Logistic Regression

DR.ALAA MOHAMMED

3RD LEVEL, 2ND SEMESTER, BIOMEDICAL INSTRUMENTATION AND BIOMECHANICS BRANCHES, BIOMEDICAL ENG. DEPARTMENT

General linear models (GLM)

The most commonly used statistical tests are developed from models that assume independent responses and follow a normal distribution with a constant variance.

The mean is the deterministic component of these models and may be expected to vary according to the independent variables in the model.

The variance and parameters of the deterministic component are usually unknown. The term, general linear model (GLM), refers to linear models having the normal distribution as the stochastic component.

General linear models

All generalized linear models have the following three characteristics:

- **1** A probability distribution describing the outcome variable
- **2** A linear model

$$\eta = \beta_0 + \beta_1 X_1 + \cdots + \beta_n X_n$$

- **3** A link function that relates the linear model to the parameter of the outcome distribution

$$g(p) = \eta \text{ or } p = g^{-1}(\eta)$$

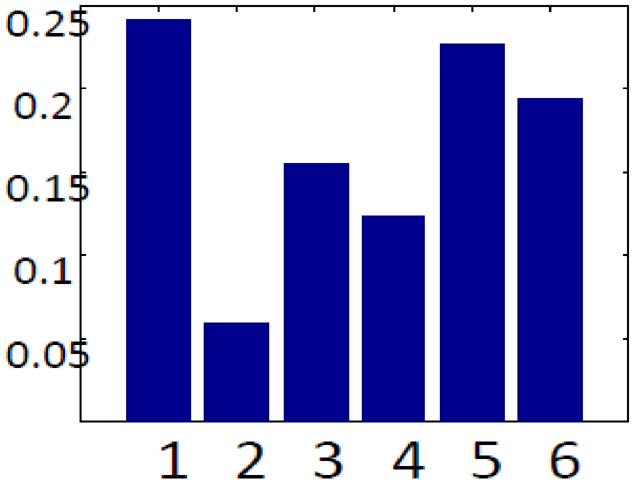
- **Logistic regression** is just one example of **generalized linear models** (**GLMs**). .

Discrete Probability Distributions

- The random variables only take on **discrete** values

  e.g., throwing dice: possible values

$$v_i \in \{1,2,3,4,5,6\}$$



- The probabilities sum to 1

$$\sum_i P(v_i) = 1$$

- Discrete distributions are particularly important in classification
- Probability Mass Function or Frequency Function (normalized histogram)

Classic Discrete Distributions (I)

- **Bernoulli Distribution**

If a random variable X can take on only two values (labelled $0 =$ "failure" and $1 =$ "success"), then the distribution of X is, by definition, Bernoulli(p), where p is the probability of success.

$P(0) = p$ and $P(1) =$ 1-p, or in compact notation:

$$P(x) = \begin{cases} p^x(1-p)^{1-x}, \text{if } x = 0 \text{ or } x = 1 \\ 0, \text{ otherwise} \end{cases}$$

Bernoulli distributions are naturally modeled by sigmoidal activation functions in neural networks with binary inputs.

Classic Discrete Distributions (II)

Binomial Distribution

Like Bernoulli distribution, binary input variables: 0 or 1, and probability $P(0) = p$ and $P(1) = 1\text{-}p$

What is the probability of *k* successes, *P*(*x*), in a series of *n* independent trials? $(n \geq x)$

$P(x)$ is a binomial random variable:

$$p(x) = P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \qquad x = 0, 1, \ldots, n.$$

$$\binom{\boldsymbol{n}}{\boldsymbol{x}} = \frac{\boldsymbol{n}!}{\boldsymbol{x}!(\boldsymbol{n}-\boldsymbol{x})!}$$

Binomial variables are important for density estimation networks, e.g. “what is the probability that *k* data points fall into region R?”

Bernoulli distribution is a subset of binomial distribution (i.e., n=1)

Classic Discrete Distributions (III)

Geometric Distribution

If we again observe *n* independent trials, each of which results in “success" with probability *p* and “Failure" with probability $1-p$, then the number of *trials* required to observe the first success, X has, by definition, a Geometric(*p*) distribution.

Specifically, the pmf of *X* is:

$$p(x) = P(X = x) = (1-p)^{x-1}p, \qquad x = 1, 2, \ldots .$$

Logistic Regression

Logistic regression is a predictive analysis like all regression analysis.

Logistic regression is a **GLM** used to model a binary categorical variable using numerical and categorical predictors.

Logistic regression analysis studies the association between a categorical dependent variable and a set of independent (explanatory) variables.

