Logistic regression

Part 3: Odds ratio and logistic regression

By: Noureddin Sadawi, PhD

University of London

Odds ratio

Odds Ratio: odds of an event in one group divided by the odds of the event in another group.

Ratio of probabilities:

- Odds=1 the event is as likely to occur as not to occur.
- Odds>1 the event is more likely to occur than not to occur.
- Odds<1 the event is more likely not to occur than to occur.

Odds ratio example 1

	No Covid-19	**Covid-19**	**Total**
Smoker	187 (62%) [187/304]	117 (38%) [117/304]	304
Non-smoker	192 (69%) [192/278]	86 (31%) [86/278]	278
Total	379	203	582

- Odds ratio = Odds of Covid-19 positive given non-smoker / Odds of Covid-19 positive given smoker = (86/192) / (117/187) = 0.72.
- The odds of becoming Covid-19 positive for non-smoker are 0.71 the odds for smoker.
- The odds of becoming Covid-19 positive for smoker are 1 / odds ratio
- 1 / 0.72 = about 1.39 times as large as the odds for non-smoker.
- An increase of about 39%.

Odds ratio example 2

Example case control study:

- Odds of survival in the treatment group: 6/5.
- Odds of survival in the control group: 4/7.
- Odds ratio: (4/7)/(6/5) = 0.48.
- The odds of surviving in the control group are **less then half** the odds of surviving in the treatment group.
- Or, an individual in the treatment group has odds about 2.1 as high [(6/5)/(4/7) = 2.1] of surviving than individuals from the control group.

Odds ratio

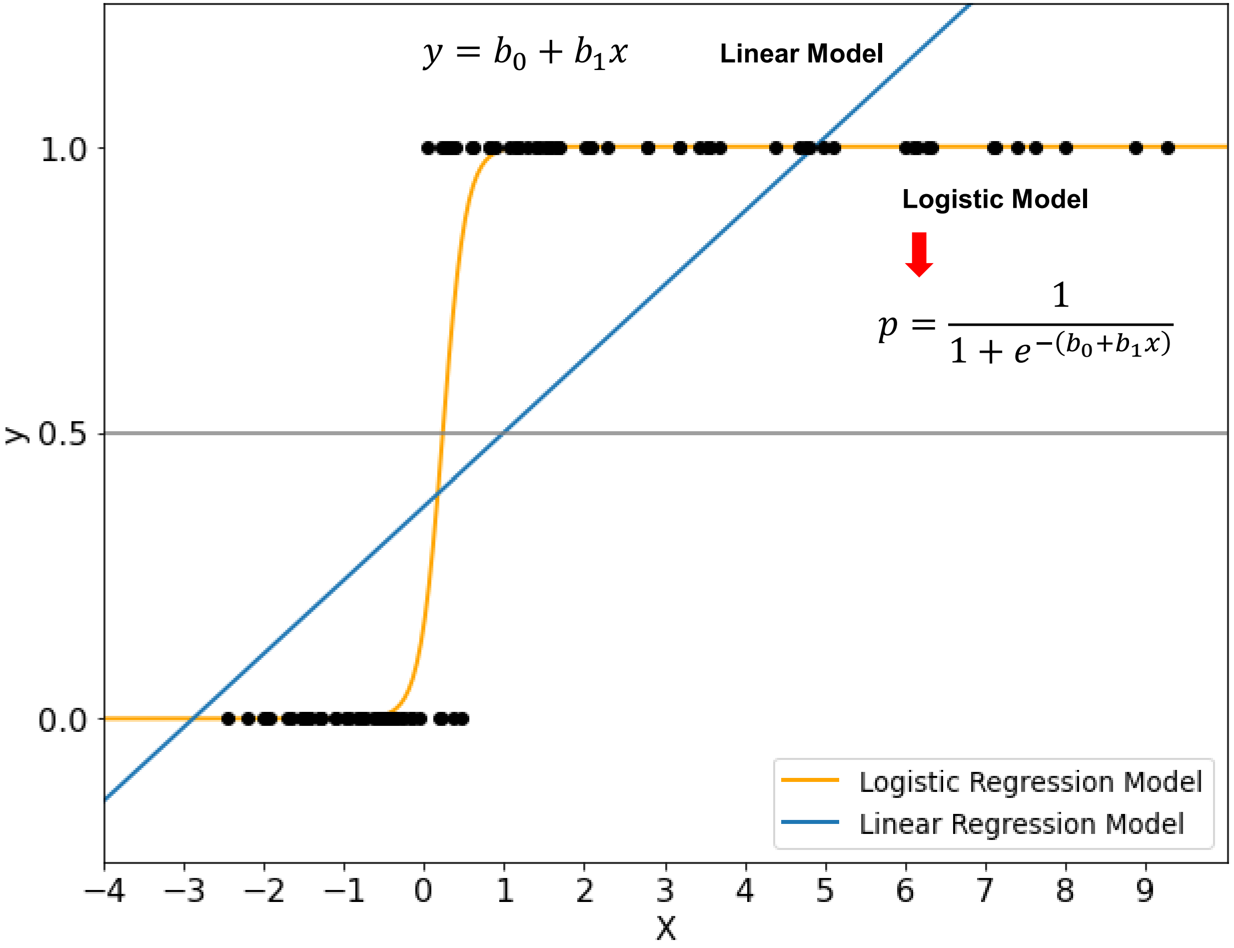
- Odds ratio < 1: odds of success in the first group are lower than in the second group.
- The closer the odds ratio to 0, the lower the odds of the first group to the second.
- Odds ratio $= 1$: the odds of both groups are the same.
- Odds ratio > 1: odds of the first group are higher than the second group.
- The higher the odds ratio, the higher the odds of the first group to the second.

