

ΙΙΜΣ ΕΠΑΓΓΕΛΜΑΤΙΚΗ ΚΑΙ ΠΕΡΙΒΑΛΛΟΝΤΙΚΗ ΥΓΕΙΑ, ΔΙΑΧΕΙΡΙΣΗ ΚΑΙ ΟΙΚΟΝΟΜΙΚΗ ΔΡΟΤΙΜΩΣΗ
ΑΚ. ΕΤΟΣ 2006-2007, 3ο εξάμηνο

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ

ΜΑΘΗΜΑ 14

Λογιστική Παλινδρόμηση

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ Διαφάνεια 14-1

14. ΛΟΓΙΣΤΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ

ΠΕΡΙΕΧΟΜΕΝΑ

- ΤΟ ΓΕΝΙΚΟ ΜΟΝΤΕΛΟ
- ΕΦΑΡΜΟΓΗ ΜΕ ΤΗ ΧΡΗΣΗ SPSS (ΔΙΤΙΜΑ ΔΕΔΟΜΕΝΑ)
- ΕΡΜΗΝΕΙΑ ΠΑΡΑΜΕΤΡΩΝ
- ΕΛΕΓΧΟΙ ΥΠΟΘΕΣΕΩΝ
- ΚΑΛΗ ΠΡΟΣΑΡΜΟΓΗ ΤΟΥ ΜΟΝΤΕΛΟΥ
- ΠΡΟΒΛΕΨΗ ΜΕ ΤΗ ΧΡΗΣΗ ΤΗΣ ΛΟΓΙΣΤΙΚΗΣ ΠΑΛΙΝΔΡΟΜΗΣΗΣ
- ΣΧΕΣΗ ΛΟΓΙΣΤΙΚΗΣ ΠΑΛΙΝΔΡΟΜΗΣΗΣ ΜΕ ΟΡ ΣΕ 2x2 ΠΙΝΑΚΕΣ
- ΠΑΡΑΔΕΙΓΜΑ ΠΟΛΛΑΠΛΗΣ ΠΑΛΙΝΔΡΟΜΗΣΗΣ & ΕΠΙΛΟΓΗ ΕΠΕΞΗΓΗΜΑΤΙΚΩΝ ΜΕΤΑΒΛΗΤΩΝ

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ Διαφάνεια 14-2

14. ΛΟΓΙΣΤΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ

14.1. Το γενικό μοντέλο

ΠΑΡΑΔΕΙΓΜΑ

X =Σκορ Ευφυΐας WAIS και Y =Νόσος γηρατειών
(1=Νόσος παρούσα)

| X | Y | X | Y | X | Y | X | Y | X | Y |
|----|---|----|---|----|---|----|---|----|---|
| 9 | 1 | 7 | 1 | 7 | 0 | 17 | 0 | 13 | 0 |
| 13 | 1 | 5 | 1 | 16 | 0 | 14 | 0 | 13 | 0 |
| 6 | 1 | 14 | 1 | 9 | 0 | 19 | 0 | 9 | 0 |
| 8 | 1 | 13 | 0 | 9 | 0 | 9 | 0 | 15 | 0 |
| 10 | 1 | 16 | 0 | 11 | 0 | 11 | 0 | 10 | 0 |
| 4 | 1 | 10 | 0 | 13 | 0 | 14 | 0 | 11 | 0 |
| 14 | 1 | 12 | 0 | 15 | 0 | 10 | 0 | 12 | 0 |
| 8 | 1 | 11 | 0 | 13 | 0 | 16 | 0 | 4 | 0 |
| 11 | 1 | 14 | 0 | 10 | 0 | 10 | 0 | 14 | 0 |
| 7 | 1 | 15 | 0 | 11 | 0 | 16 | 0 | 20 | 0 |
| 9 | 1 | 18 | 0 | 6 | 0 | 14 | 0 | | |

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ Διαφάνεια 14-3

14. ΛΟΓΙΣΤΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ

14.1. Το γενικό μοντέλο

ΠΑΡΑΔΕΙΓΜΑ

(Agresti, 1991, σελ. 122-123)

Στη μελέτη αυτή έχουμε δείγμα ηλικιωμένων οι οποίοι πέρασαν από ψυχιατρική εξέταση για την οποία εξετάστηκαν για τη νόσο του γήρατος (senility symptoms).

Εδώ θέλουμε να δούμε κατά πόσο μπορούμε να προβλέψουμε αν κάποιος ηλικιωμένος έχει τη νόσο χρησιμοποιώντας μόνο την κλίμακα ευφυΐας του Wechsler (Wechsler adult intelligence scale, WAIS)

1. Να βρεθεί ποια τιμή του WAIS που αντιστοιχεί σε πιθανότητα $> 1/2$ (δηλαδή τότε υπάρχει αυξημένος κίνδυνος)
2. Να ερμηνεύσουμε τη σχέση WAIS και νόσου

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ Διαφάνεια 14-4

14. ΛΟΓΙΣΤΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ
14.1. Το γενικό μοντέλο

Μοντέλο (με τις γνώσεις μας μέχρι τώρα)

$$P(\text{Senility}=1) = a + b \text{ WAIS} + \epsilon, \quad \epsilon \sim N(0, \sigma^2)$$

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ Διαφάνεια 14-5

14. ΛΟΓΙΣΤΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ
14.1. Το γενικό μοντέλο

Μοντέλο (με τις γνώσεις μας μέχρι τώρα)

$$P(\text{Senility}=1) = a + b \text{ WAIS} + \epsilon, \quad \epsilon \sim N(0, \sigma^2)$$

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ Διαφάνεια 14-7

14. ΛΟΓΙΣΤΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ
14.1. Το γενικό μοντέλο

Μοντέλο (με τις γνώσεις μας μέχρι τώρα)

$$P(\text{Senility}=1) = a + b \text{ WAIS} + \epsilon, \quad \epsilon \sim N(0, \sigma^2)$$

Model Summary^a

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
|-------|-------------------|----------|-------------------|----------------------------|
| 1 | .426 ^a | .181 | .166 | .40405 |

a. Predictors: (Constant), wais Wechsler Adult

Coefficients^a

| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
|-------|--|-----------------------------|------------|---------------------------|--------|------|
| | | B | Std. Error | Beta | | |
| 1 | (Constant) | .847 | .182 | | 4.662 | .000 |
| | wais Wechsler Adult Intelligence Scale | -.051 | .015 | -.426 | -3.395 | .001 |

a. Dependent Variable: senility_Senility Symptoms

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ Διαφάνεια 14-8

14. ΛΟΓΙΣΤΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ
14.1. Το γενικό μοντέλο

ΠΡΟΒΛΗΜΑΤΑ????