Description of the relationship between an outcome and risk factor(s)

The name logistic regression is used when the dependent variable has only two values, such as 0 and 1 or Yes and No.

The name multinomial logistic regression is usually reserved for the case when the dependent variable has three or more unique values

Logistic regression is a predictive analysis like all regression analysis.

The Logit and Logistic Transformations

In logistic regression, a mathematical model of a set of explanatory variables is used to predict a logit transformation of the dependent variable

Suppose the numerical values of **0** and **1** are assigned to the two outcomes of a binary variable.

- Often, the **0** represents a negative response and the **1** represents a positive response.
- The mean of this variable will be the proportion of positive responses. If $\boldsymbol{p}$ is the proportion of observations with an outcome of **1**, then $\mathbf{1}-\boldsymbol{p}$ is the probability of a outcome of **0**.
- The ratio $\boldsymbol{p}/(\mathbf{1}-\boldsymbol{p})$ is called the odds and the logit is the logarithm of the odds, or just log odds.

Logistic Function

- Mathematically, the logit function is written

Logit function

$$logit(p) = \log\left(\frac{p}{1-p}\right), \text{ for } 0 \leq p \leq 1$$

- The **logit** function takes a value between 0 and 1 and maps it to a value between ∞ and ∞.

- The inverse **logit** function takes a value between $-\infty$ and ∞ and maps it to a value between 0 and 1.

Inverse logit (logistic) function

$$g^{-1}(x) = \frac{\exp(x)}{1+\exp(x)} = \frac{1}{1+\exp(-x)}$$

This formulation also has some use when it comes to interpreting the model as logit can be interpreted as the log odds of a success.

The logistic regression model

The three GLM criteria give us:

$$y_i \sim \text{Binom}(p_i)$$

$$\eta = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n$$

$$\text{logit}(p) = \eta$$

From which we arrive at,

$$p_i = \frac{\exp(\beta_0 + \beta_1 x_{1,i} + \cdots + \beta_n x_{n,i})}{1 + \exp(\beta_0 + \beta_1 x_{1,i} + \cdots + \beta_n x_{n,i})}$$

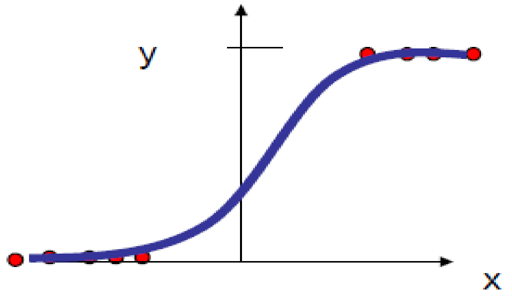
The logistic regression model

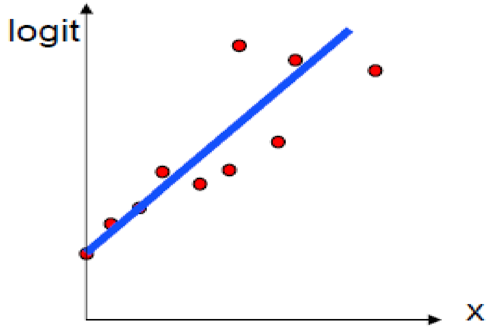
Logit

$$p_i = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_i)}}$$

$$p_i = \frac{e^{\beta_0 + \beta_1 x_i}}{e^{\beta_0 + \beta_1 x_i} + 1}$$

$$\log\left[\frac{p_i}{1 - p_i}\right] = \beta_0 1 + \beta_1 x_i$$





- The linear logistic regression model is defined as:

$$\text{logit}(\mathrm{p}) = \alpha + \beta X$$

$$\log\left(\frac{p}{1-p}\right) = \alpha + \beta X$$

That means:

$$p = \frac{e^{\alpha+\beta X}}{1+e^{\alpha+\beta X}}$$

- α is the log odds of the outcome when X = 0
- β is log odds ratio (OR) associated with one unit increase in X

Advantages of logistic regression model

- Probability of outcome varies with values of X
- Regression coefficient can be interpreted as log odds ratio
- Can be applied to many research designs
- Many software implement the linear logistic regression model

Odds

Odds are another way of quantifying the probability of an event, commonly used in gambling (and logistic regression).