Logistic regression

- Mathematical modelling approach that can be used to describe the relationship between several input variables (i.e. predictors) and a binary outcome.
- In more detail, logistic regression predicts the probability of an outcome that can only have two values (i.e. a dichotomy).
  - The prediction is based on the use of one or several predictors (numerical and categorical).

Logistic regression

- Probability values are always in the range 0 to 1 and linear regression predicts values outside it.
- As the outcome can only have one of two possible values for each data point, the residuals will not be normally distributed around the predicted line.
- Logistic regression produces a logistic curve, which is limited to values between 0 and 1.
- The curve is constructed using the natural logarithm of the 'odds' of the target variable.



Logistic regression

- Combination of natural log transformations and odds ratios.
- Remember: Outcome variable can have **only one of two** values – {0,1}.
- The main idea is to model the probability of being in one of the two categories.
- Logit transformation = Natural log of the odds.
- Logit makes probabilities into odds.
- If *p* is the probability then *p/(1-p)* is the odds.
- The natural logarithm of the odds is the logit of the probability.

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

Logistic regression

- The logistic regression equation can be written in terms of an odds ratio.
- The constant (b_0) moves the curve right and left and the slope (b_1) defines the steepness of the curve.
- The equation can be written in terms of log-odds (logit) which is a linear function of the predictors.
- The coefficient (b_1) is the amount the logit (log-odds) changes with a one unit change in *x*.
- Logistic regression can handle any number of numerical and/or categorical variables.

$$\frac{p}{1-p} = \exp(b_0 + b_1 x)$$

$$ln\frac{p}{1-p} = b_0 + b_1 x$$

$$p = \frac{1}{1 + e^{-(b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_p x_p)}}$$