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ Διαφάνεια 14-8

14. ΛΟΓΙΣΤΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ

14.1. Το γενικό μοντέλο

ΠΡΟΒΛΗΜΑΤΑ:

- Παραβίαση των προϋποθέσεων (κανονικότητα, γραμμικότητα, ομοσκεδαστικότητα)
- Οι προβλέψεις για την Y δεν είναι στο διάστημα $[0,1]$!!!???

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ

Διαφάνεια 14-9

14. ΛΟΓΙΣΤΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ

14.1. Το γενικό μοντέλο

ΠΩΣ ΛΥΝΟΝΤΑΙ ΤΑ ΠΡΟΒΛΗΜΑΤΑ:

1. Υποθέτοντας Διωνυμική Κατανομή
2. Συνδέοντας την πιθανότητα p με τις επεξηγηματικές μεταβλητές (γραμμικό συνδυασμό $a+bx$) μέσω μιας συνάρτησης που θα δίνει τιμές στο p στο διάστημα $[0,1]$

ΣΥΝΑΡΤΗΣΗ

$$\text{logit}(p) = \log(p/(1-p)) = a+bx$$

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ

Διαφάνεια 14-10

14. ΛΟΓΙΣΤΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ

14.1. Το γενικό μοντέλο

ΜΟΝΤΕΛΟ

$$Y_i \sim \text{Binomial}(p_i, N_i)$$

$$\log(p_i/(1-p_i)) = b_0 + b_1X_{i1} + \dots + b_pX_{ip}$$

Όταν $Y =$ δίτιμη τότε $n_i = 1$ (δηλαδή κατανομή Bernoulli)

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ

Διαφάνεια 14-11

14. ΛΟΓΙΣΤΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ

14.1. Το γενικό μοντέλο

ΜΟΝΤΕΛΟ

$$p_i = e^{b_0 + b_1X_{i1} + \dots + b_pX_{ip}} / [1 + e^{b_0 + b_1X_{i1} + \dots + b_pX_{ip}}]$$

Αναμενόμενες τιμές: $E(Y_i | X_{i1}, \dots, X_{ip}) = n_i p_i$

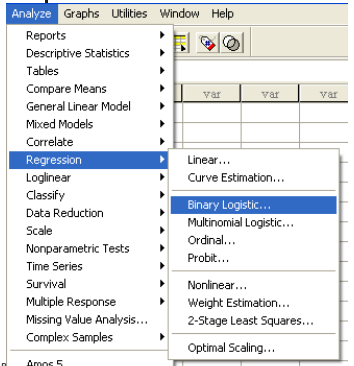
ΣΗΜΑΝΤΙΚΟ ΜΟΝΤΕΛΟ ΓΙΑΤΙ

ΚΑΝΕΙ ΧΡΗΣΗ ΤΩΝ ODDS + ODDS RATIO

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ

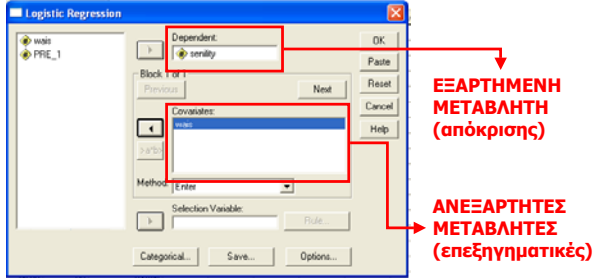
Διαφάνεια 14-12

14. ΛΟΓΙΣΤΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ
14.2. Εφαρμογή με τη χρήση του SPSS
(Δίτιμα δεδομένα)



Amos 5 Διαφάνεια 14-13

14. ΛΟΓΙΣΤΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ
14.2. Εφαρμογή με τη χρήση του SPSS
(Δίτιμα δεδομένα)



ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ Διαφάνεια 14-14

14. ΛΟΓΙΣΤΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ
14.2. Εφαρμογή με τη χρήση του SPSS
(Δίτιμα δεδομένα)

Variables in the Equation

| Step | | B | S.E. | Wald | df | Sig. | Exp(B) |
|----------------|----------|-------|-------|-------|----|------|--------|
| 1 ^a | wais | -.324 | .114 | 8.057 | 1 | .005 | .724 |
| | Constant | 2.404 | 1.192 | 4.069 | 1 | .044 | 11.068 |

a. Variable(s) entered on step 1: wais.

ΜΟΝΤΕΛΟ

Log(ODDS) = 2.404 – 0.324 WAIS

$p = e^{2.404 - 0.324 WAIS} / (1 + e^{2.404 - 0.324 WAIS})$

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ Διαφάνεια 14-15

14. ΛΟΓΙΣΤΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ
14.3. Ερμηνεία μοντέλου

Έστω ότι έχουμε 2 άτομα A και B με $X = X_A$ και $X = X_B$

$$\text{Log}(\text{ODDS}_A) = a + b X_A$$

$$\text{Log}(\text{ODDS}_B) = a + b X_B$$

$$\text{LOG}(\text{OR}_{A-B}) = a + b X_A - a - b X_B = b(X_A - X_B)$$

Αν τα δύο άτομα διαφέρουν ως προς το X κατά μία μονάδα ΤΟΤΕ

$$\text{LOG}(\text{OR}_{A-B}) = b \Rightarrow \text{OR}_{A-B} = e^b$$

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ Διαφάνεια 14-16

14. ΛΟΓΙΣΤΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ

14.3. Ερμηνεία μοντέλου

Όμοια είναι η ερμηνεία στην πολλαπλή παλινδρόμηση όμως υποθέτουμε ότι μόνο μια μεταβλητή αλλάζει κάθε φορά.

Συνεπώς ο συντελεστής συνδέεται με το OR διορθωμένο ως προς τις υπόλοιπες επεξηγηματικές μεταβλητές.

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ

Διαφάνεια 14-17

14. ΛΟΓΙΣΤΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ

14.3. Ερμηνεία μοντέλου

Variables in the Equation

| | B | S.E. | Wald | df | Sig. | Exp(B) |
|------------------------|-------|-------|-------|----|------|--------|
| Step 1 ^a | | | | | | |
| wais | -.324 | .114 | 8.057 | 1 | .005 | .724 |
| Constant | 2.404 | 1.192 | 4.069 | 1 | .044 | 11.068 |

a. Variable(s) entered on step 1: wais.