Odds

For some event E,

$$\text{odds}(E) = \frac{P(E)}{P(E^c)} = \frac{P(E)}{1 - P(E)}$$

Similarly, if we are told the odds of E are x to y then

$$\text{odds}(E) = \frac{x}{y} = \frac{x/(x+y)}{y/(x+y)}$$

which implies

$$P(E) = x/(x+y), \quad P(E^c) = y/(x+y)$$

Risk: probability (P) of an event during a period

Odds: ratio of probability of success over the probability of failure:

$$\boldsymbol{Odds = P/(1-P)}$$

- n =5 patients followed, 1 patient suffered from stroke:

$$\boldsymbol{P = 1/5 = 0.2}$$

Odds=0.2/0.8=0.25

Example: Donner Party

The time the last survivor was rescued on April 21, 1847, 40 of the 87 members had died from famine and exposure to extreme cold.

	Age	Sex	Status
1	23.00	Male	Died
2	40.00	Female	Survived
3	40.00	Male	Survived
4	30.00	Male	Died
5	28.00	Male	Died
⋮	⋮	⋮	⋮
43	23.00	Male	Survived
44	24.00	Male	Died
45	25.00	Female	Survived

Example (cont.)

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	1.8185	0.9994	1.82	0.0688
Age	-0.0665	0.0322	-2.06	0.0391

Model:

$$\log\left(\frac{p}{1-p}\right) = 1.8185 - 0.0665 \times \text{Age}$$

Odds / Probability of survival for a newborn (Age=0):

$$\log\left(\frac{p}{1-p}\right) = 1.8185 - 0.0665 \times 0$$

$$\frac{p}{1-p} = \exp(1.8185) = 6.16$$

$$p = 6.16/7.16 = 0.86$$

Example (cont.)

Model:

$$\log\left(\frac{p}{1-p}\right) = 1.8185 - 0.0665 \times \text{Age}$$

Odds / Probability of survival for a 25 year old:

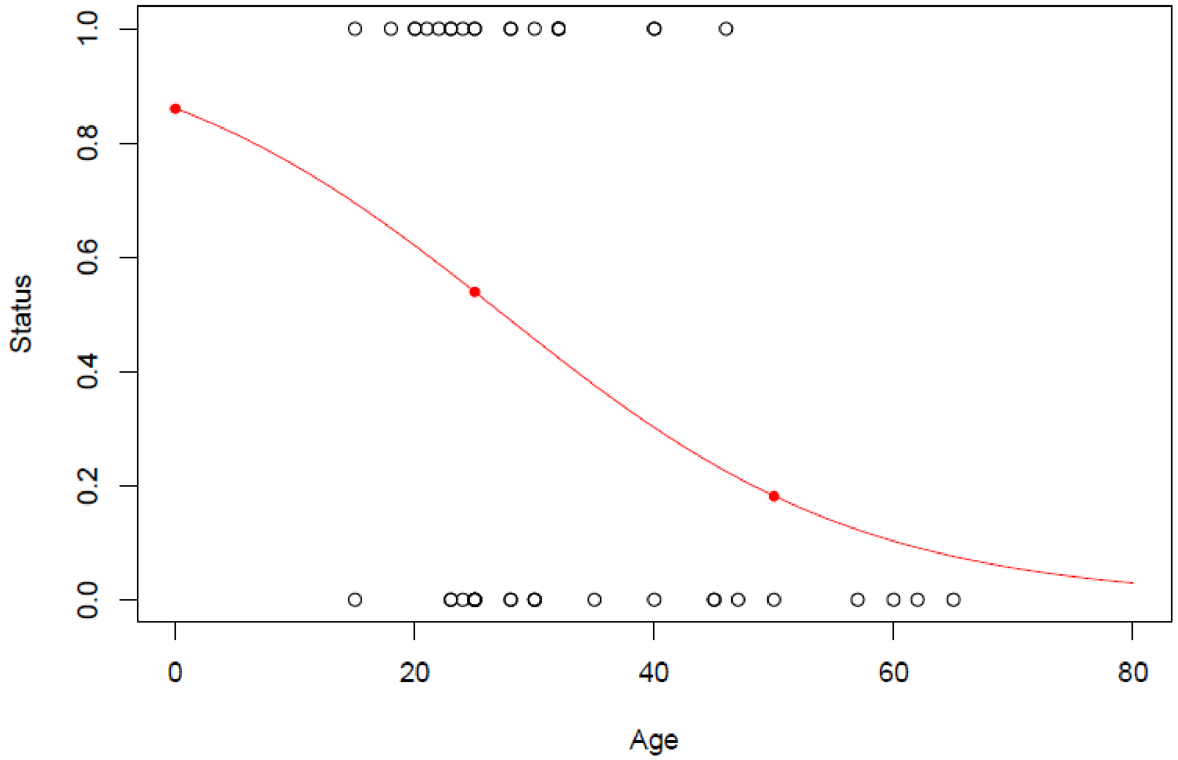
$$\begin{aligned}\log\left(\frac{p}{1-p}\right) &= 1.8185 - 0.0665 \times 25 \\ \frac{p}{1-p} &= \exp(0.156) = 1.17 \\ p &= 1.17/2.17 = 0.539\end{aligned}$$

