

# Logistic Regression

Example: Phone Service

# Logistic Regression Model

Two equivalent formulations

$$p(Y = 1 | X_1 = v_1, \dots, X_n = v_n) = \frac{\exp(a_0 + a_1 v_1 + \dots + a_n v_n)}{1 + \exp(a_0 + a_1 v_1 + \dots + a_n v_n)}$$

$$\log\left(\frac{p(Y = 1 | X_1 = v_1, \dots, X_n = v_n)}{p(Y = 0 | X_1 = v_1, \dots, X_n = v_n)}\right) = a_0 + a_1 v_1 + \dots + a_n v_n$$

In the second form we use the concept of odds, ratio between probability of success and probability of failure.

This is the “fair” bet versus 1 if we are betting on success.

If  $p(\text{success}) = 0.8$  then  $\text{odds} = 0.8 / 0.2 = 4$ .

Indeed, it is fair to bet 4 against 1 in this case.

Odds and probability are interchangeable, but thinking in terms of odds is often more intuitive.

# A market research

The answer variable  $Y$  is the purchase of a new phone service by a customer.

Predictors are education, mobility and income.

All variables are binary: 0 or 1.

We have already discretized them in some way.

E.g. we have set thresholds on years of scholarship, number of movements from a city to another in the last 10 years, year income.

The dataset is about a sample of 10524 customers.

Each cell of the table shows the number of customers receiving an offer and number of customer accepting it.

	Low education		High education	
	Low mobility	High mobility	Low mobility	High mobility
Low income	153/2160 = 0.071	226/1137 = 0.199	61/886 = 0.069	233/1091 = 0.214
High income	147/1363 = 0.108	139/547 = 0.254	287/1925 = 0.149	382/1415 = 0.270

Let us rewrite the model for readability

$$p(A = 1 | I = i, M = m, R = r) = \frac{\exp(a_0 + a_I I + a_M M + a_R R)}{1 + \exp(a_0 + a_I I + a_M M + a_R R)}$$

With some algebraic manipulation we get

$$\exp(a_0) = \frac{p(A = 1 | I = 0, M = 0, R = 0)}{p(A = 0 | I = 0, M = 0, R = 0)}$$

**= odds of the basic case**

i.e. the case with all predictors set to 0 (Low).

In the basic case successes are 153 out of 2160, then frequency is 0,071.

We estimate the odds using empirical frequencies

$$\frac{0,071}{1 - 0,071} = 0,076$$

We expect that a software tool will set the coefficient  $a_0$  in such a way that  $\exp(a_0) = 0,076$  (or close to this value).

$$\exp(a_0) = \frac{p(A = 1 | I = 0, M = 0, R = 0)}{p(A = 0 | I = 0, M = 0, R = 0)} = \frac{0,071}{1 - 0,071} = 0,076$$

In other words  $a_0 = \log(0,076) = -2,571$ .

Indeed, a real software tool chose  $a_0 = -2,500$ , close to  $-2,571$  as suggested by frequencies.

Summarizing: **coefficient  $a_0$  is the logarithm of the odds of the basic case, in which all explicative variables have value 0.**

$$p(A = 1 | I = i, M = m, R = r) = \frac{\exp(a_0 + a_I I + a_M M + a_R R)}{1 + \exp(a_0 + a_I I + a_M M + a_R R)}$$

$$\exp(a_I) = \frac{\text{odds}(I = 1, M = 0, R = 0)}{\text{odds}(I = 0, M = 0, R = 0)}$$

In a similar way : coefficient  $a_I$  is the logarithm of the ratio between the odds of the case where only variable  $I$  equals 1 and the odds of the basic case.

A bit more intuitively: coefficient  $a_I$  is the multiplier of the logarithm of the basic case odds when we lift the education level.

This way of computing coefficients assumes that predictors (explicative variables) are independent.

Each coefficient shows the impact of a change in only one variable from the basic case.

There is no coefficient for joint changes in 2 or 3 variables. Indeed, the structure of the equation is linear.

Note that the basic case is chosen in arbitrary way. We could have chosen value 0 for high education as well.

Note also that the initial choice of discretization criteria is very important. We can state that 12 years of education is Low and 13 is High, or we can set the threshold at 15. It is up to the analyst to choose those criteria in a productive way.