ΜΟΝΤΕΛΟ

$$\text{Log(ODDS)} = 2.404 - 0.324 \text{ WAIS}$$

$$\text{OR} = e^{-0.324} = 0.724$$

Συνεπώς αύξηση των θετικών απαντήσεων στην κλίμακα WAIS κατά μία συνεπάγεται μείωση της σχετικής πιθανότητας της νόσου να είναι παρούσα κατά 27.6% [(1-0.724)100%]

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ

Διαφάνεια 14-18

14. ΛΟΓΙΣΤΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ

14.3. Ερμηνεία μοντέλου

Έστω ότι έχουμε 1 άτομο με $X=0$ τότε

$$\text{Log(ODDS}_A) = a + b \cdot 0 = a$$

$$p = e^a / (1 + e^a)$$

Συνεπώς η σταθερά a συνδέεται με το λογάριθμο του ODDS όταν $X=0$ (ή όταν όλα τα $X=0$ όταν έχουμε πολλαπλή παλινδρόμηση).

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ

Διαφάνεια 14-19

14. ΛΟΓΙΣΤΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ

14.3. Ερμηνεία μοντέλου

Variables in the Equation

| | B | S.E. | Wald | df | Sig. | Exp(B) |
|------------------------|-------|-------|-------|----|------|--------|
| Step 1 ^a | | | | | | |
| wais | -.324 | .114 | 8.057 | 1 | .005 | .724 |
| Constant | 2.404 | 1.192 | 4.069 | 1 | .044 | 11.068 |

a. Variable(s) entered on step 1: wais.

ΜΟΝΤΕΛΟ

$$\text{Log(ODDS)} = 2.404 - 0.324 \text{ WAIS}$$

$$\text{ODDS}(X=0) = e^{2.404} = 11.07$$

Συνεπώς αν κάποιος ηλικιωμένος δε δώσει καμία θετική απάντηση στην κλίμακα WAIS τότε πιθανότητα η νόσος να είναι παρούσα είναι 11πλάσια της πιθανότητας η νόσος να είναι απύουσα

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ

Διαφάνεια 14-20

14. ΛΟΓΙΣΤΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ
14.3. Ερμηνεία μοντέλου

| Variables in the Equation | | | | | | | |
|---------------------------|----------|-------|-------|-------|----|------|--------|
| Step | | B | S.E. | Wald | df | Sig. | Exp(B) |
| 1 ^a | weis | -.324 | .114 | 8.057 | 1 | .005 | .724 |
| | Constant | 2.404 | 1.192 | 4.069 | 1 | .044 | 11.068 |

a. Variable(s) entered on step 1: weis.

ΜΟΝΤΕΛΟ

Log(ODDS) = 2.404 – 0.324 WAIS
P(X=0) = e^{2.404} / (1+e^{2.404}) = 0.9171318

Συνεπώς αν κάποιος ηλικιωμένος δε δώσει καμία θετική απάντηση στην κλίμακα WAIS τότε πιθανότητα η νόσος να είναι παρούσα είναι ίση με 91.7%.

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ Διαφάνεια 14-21

14. ΛΟΓΙΣΤΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ
14.3. Ερμηνεία μοντέλου

| Variables in the Equation | | | | | | | | | |
|---------------------------|----------|-------|-------|-------|----|------|--------|-----------------------|------|
| Step | | B | S.E. | Wald | df | Sig. | Exp(B) | 95.0% C.I. for EXP(B) | |
| 1 ^a | weis | -.324 | .114 | 8.057 | 1 | .005 | .724 | | |
| | Constant | 2.404 | 1.192 | 4.069 | 1 | .044 | 11.068 | .579 | .905 |

a. Variable(s) entered on step 1: weis.

95% ΔΕ για b => b ± 1.96 se(b)
95% ΔΕ για OR => e^b ± 1.96 se(b)

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ Διαφάνεια 14-23

14. ΛΟΓΙΣΤΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ
14.3. Ερμηνεία μοντέλου

Δ.Ε. Για τα ODDS RATIOS

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ Διαφάνεια 14-22

14. ΛΟΓΙΣΤΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ
14.4. Έλεγχος Υποθέσεων

| Variables in the Equation | | | | | | | | | |
|---------------------------|----------|-------|-------|-------|----|------|--------|-------|-------|
| Step | | B | S.E. | Wald | df | Sig. | Exp(B) | Lower | Upper |
| 1 ^a | weis | -.324 | .114 | 8.057 | 1 | .005 | .724 | | |
| | Constant | 2.404 | 1.192 | 4.069 | 1 | .044 | 11.068 | .579 | .905 |

a. Variable(s) entered on step 1: weis.

H₀: β=0 vs. H₁: β≠0
p.value = 0.005 < 0.05
 => απορρίπτουμε H₀
 => άρα κάθε επιπλέον θετική απάντηση στην WAIS μειώνει σημαντικά τη σχετική πιθανότητα της νόσου
 => άρα το WAIS σχετίζεται με τη νόσο

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ Διαφάνεια 14-24

14. ΛΟΓΙΣΤΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ

14.4. Έλεγχος Υποθέσεων

| Step | Variable | B | S.E. | Wald | df | Sig. | Exp(B) | 95.0% C.I. for EXP(B) | |
|----------------|----------|-------|-------|-------|----|------|--------|-----------------------|-------|
| | | | | | | | | Lower | Upper |
| 1 ^a | waiss | -.324 | .114 | 8.057 | 1 | .005 | .724 | .579 | .905 |
| | Constant | 2.404 | 1.192 | 4.069 | 1 | .044 | 11.068 | | |

a. Variable(s) entered on step 1: waiss.

H₀: α=0 vs. H₁: α≠0
p.value = 0.044 < 0.05
 => απορρίπτουμε H₀
 => άρα αν κάποιος δεν δώσει καμία θετική απάντηση τότε θα έχει πιθανότητα εμφάνισης της νόσου σημαντικά διαφορετική του 50%

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ Διαφάνεια 14-25

14. ΛΟΓΙΣΤΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ

14.5. Έλεγχος καλής προσαρμογής

DEVIANCE του μοντέλου m

$$l = \log f(\mathbf{y} | \boldsymbol{\beta})$$

$$D_m = -2 \log f(\mathbf{y} | \boldsymbol{\beta}) = -2 \sum_{i=1}^n \{y_i \log(p_i) + (1 - y_i) \log(1 - p_i)\}$$

D_c = Deviance του σταθερού μοντέλου
 Δηλαδή $\log(\text{ODDS}) = a$ [σταθερή πιθανότητα]

D_s = Deviance του κορεσμένου μοντέλου
 Κορεσμένο μοντέλο => μέσοι = data

Πολλές φορές από όλα τα deviance αφαιρούμε αυτή του σταθερού έτσι ώστε $\Delta_m = D_m - D_s => \Delta_s = 0$