Odds / Probability of survival for a 50 year old:

$$\begin{aligned}\log\left(\frac{p}{1-p}\right) &= 1.8185 - 0.0665 \times 0 \\ \frac{p}{1-p} &= \exp(-1.5065) = 0.222 \\ p &= 0.222/1.222 = 0.181\end{aligned}$$

Example (cont.)

$$\log\left(\frac{p}{1-p}\right) = 1.8185 - 0.0665 \times \text{Age}$$



Example (cont.)

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	1.8185	0.9994	1.82	0.0688
Age	-0.0665	0.0322	-2.06	0.0391

Simple interpretation is only possible in terms of log odds and log odds ratios for intercept and slope terms.

Intercept: The log odds of survival for a party member with an age of 0.

From this we can calculate the odds or probability, but additional calculations are necessary.

Slope: For a unit increase in age (being 1 year older) how much will the log odds ratio change, not particularly intuitive.

More often then not we care only about sign and relative magnitude.

Example (cont.)

$$\log\left(\frac{p_1}{1-p_1}\right) = 1.8185 - 0.0665(x+1)$$
$$= 1.8185 - 0.0665x - 0.0665$$
$$\log\left(\frac{p_2}{1-p_2}\right) = 1.8185 - 0.0665x$$

$$\log\left(\frac{p_1}{1-p_1}\right) - \log\left(\frac{p_2}{1-p_2}\right) = -0.0665$$
$$\log\left(\frac{p_1}{1-p_1}\Big/\frac{p_2}{1-p_2}\right) = -0.0665$$
$$\frac{p_1}{1-p_1}\Big/\frac{p_2}{1-p_2} = \exp(-0.0665) = 0.94$$

Example 1 (HW)

Table 1 is a cross tabulation of two binary variables for a sample of 172 boys in reception classes.

- Whether or not the child is *perceived* by their teacher to have a behavior problem (which we will later model as the response).
- Ethnic group (which we will later model as the explanatory variable).

Table 1

	Ethnic Group		
Behaviour Problems	White	Black	Total
NO	90 [0.83]	30 [0.48]	120 (70%)
YES	19 [0.17]	33 [0.52]	52 (30%)
Total	109 (63%)	63 (37%)	172 (100%)

Example 1 (cont.)

We can see that the majority of the sample of boys (70%) are not perceived to have a behavior problem and that 63% of them are white.

The conditional probabilities of having a behavior problem, given ethnic group are shown in square brackets after each of the cell frequencies.

For example the probability of being perceived to have a behavior problem for white boys is 0.17, and for black boys is 0.52.

Example 1 (cont.)

- **Odds**

In the above table, the odds of a white boy being seen to have a behavior problem are 19/90 = 0.21 or 0.21 to 1. In betting terms that is about 5 : 1 against – much less than even money.

For black boys, the corresponding odds are 33/30 = 1.1, or 1.1 to 1.

Equivalent to 11 to 10 on, (or a little better than even money.). Note that odds are not the same as probabilities – they are not restricted to the range 0 to 1.

Example 1 (cont.)

- **logistic regression model**

We can fit a logistic regression model to the data in Table 1.

We get:

$$\boldsymbol{Logit\ P\ =\ -1.56\ +\ 1.65EG}$$

Which we can interpret as the log odds of a white boy (EG=0) seen as having a behavior problem being equal to –1.56, hence the odds of a white boy having a behavior problem are:

$$\exp(-1.56)\ =\ 0.21$$

The log odds of a black boy (EG=1) having a perceived behaviour problem are

$$-1.56 + 1.65 = 0.09$$

Hence the odds of a black boy having a perceived behavior problem are $\boldsymbol{exp}(\mathbf{0.09})$ $= \mathbf{1.1}$ Alternatively we can say that the odds for black boys are

$\boldsymbol{exp}(\mathbf{1.65}) = \mathbf{5.21}$ times as high as they are for white boys.

That is, the relative odds of a teacher perceiving a black boy to have behavioural problems compared with a black boy are 5.21.

Notice that these results correspond exactly to the results in Table 1.

THANKS FOR LISTENING