by odd-ratios II

For an individual x, the odd to get the answer $Y = 1$ is defined by:

$$odd(x) = \frac{p(x)}{1 - p(x)}$$

where $p(x) = \mathbb{P}\left(Y = 1 \,/ X = x\right)$.

The odds ratio (OR) between 2 individuals x and x' is:

$$OR = \frac{odd(x)}{odd(x')} = \frac{\frac{p(x)}{1-p(x)}}{\frac{p(x')}{1-p(x')}}$$

Issue
Idea
Model
Estimation
Parameters interpretation
References

Interpretation by odd-ratios II

For 2 categories j and j':

1. Interpretation rules of odd-ratios

$$
\begin{aligned}
OR_{j,j'} > 1 &\Leftrightarrow p_j > p_{j'} \ , \\
OR_{j,j'} = 1 &\Leftrightarrow p_j = p_{j'} \ , \\
OR_{j,j'} < 1 &\Leftrightarrow p_j < p_{j'} \ .
\end{aligned}
$$

2. Odd-ratios and relative risk (or risk ratio: RR)
   If $p_{j'} = p_j \approx 0$ then:

$$
OR_{j,j'} \approx p_j / p_{j'} := RR_{j,j'} \ .
$$

3. Odd-ratios and parameters:

$$
\begin{aligned}
OR_{j,\mathrm{ref}} &= \exp\left(\beta_j \left(x_j - x_{\mathrm{ref}}\right)\right) \\
OR_{j,j'} &= \exp\left(\beta_j - \beta_{j'}\right)
\end{aligned}
$$

References

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The elements of statistical learning. Data Mining, inference, and prediction*. Springer Series in Statistics. Springer, 2 edition.

McCullagh, P. and Nelder, J. A. (1989). *Generalized linear models*. Monographs on Statistics & Applied Probability (37). Chapman & Hall.

Issue
Idea
Model
Estimation
Parameters interpretation
References