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ Διαφάνεια 14-26

14. ΛΟΓΙΣΤΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ

14.5. Έλεγχος καλής προσαρμογής

Για κάθε μοντέλο m διαφορετικό του σταθερού ή του κορεσμένου ΙΣΧΥΕΙ

$$D_s < D_m < D_c$$

$$\Delta_s < \Delta_m < \Delta_c$$

ΕΠΙΠΛΕΟΝ (για μεγάλα δείγματα)

$$\Delta_m = D_m - D_s \sim \chi_{n-pm}^2$$

$$\Delta_0 - \Delta_m = D_0 - D_m \sim \chi_{pm-1}^2$$

Γενικότερα αν m₁ είναι φωλιασμένο (nested) στο μοντέλο m₂ ισχύει

$$\Delta_{m1} - \Delta_{m2} = D_{m1} - D_{m2} \sim \chi_{pm2-pm1}^2$$

p_m = αριθμός παραμέτρων στο μοντέλο m

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ Διαφάνεια 14-27

14. ΛΟΓΙΣΤΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ

14.5. Έλεγχος καλής προσαρμογής

D₁ = 51.01738
D₀ = 61.80632
D₀-D₁ = 10.78894

Omnibus Tests of Model Coefficients

| | | Chi-square | df | Sig. |
|--------|-------|------------|----|------|
| Step 1 | Step | 10.789 | 1 | .001 |
| | Block | 10.789 | 1 | .001 |
| | Model | 10.789 | 1 | .001 |

Διαφορά deviance από προηγούμενο βήμα
Εδώ το προηγούμενο βήμα είναι το σταθερό
Άρα διαφορά από το σταθερό μοντέλο
H₀: β₁ = β₂ = ... = β_n
Βλέπουμε κατά πόσο οι X βελτιώνουν την εκτίμηση

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ Διαφάνεια 14-28

14. ΛΟΓΙΣΤΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ
14.5. Έλεγχος καλής προσαρμογής

$D_m = 51.01738$
 $D_0 = 61.80632$
 $D_0 - D_m = 10.78894$

Model Summary

| Step | -2 Log likelihood | Cox & Snell R Square | Nagelkerke R Square |
|------|---------------------|----------------------|---------------------|
| 1 | 51.017 ^a | .181 | .266 |

a. Estimation terminated at iteration number 5 because parameter estimates changed by less than .001.

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ Διαφάνεια 14-29

14. ΛΟΓΙΣΤΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ
14.5. Έλεγχος καλής προσαρμογής

$D_m = 51.01738$
 $D_0 = 61.80632$
 $D_0 - D_m = 10.78894$

Model Summary

| Step | -2 Log likelihood | Cox & Snell R Square | Nagelkerke R Square |
|------|---------------------|----------------------|---------------------|
| 1 | 51.017 ^a | .181 | .266 |

a. Estimation terminated at iteration number 5 because parameter estimates changed by less than .001.

$R_{CS}^2 = 1 - (l_0 / l_m)^{2/N} = 1 - e^{(D_m - D_0)/N}$
 $N = \sum n_i$
Στο τρέχων παράδειγμα $R_{CS}^2 = 1 - e^{(51.02 - 61.81)/54} = 0.181$

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ Διαφάνεια 14-30

14. ΛΟΓΙΣΤΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ
14.5. Έλεγχος καλής προσαρμογής

$D_m = 51.01738$
 $D_0 = 61.80632$
 $D_0 - D_m = 10.78894$

Model Summary

| Step | -2 Log likelihood | Cox & Snell R Square | Nagelkerke R Square |
|------|---------------------|----------------------|---------------------|
| 1 | 51.017 ^a | .181 | .266 |

a. Estimation terminated at iteration number 5 because parameter estimates changed by less than .001.

$R_N^2 = R_{CS}^2 / (1 - l_0^{2/N}) = 1 - e^{(D_m - D_0)/N} / (1 - e^{-D_0/N})$
Στο τρέχων παράδειγμα $R_N^2 = 0.181 / (1 - e^{-61.81/54}) = 0.266$

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ Διαφάνεια 14-31

14. ΛΟΓΙΣΤΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ
14.5. Έλεγχος καλής προσαρμογής

ΣΤΑ GLM και στα μοντέλα λογιστικής παλινδρόμησης με βάση διωνυμική με $N_i > 1$ μπορούμε να χρησιμοποιήσουμε τη διαφορά των deviance από το κορεσμένο μοντέλο

Στα δίτιμα όμως αυτή η ποσότητα δεν μπορεί να υπολογιστεί άρα πρέπει να βρούμε εναλλακτικούς τρόπους

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ Διαφάνεια 14-32

14. ΛΟΓΙΣΤΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ
14.5. Έλεγχος καλής προσαρμογής

ΕΛΕΓΧΟΣ ΚΑΛΗΣ ΠΡΟΣΑΡΜΟΓΗΣ Hosmer+Lemeshow

1. Χωρίζουμε τα fitted p σε g (συνήθως με ίσο μέγεθος) ομάδες και φτιάχνουμε τον πίνακα (συνήθως g=10)

G Ομάδες x 2 κατηγορίες της Y με συχνότητες O_{ij}

2. Από αυτό τον πίνακα υπολογίζουμε

$$x_{HL}^2 = \sum_{k=1}^g \frac{(O_{k1} - E_{k1})^2}{E_{k1}(1 - E_{k1}/m_k)}$$

O_{k1} = παρατηρούμενες τιμές στην k ομάδα με $Y=1$

E_{k1} = αναμενόμενες τιμές (με βάση το μοντέλο) στην k ομάδα με $Y=1$ (προκύπτει ως άθροισμα των fitted πιθανοτήτων)

m_k = συνολικός αριθμός ατόμων στην k ομάδα

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ

Διαφάνεια 14-33

14. ΛΟΓΙΣΤΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ
14.5. Έλεγχος καλής προσαρμογής

Contingency Table for Hosmer and Lemeshow Test

| Step | 1 | senility Symptom = .00 not present | | senility Symptom = 1.00 present | | Total |
|------|---|------------------------------------|----------|---------------------------------|----------|-------|
| | | Observed | Expected | Observed | Expected | |
| | | 4 | 3.885 | 0 | .115 | |
| 2 | 4 | 3.765 | 0 | .235 | 4 | |
| 3 | 3 | 2.761 | 0 | .239 | 3 | |
| 4 | 5 | 6.253 | 2 | .747 | 7 | |
| 5 | 5 | 5.150 | 1 | .850 | 6 | |
| 6 | 7 | 6.191 | 1 | 1.809 | 8 | |
| 7 | 5 | 4.180 | 1 | 1.820 | 6 | |
| 8 | 4 | 3.746 | 2 | 2.254 | 6 | |
| 9 | 1 | 2.488 | 4 | 2.512 | 5 | |
| 10 | 2 | 1.581 | 3 | 3.419 | 5 | |