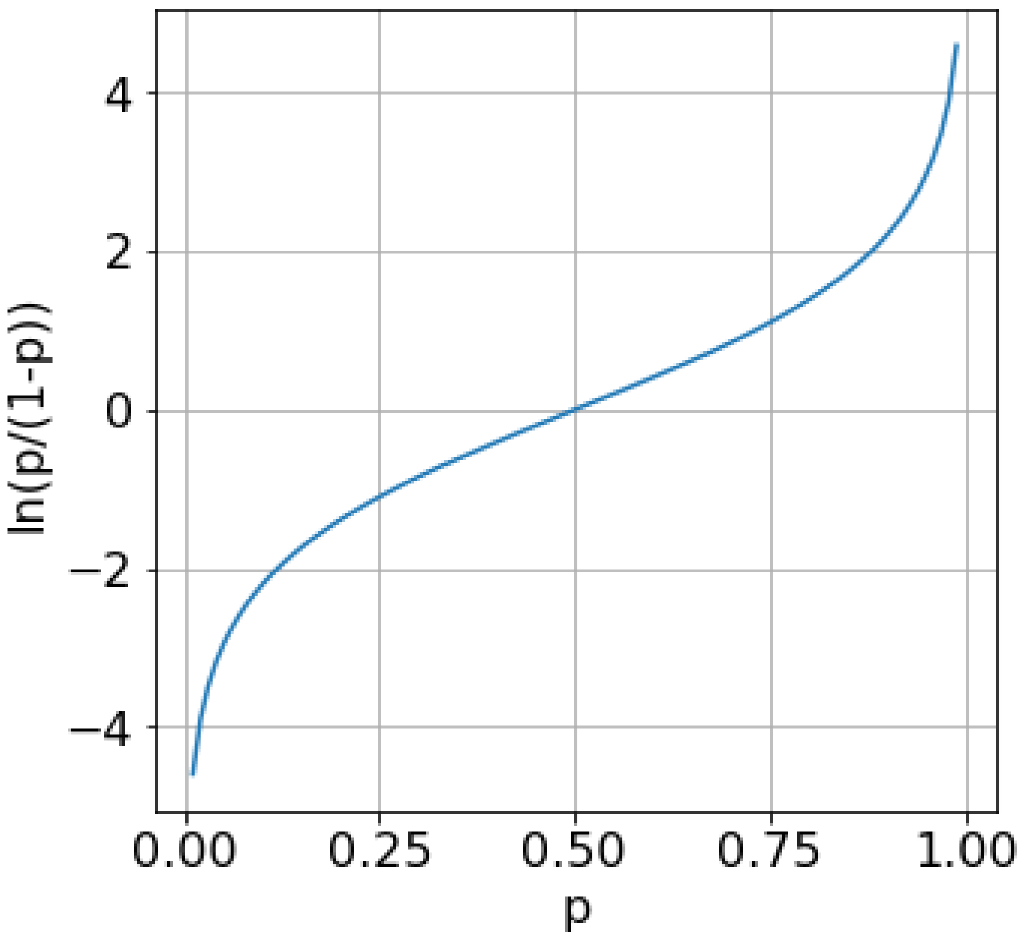
Logistic regression

- Y = f(X)
- logit (p) = f(X)
- ln (odds of event) = ln (p /(1 - p)) = f(X)
- Odds of event = exp(f(X))
- If you invert the logit transformation by exponentiating then you have the odds of the event of interest .
- p /(1 - p) = exp(f(X))
- p = exp(f(X)) / (1+ exp(f(X)))

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

The logit function

- Probability values are always in the range 0 to 1 and linear regression predicts values outside it.
- Notice the y axis is the natural log of the odds.
- The logit at probability 0.5 is 0.
- The logit at probability 0 is -∞.
- The logit at probability 0 is +∞.



Logistic regression

- Outcome variable: result of diabetes test.
- logit(probability of positive diabetes test) =

  -0.268 + 0.018 * BMI

- -0.268: logged odds of being diabetic for BMI=0 (no clinical interpretation).
- 0.018: change in the logged odds of diabetes **per unit** change in BMI.
- exp(0.018)=1.018: for each unit change in BMI the odds of becoming diabetic increase by approximately 1.8%.
- For a 5 unit increase in BMI, the odds of becoming diabetic increases by approx exp(0.018)=(1.018)^5=1.093 (i.e. 9.3%).

BMI	Outcome
28.1	0
43.1	1
31.0	1
30.5	1
30.1	1
…	…
43.3	1
36.5	1
28.4	0
32.9	0
26.2	0