LOGISTIC REGRESSION

- ΠΟΤΕ ΚΑΙ ΓΙΑΤΙ ΧΡΗΣΙΜΟΠΟΙΕΙΤΑΙ
- BINARY, MULTINOMIAL AND ORDINAL LOGISTIC REGRESSION
- ΣΥΝΟΠΤΙΚΗ ΘΕΩΡΙΑ
- LOGISTIC REGRESSION IN SPSS
- ΠΑΡΑΔΕΙΓΜΑ

ΠΟΤΕ

Για την πρόβλεψη των τιμών μίας κατηγορικής μεταβλητής (εξαρτημένη) από τις τιμές μίας ή περισσότερων κατηγορικών ή/και ποσοτικών ανεξάρτητων μεταβλητών.

- Binary: δίτιμη κατηγορική. 1=παρουσία γνωρίσματος 0=απουσία
- Multinomial: κατηγορική με περισσότερες των δύο τιμές
- Ordinal: όταν οι τιμές της ανεξάρτητης μπορούν να διαταχθούν.

ΓΙΑΤΙ ΟΧΙ ΚΛΑΣΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ

Δεν ικανοποιούνται οι υποθέσεις του μοντέλου της γραμμικής παλινδρόμησης:

1. Ομοσκεδαστικότητα
2. Κανονικότητα
3. Γραμμικότητα

π.χ. Εξαρτημένη= Ιδιοκτησία σπιτιού
Ανεξάρτητη = Εισόδημα

MODELS FOR BINARY DATA

- Υποθέτουμε ότι

  1. Logit

$$P(Y_i = 1 / X_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 X_{1i} + \ldots + \beta_p X_{pi}$$

2. Probit

$$\Phi^{-1}(p_i) = \beta_0 + \beta_1 X_{1i} + \ldots + \beta_p X_{pi}$$

LOGIT MODEL

- Logit:

$$\ln\left(\frac{p_i}{1-p_i}\right)$$

- Linear predictor:

$$\beta_0+\beta_1 X_{1i}+\ldots+\beta_p X_{pi}$$

- Logistic function:

$$p_i=\frac{\exp\left(\beta_0+\beta_1 X_{1i}+\ldots+\beta_p X_{pi}\right)}{1+\exp\left(\beta_0+\beta_1 X_{1i}+\ldots+\beta_p X_{pi}\right)}=\frac{1}{1+\exp\left(-\left(\beta_0+\beta_1 X_{1i}+\ldots+\beta_p X_{pi}\right)\right)}$$

Παρατηρήσεις

- Οι κατηγορικές ανεξάρτητες μεταβλητές με k τιμές υπεισέρχονται στο μοντέλο της λογιστικής παλινδρόμησης με χρήση k-1 το πλήθος δείκτριες μεταβλητές.

Odds

- Odds: παριστάνουν την σχετική συχνότητα με την οποία διαφορετικά ενδεχόμενα πραγματοποιούνται. Έτσι γράφοντας α:β σημαίνει για το πρώτο ενδεχόμενο πιθανότητα α/(α+β). Επιπλέον odds=2.5 σημαίνει πιθανότητα 2.5/(2.5+1). Τέλος odds=probability/(1-probability).
  β_i

η αλλαγή στο Log odds για μοναδιαία αύξηση στο

$$X_i$$

ενώ οι υπόλοιπες ανεξάρτητες παραμένουν σταθερές

Ε.Μ.Π.

$$L=\prod_{i=1}^{n}\left[\frac{\exp(A)}{1+\exp(A)}\right]^{Y_i}\left[\frac{1}{1+\exp(A)}\right]^{1-Y_i}$$

$$A=\beta_0+\beta_1X_{1i}+\ldots+\beta_pX_{pi}$$

Ε.Μ.Π.

- Iterative solutions. The general approach involves establishing initial guesses for the unknown parameters and then continuously adjusting these estimates until the maximum value of L is found. (see among others Hosmer and Lemeshow, 2000, p.9).

Logistic Regression in SPSS

- Στο αρχείο lowbirthweight.sav (πηγή Hosmer and Lemeshow (2000)) καταγράφονται πληροφορίες για 189 γεννήσεις καθώς και για τις μητέρες των νεογνών. Το ενδιαφέρον επικεντρώνεται στη μελέτη του φαινομένου της γέννησης νεογνών με βάρος μικρότερου των 2.500 γραμμαρίων. Το ενδιαφέρον εξηγείται διότι η θνησιμότητα των νεογνών σε τέτοιες περιπτώσεις είναι πολύ υψηλή.

Logistic Regression in SPSS

- Παράδειγμα 2

  Στο αρχείο hosmerp.3.sav καταγράφεται η ηλικία (age), η κατάταξη σε ηλικιακό γκρουπ (agerp) και η παρουσία ή όχι σημαντικών στεφανιαίων διαταραχών. Το ενδιαφέρον επικεντρώνεται στη μελέτη της σχέσης της ηλικίας με την παρουσία ή όχι διαταραχών.