$D_m = 51.01738$

$D_0 = 61.80632$

$D_0 - D_m = 10.78894$

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ

Διαφάνεια 14-34

14. ΛΟΓΙΣΤΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ
14.5. Έλεγχος καλής προσαρμογής

| Contingency Table for Hosmer and Lemeshow Test | | | | | | | $\frac{(O_{k1} - E_{k1})^2}{E_{k1}(1 - E_{k1}/m_k)}$ | HL |
|--|---|------------------------------------|----------|------------------------|----------|-------------|--|----|
| Step | 1 | senility Symptom = .00 not present | | Symptom = 1.00 present | | Total | | |
| | | Observed | Expected | Observed | Expected | | | |
| 1 | 4 | 3.885 | 0 | 0.115 | 4 | 0.118 | | |
| 2 | 4 | 3.765 | 0 | 0.235 | 4 | 0.250 | | |
| 3 | 3 | 2.761 | 0 | 0.239 | 3 | 0.259 | | |
| 4 | 5 | 6.253 | 2 | 0.747 | 7 | 2.356 | | |
| 5 | 5 | 5.150 | 1 | 0.850 | 6 | 0.031 | | |
| 6 | 7 | 6.191 | 1 | 1.809 | 8 | 0.468 | | |
| 7 | 5 | 4.180 | 1 | 1.820 | 6 | 0.531 | | |
| 8 | 4 | 3.746 | 2 | 2.254 | 6 | 0.046 | | |
| 9 | 1 | 2.488 | 4 | 2.512 | 5 | 1.770 | | |
| 10 | 2 | 1.581 | 3 | 3.419 | 5 | 0.162 | | |
| | | O_{k1} | | E_{k1} | | $x^2(HL) =$ | 5.991 | |

$$x_{HL}^2 = \sum_{k=1}^g \frac{(O_{k1} - E_{k1})^2}{E_{k1}(1 - E_{k1}/m_k)}$$

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ

Διαφάνεια 14-35

14. ΛΟΓΙΣΤΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ
14.5. Έλεγχος καλής προσαρμογής

| Contingency Table for Hosmer and Lemeshow Test | | | | | | | $\frac{(O_{k1} - E_{k1})^2}{E_{k1}(1 - E_{k1}/m_k)}$ | HL |
|--|---|------------------------------------|----------|------------------------|----------|---------|--|----|
| Step | 1 | senility Symptom = .00 not present | | Symptom = 1.00 present | | Total | | |
| | | Observed | Expected | Observed | Expected | | | |
| 1 | 4 | 3.885 | 0 | 0.115 | 4 | 0.118 | | |
| 2 | 4 | 3.765 | 0 | 0.235 | 4 | 0.250 | | |
| 3 | 3 | 2.761 | 0 | 0.239 | 3 | 0.259 | | |
| 4 | 5 | 6.253 | 2 | 0.747 | 7 | 2.356 | | |
| 5 | 5 | 5.150 | 1 | 0.850 | 6 | 0.031 | | |
| 6 | 7 | 6.191 | 1 | 1.809 | 8 | 0.468 | | |
| 7 | 5 | 4.180 | 1 | 1.820 | 6 | 0.531 | | |
| 8 | 4 | 3.746 | 2 | 2.254 | 6 | 0.046 | | |
| 9 | 1 | 2.488 | 4 | 2.512 | 5 | 1.770 | | |
| 10 | 2 | 1.581 | 3 | 3.419 | 5 | 0.162 | | |
| | | O_{k1} | | E_{k1} | | $x^2 =$ | 5.991 | |

Hosmer and Lemeshow Test

| Step | Chi-square | df | Sig. |
|------|------------|----|------|
| 1 | 5.991 | 8 | .648 |

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ

Διαφάνεια 14-36

14. ΛΟΓΙΣΤΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ
14.5. Έλεγχος καλής προσαρμογής

Classification Table ^a

| Observed | | Predicted | | Percentage Correct |
|--------------------|-------------------------------|-------------------------------|---------------------------|--------------------|
| | | Senility Symptoms not present | Senility Symptoms present | |
| Step 1 | Senility Symptoms not present | 37 | 3 | 92.5 |
| | Senility Symptoms present | 9 | 5 | 35.7 |
| Overall Percentage | | | | 77.8 |

a. The cut value is .500

ΠΙΘΑΝΟΤΗΤΑ ΣΩΣΤΗΣ ΠΡΟΒΛΕΨΗΣ ΜΕ ΤΟ ΤΡΕΧΩΝ ΜΟΝΤΕΛΟ = 77.8%

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ

Διαφάνεια 14-37

14. ΛΟΓΙΣΤΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ
14.6. Πρόβλεψη με τη χρήση της λογιστικής παλινδρόμησης

ΤΙ ΜΠΟΡΟΥΜΕ ΝΑ ΚΑΝΟΥΜΕ

1. ΠΙΝΑΚΑΣ ΠΡΟΒΛΕΨΕΩΝ – ΠΙΘΑΝΟΤΗΤΩΝ
2. ΔΙΑΓΡΑΜΜΑ
3. ΤΙΜΗ ΔΙΑΧΩΡΙΣΜΟΥ

ΥΠΕΝΘΥΜΙΖΟΥΜΕ ΟΤΙ ΠΡΟΒΛΕΨΕΙΣ ΕΚΤΟΣ ΤΩΝ ΠΑΡΑΤΗΡΟΥΜΕΝΩΝ ΤΙΜΩΝ ΤΗΣ Χ ΔΕΝ ΕΙΝΑΙ ΑΞΙΟΠΙΣΤΕΣ

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ

Διαφάνεια 14-38

14. ΛΟΓΙΣΤΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ
14.6. Πρόβλεψη με τη χρήση της λογιστικής παλινδρόμησης

ΠΙΝΑΚΑΣ ΠΡΟΒΛΕΨΕΩΝ – ΠΙΘΑΝΟΤΗΤΩΝ

$$p = e^{2.404 - 0.324 W} / (1 + e^{2.404 - 0.324 W})$$

W = WAIS = 0, 1, 2, ..., 20

Αν θέλουμε για κάποιες τιμές μπορούμε να επιλέξουμε

min, Q1, median, mean, Q3, max

| WAIS | p |
|------|-------|
| 0 | 0.917 |
| 1 | 0.889 |
| 2 | 0.853 |
| 3 | 0.807 |
| 4 | 0.752 |
| 5 | 0.687 |
| 6 | 0.613 |
| 7 | 0.534 |
| 8 | 0.453 |
| 9 | 0.375 |
| 10 | 0.302 |
| 11 | 0.239 |
| 12 | 0.185 |
| 13 | 0.141 |
| 14 | 0.106 |
| 15 | 0.079 |
| 16 | 0.058 |
| 17 | 0.043 |
| 18 | 0.031 |
| 19 | 0.023 |
| 20 | 0.017 |

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ

Διαφάνεια 14-39

14. ΛΟΓΙΣΤΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ
14.6. Πρόβλεψη με τη χρήση της λογιστικής παλινδρόμησης

ΠΙΝΑΚΑΣ ΠΡΟΒΛΕΨΕΩΝ – ΠΙΘΑΝΟΤΗΤΩΝ

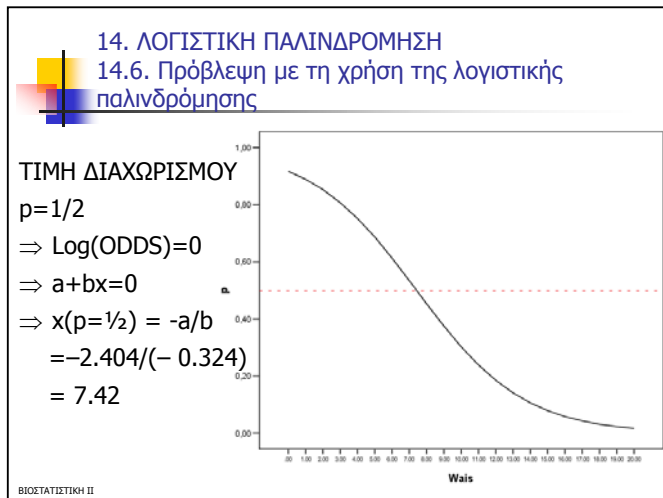
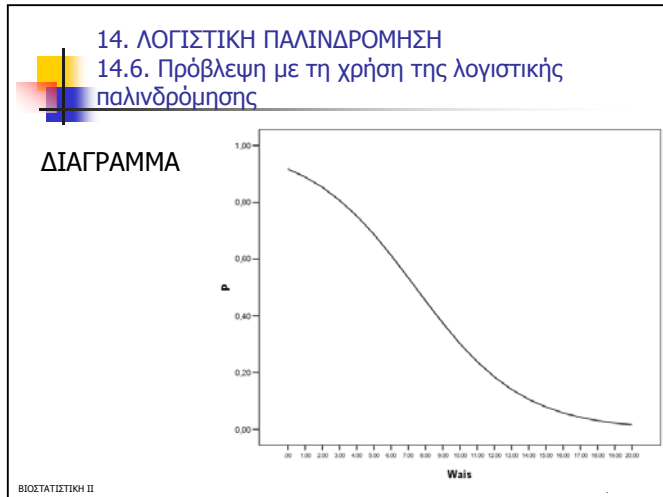
Αν θέλουμε για κάποιες τιμές μπορούμε να επιλέξουμε

min, Q1, median, mean, Q3, max

| | Προβλεπόμενη Πιθανότητα |
|-----------------|-------------------------|
| Ελάχιστη Τιμή | 4 0.752 |
| 1ο Τεταρτημόριο | 9 0.375 |
| Διάμεσος | 11 0.239 |
| Μέσος | 11.57407 0.207 |
| 3ο Τεταρτημόριο | 14 0.106 |
| Μέγιστη τιμή | 20 0.017 |

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ

Διαφάνεια 14-40



14. ΛΟΓΙΣΤΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ
14.7. Σχέση Λογιστικής παλινδρόμησης με OR σε 2x2 Πίνακες

**ΠΑΡΑΔΕΙΓΜΑ ΜΕ 2x2 ΠΙΝΑΚΑ
ΚΑΙ ΟΙ 2 ΜΕΤΑΒΛΗΤΕΣ ΕΙΝΑΙ 0-1 (1= ΝΑΙ)**

2x2_contraception.sav - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Util

ΚΑΝΟΥΜΕ weight cases ως προς COUNTS

12 :

| | contraception | myocardial | counts | Var |
|---|---------------|------------|--------|-----|
| 1 | 1 | 1 | 23 | |
| 2 | 1 | 0 | 34 | |
| 3 | 0 | 1 | 35 | |
| 4 | 0 | 0 | 132 | |
| 5 | | | | |

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ Διαφάνεια 14-44

14. ΛΟΓΙΣΤΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ
14.7. Σχέση Λογιστικής παλινδρόμησης με OR σε 2x2 Πίνακες

ΦΤΙΑΧΝΟΥΜΕ ΜΟΝΤΕΛΟ ΜΕ
Y=myocardial
X=χρήση αντισυλληπτικού χαπιού

Variables in the Equation

| Step | myocardial | B | S.E. | Wald | df | Sig. | Exp(B) | 95.0% C.I. for EXP(B) | |
|------|------------|--------|------|--------|----|------|--------|-----------------------|-------|
| | | | | | | | | Lower | Upper |
| 1 | Constant | -1.356 | .192 | 49.745 | 1 | .000 | .258 | 1.336 | 4.873 |

a. Variable(s) entered on step 1: myocardial.

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II Διαφάνεια 14-45

14. ΛΟΓΙΣΤΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ
14.7. Σχέση Λογιστικής παλινδρόμησης με OR σε 2x2 Πίνακες

Risk Estimate

| | Value | 95% Confidence Interval | |
|---|-------|-------------------------|-------|
| | | Lower | Upper |
| Odds Ratio for contraception Oral Contraception (0 / 1) | 2.551 | 1.336 | 4.873 |
| For cohort myocardial Myocardial Infraction = 0 | 1.525 | 1.056 | 1.663 |
| For cohort myocardial Myocardial Infraction = 1 | .519 | .337 | .800 |
| N of Valid Cases | 224 | | |

Variables in the Equation

| Step | myocardial | B | S.E. | Wald | df | Sig. | Exp(B) | 95.0% C.I. for EXP(B) | |
|------|------------|--------|------|--------|----|------|--------|-----------------------|-------|
| | | | | | | | | Lower | Upper |
| 1 | Constant | -1.356 | .192 | 49.745 | 1 | .000 | .258 | 1.336 | 4.873 |

a. Variable(s) entered on step 1: myocardial.

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II Διαφάνεια 14-46

14. ΛΟΓΙΣΤΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ
14.7. Σχέση Λογιστικής παλινδρόμησης με OR σε 2x2 Πίνακες

Άρα όταν έχουμε μία X κατηγορική τότε είναι το ίδιο σαν να εκτιμάμε το OR όπως μάθαμε σε απλούς πίνακες 2x2

ΓΙΑΤΙ ΕΙΝΑΙ ΚΑΛΥΤΕΡΟΣ ΑΥΤΟΣ Ο ΤΡΟΠΟΣ

- Μπορούμε να έχουμε X με πολλά επίπεδα
- Μπορούμε να βάλουμε πολλά X (ποσοτικά+κατηγορικά)
- Να κάνουμε προβλέψεις

Variables in the Equation

| Step | myocardial | B | S.E. | Wald | df | Sig. | Exp(B) | 95.0% C.I. for EXP(B) | |
|------|------------|--------|------|--------|----|------|--------|-----------------------|-------|
| | | | | | | | | Lower | Upper |
| 1 | Constant | -1.356 | .192 | 49.745 | 1 | .000 | .258 | 1.336 | 4.873 |

a. Variable(s) entered on step 1: myocardial.

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II Διαφάνεια 14-47

14. ΛΟΓΙΣΤΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ
14.8. Παράδειγμα πολλαπλής παλινδρόμησης

ΜΟΛΥΒΔΟΣ & ΨΥΧΟΛΟΓΙΚΗ – ΝΕΥΡΟΛΟΓΙΚΗ ΚΑΤΑΣΤΑΣΗ ΠΑΙΔΙΩΝ – LEAD.DAT

Το 1975 δημοσιεύθηκαν τα αποτελέσματα μιας μεγάλης μελέτης που σκοπό είχε να διερευνήσει την επίδραση της παρατεταμένης έκθεσης στο μόλυβδο στην ψυχολογική και νευρολογική κατάσταση παιδιών (για λεπτομέρειες βλ. Landigram *et al.*, 1975, *Lancet*, 1, σελ. 708 - 715). Τα δεδομένα είναι διαθέσιμα στο αρχείο LEAD.DAT.

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II Διαφάνεια 14-48

14. ΛΟΓΙΣΤΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ

14.8. Παράδειγμα πολλαπλής παλινδρόμησης

ΜΟΛΥΒΑΟΣ & ΨΥΧΟΛΟΓΙΚΗ – ΝΕΥΡΟΛΟΓΙΚΗ ΚΑΤΑΣΤΑΣΗ ΠΑΙΔΙΩΝ – LEAD.DAT

Περίληπτικά να αναφέρουμε ότι η μελέτη εξετάζει μια ομάδα παιδιών που διέμεναν κοντά σε ένα εργοστάσιο τήξεως μολύβδου στο El Paso του Texas.

Στα παιδιά αυτά καταγράφηκε το επίπεδο μολύβδου στο αίμα τους το 1972 και 1973. Μια ομάδα έκθεσης αποτελείται από 46 παιδιά με επίπεδο μολύβδου 40 µg/ml στο αίμα το 1972 (ή σε μερικές περιπτώσεις το 1973).

Αυτή ομάδα αναφέρεται ως GROUP = 2 ή 3. Μια ομάδα ελέγχου (μαρτύρων) αποτελείται από 78 παιδιά με επίπεδο μολύβδου < 40 µg/ml στο αίμα και τις 2 χρονιές (1972 & 1973).

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ

Διαφάνεια 14-49

14. ΛΟΓΙΣΤΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ

14.8. Παράδειγμα πολλαπλής παλινδρόμησης

ΜΟΛΥΒΑΟΣ & ΨΥΧΟΛΟΓΙΚΗ – ΝΕΥΡΟΛΟΓΙΚΗ ΚΑΤΑΣΤΑΣΗ ΠΑΙΔΙΩΝ – LEAD.DAT

Σημαντικές μεταβλητές για τη μελέτη είναι και οι ακόλουθες

1. αριθμός κτυπημάτων δακτύλου – καρπού (finger-wrist taps test) του κυρίως χεριού το οποίο αποτελεί ένα σύνηθες τεστ καλής νευρολογικής λειτουργίας.

2. Η κλίμακα ευφυΐας του Wechsler .

ΕΛΩ ΘΑ ΜΕΛΕΤΗΣΟΥΜΕ ΠΟΙΟΙ ΠΑΡΑΓΟΝΤΕΣ ΕΠΗΡΕΑΖΟΥΝ ΤΗΝ ΕΜΦΑΝΙΣΗ ΥΨΗΛΗΣ ΠΕΡΙΕΚΤΙΚΟΤΗΤΑΣ ΜΟΛΥΒΑΟΥ ΣΤΟ ΑΙΜΑ.

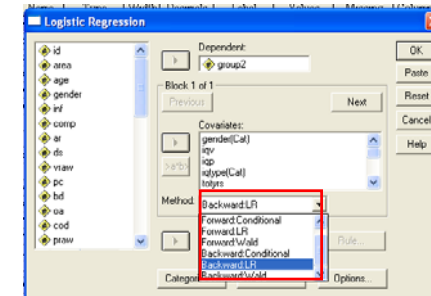
ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ

Διαφάνεια 14-50

14. ΛΟΓΙΣΤΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ

14.8. Παράδειγμα πολλαπλής παλινδρόμησης

ΧΡΗΣΙΜΟΠΟΙΟΥΜΕ BACKWARD/FORWARD STEPWISE METHODS



ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ

Διαφάνεια 14-51

14. ΛΟΓΙΣΤΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ

14.8. Παράδειγμα πολλαπλής παλινδρόμησης

Omnibus Tests of Model Coefficients

| | | Chi-square | df | Sig. |
|----------------------|-------|------------|----|------|
| Step 1 ^a | Step | 23.011 | 13 | .042 |
| | Block | 23.011 | 13 | .042 |
| | Model | 23.011 | 13 | .042 |
| Step 2 ^a | Step | -.016 | 1 | .899 |
| | Block | 22.995 | 12 | .028 |
| | Model | 22.995 | 12 | .028 |
| ... | | | | |
| Step 9 ^a | Step | -1.736 | 1 | .188 |
| | Block | 17.454 | 4 | .002 |
| | Model | 17.454 | 4 | .002 |
| Step 10 ^a | Step | -1.703 | 1 | .192 |
| | Block | 15.752 | 3 | .001 |
| | Model | 15.752 | 3 | .001 |

a. A negative Chi-squares value indicates that the Chi-squares value has decreased from the previous step.

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ

Διαφάνεια 14-52

14. ΛΟΓΙΣΤΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ
14.8. Παράδειγμα πολλαπλής παλινδρόμησης

ΧΡΗΣΙΜΟΠΟΙΟΥΜΕ BACKWARD/FORWARD STEPWISE METHODS

Model Summary

| Step | -2 Log likelihood | Cox & Snell R Square | Nagelkerke R Square |
|------|----------------------|----------------------|---------------------|
| 1 | 140.537 ^a | .169 | .231 |
| 2 | 140.553 ^a | .169 | .231 |
| 3 | 140.594 ^a | .169 | .231 |
| 4 | 140.708 ^a | .168 | .230 |
| 5 | 141.103 ^a | .166 | .226 |
| 6 | 142.870 ^a | .154 | .210 |
| 7 | 143.656 ^a | .148 | .202 |
| 8 | 144.358 ^a | .143 | .196 |
| 9 | 146.094 ^b | .131 | .179 |
| 10 | 147.796 ^b | .119 | .163 |

a. Estimation terminated at iteration number 5 because parameter estimates changed by less than .001.
b. Estimation terminated at iteration number 4 because parameter estimates changed by less than .001.

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ Διαφάνεια 14-53

14. ΛΟΓΙΣΤΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ
14.8. Παράδειγμα πολλαπλής παλινδρόμησης

Model if Term Removed

| Variable | Model Log Likelihood | Change in -2 Log Likelihood | df | Sig. of the Change |
|-------------|----------------------|-----------------------------|----|--------------------|
| Step 10 iqp | -77.857 | 7.917 | 1 | .005 |
| pica | -76.685 | 5.574 | 1 | .018 |
| fst2yrs | -76.319 | 4.841 | 1 | .028 |

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ Διαφάνεια 14-55

14. ΛΟΓΙΣΤΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ
14.8. Παράδειγμα πολλαπλής παλινδρόμησης

ΧΡΗΣΙΜΟΠΟΙΟΥΜΕ BACKWARD/FORWARD STEPWISE METHODS

Hosmer and Lemeshow Test

| Step | Chi-square | df | Sig. |
|------|------------|----|------|
| 1 | 7.529 | 8 | .481 |
| 2 | 1.785 | 8 | .987 |
| 3 | 4.043 | 8 | .853 |
| 4 | 7.003 | 8 | .536 |
| 5 | 3.471 | 8 | .901 |
| 6 | 14.575 | 8 | .068 |
| 7 | 5.172 | 8 | .739 |
| 8 | 10.164 | 8 | .254 |
| 9 | 8.225 | 8 | .412 |
| 10 | 7.565 | 8 | .477 |

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ Διαφάνεια 14-54

14. ΛΟΓΙΣΤΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ
14.8. Παράδειγμα πολλαπλής παλινδρόμησης

ΧΡΗΣΙΜΟΠΟΙΟΥΜΕ BACKWARD/FORWARD STEPWISE METHODS

Variables in the Equation

| | B | S.E. | Wald | df | Sig. | Exp.(B) |
|-----------------|-------|-------|-------|----|------|---------|
| Step 10 pica(1) | 1.597 | .693 | 5.301 | 1 | .021 | 4.937 |
| fst2yrs(1) | .960 | .438 | 4.792 | 1 | .029 | 2.611 |
| iqp | -.036 | .014 | 6.994 | 1 | .008 | .965 |
| Constant | 2.585 | 1.336 | 3.746 | 1 | .053 | 13.268 |

a. Variable(s) entered on step 1: gender, totyrs, pica, colic, clumsiness, irritab, convuls, area, fst2yrs, iqp, iap, iqtype.

LOG(ODDS) = 7.7 -0.036 IQP +1.6 PICA +0.96 AGE

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ Διαφάνεια 14-56

14. ΛΟΓΙΣΤΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ

14.8. Παράδειγμα πολλαπλής παλινδρόμησης

| Variables in the Equation | | | | | | | |
|---------------------------|------------|-------|-------|-------|----|------|--------|
| | | B | S.E. | Wald | df | Sig. | Exp(B) |
| Step | pica(1) | 1.597 | .693 | 5.301 | 1 | .021 | 4.937 |
| 10 | fst2yrs(1) | .960 | .438 | 4.792 | 1 | .029 | 2.611 |
| | iqp | -.036 | .014 | 6.994 | 1 | .008 | .965 |
| | Constant | 2.585 | 1.336 | 3.746 | 1 | .053 | 13.268 |

- Κάθε επιπλέον θετική απάντηση στο IQP μειώνει τη σχετική πιθανότητα εμφάνισης μολύβδου στο αίμα κατά 3.5% (διατηρώντας τις υπόλοιπες μεταβλητές σταθερές)

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ

Διαφάνεια 14-57

14. ΛΟΓΙΣΤΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ

14.8. Παράδειγμα πολλαπλής παλινδρόμησης

| Variables in the Equation | | | | | | | |
|---------------------------|------------|-------|-------|-------|----|------|--------|
| | | B | S.E. | Wald | df | Sig. | Exp(B) |
| Step | pica(1) | 1.597 | .693 | 5.301 | 1 | .021 | 4.937 |
| 10 | fst2yrs(1) | .960 | .438 | 4.792 | 1 | .029 | 2.611 |
| | iqp | -.036 | .014 | 6.994 | 1 | .008 | .965 |
| | Constant | 2.585 | 1.336 | 3.746 | 1 | .053 | 13.268 |

- Παιδιά με το σύμπτωμα PICA έχουν σχετική πιθανότητα εμφάνισης μολύβδου 5πλάσια της αντίστοιχης για παιδιά χωρίς το σύμπτωμα (και τα υπόλοιπα χαρακτηριστικά ίδια – σταθερά)

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ

Διαφάνεια 14-58

14. ΛΟΓΙΣΤΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ

14.8. Παράδειγμα πολλαπλής παλινδρόμησης

| Variables in the Equation | | | | | | | |
|---------------------------|------------|-------|-------|-------|----|------|--------|
| | | B | S.E. | Wald | df | Sig. | Exp(B) |
| Step | pica(1) | 1.597 | .693 | 5.301 | 1 | .021 | 4.937 |
| 10 | fst2yrs(1) | .960 | .438 | 4.792 | 1 | .029 | 2.611 |
| | iqp | -.036 | .014 | 6.994 | 1 | .008 | .965 |
| | Constant | 2.585 | 1.336 | 3.746 | 1 | .053 | 13.268 |

- Αν το παιδί διέμενε κοντά στο εργοστάσιο τα δύο πρώτα έτη της ζωής του τότε η σχετική πιθανότητα εμφάνισης μολύβδου είναι 2.6 φορές την αντίστοιχη πιθανότητα για παιδιά που δεν διέμεναν κοντά στο εργοστάσιο τα 2 πρώτα έτη της ζωής τους (και τα υπόλοιπα χαρακτηριστικά ίδια – σταθερά)

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ

Διαφάνεια 14-59

14. ΛΟΓΙΣΤΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ

14.8. Παράδειγμα πολλαπλής παλινδρόμησης

| Variables in the Equation | | | | | | | |
|---------------------------|------------|-------|-------|-------|----|------|--------|
| | | B | S.E. | Wald | df | Sig. | Exp(B) |
| Step | pica(1) | 1.597 | .693 | 5.301 | 1 | .021 | 4.937 |
| 10 | fst2yrs(1) | .960 | .438 | 4.792 | 1 | .029 | 2.611 |
| | iqp | -.036 | .014 | 6.994 | 1 | .008 | .965 |
| | Constant | 2.585 | 1.336 | 3.746 | 1 | .053 | 13.268 |

- Λόγοι σχετικών πιθανοτήτων διορθωμένοι (odds ratios adjusted) ως προς τις υπόλοιπες επεξηγηματικές μεταβλητές.

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ

Διαφάνεια 14-60