

Άσκηση 9. Σε μία case-control μελέτη συγκροτήθηκε μία ομάδα από ασθενείς με έλκος (case group) και μία ομάδα από άτομα χωρίς έλκος (control group) με παρόμοια χαρακτηριστικά με την πρώτη ομάδα (ηλικία, φύλο, κοινωνικοοικονομική θέση). Οι ασθενείς με έλκος χωρίστηκαν ανάλογα με το είδος του έλκους (Α: γαστρικό, Β: δωδεκαδάκτυλου). Εξετάσθηκε πόσα άτομα σε κάθε ομάδα κάνουν χρήση ασπιρίνης. Τα αποτελέσματα δίνονται στον παρακάτω πίνακα:

	Χρήση Ασπιρίνης		Σύνολο
	Μη χρήστης	Χρήστης	
τύπος έλκους Α			
<i>Cases</i>	39	25	64
<i>Controls</i>	62	6	68
τύπος έλκους Β			
<i>Cases</i>	49	8	57
<i>Controls</i>	53	8	61

- α) Έχει σχέση η χρήση ασπιρίνης με το έλκος τύπου Α;
 β) Έχει σχέση η χρήση ασπιρίνης με το έλκος τύπου Β;
 γ) Επιφέρει η χρήση ασπιρίνης τα ίδια αποτελέσματα και στους δύο τύπους έλκους;

Απαντήσεις. Θέτουμε

X = ασθενής (1 = case (με έλκος), 2 = control (χωρίς έλκος)),

Y = χρήση ασπιρίνης (1 = Μη χρήστης, 2 = Χρήστης),

Z = τύπος έλκους (1 = τύπος Α, 2 = τύπος Β).

Περνάμε τα δεδομένα στον SPSS Data Editor ως εξής:

n	x	y	z
39	1	1	1
62	2	1	1
49	1	1	2
53	2	1	2
25	1	2	1
6	2	2	1
8	1	2	2
8	2	2	2

Παρατηρούμε ότι όλες οι γραμμές του πίνακα συνάφειας είναι προκαθορισμένες (δηλ. Y : μεταβλητή απόκρισης, X, Z ερμηνευτικές μεταβλητές). Αρχικά ας δούμε ξεχωριστά το ερώτημα (α) και ξεχωριστά το ερώτημα (β). Στη συνέχεια θα τα εξετάσουμε και τα δύο μαζί.

Το ερώτημα (α) μπορεί να απαντηθεί χρησιμοποιώντας μόνο τον διδιάστατο υποπίνακα που αφορά το έλκος τύπου Α, δηλαδή τον πίνακα συνάφειας

	Μη χρήστης Ασπιρίνης	Χρήστης Ασπιρίνης	Σύνολο
<i>Cases έλκους Α</i>	39 (60.9%)	25 (39.1%)	64 (100%)
<i>Controls</i>	62 (91.2%)	6 (8.8%)	68 (100%)

Σε αυτόν τον πίνακα συνάφειας τα αθροίσματα των γραμμών είναι προκαθορισμένα (Y : μεταβλητή απόκρισης, X :ερμηνευτική μεταβλητή, $Z = 1$). Επομένως το ερώτημα (α) μεταφράζεται στο συγκεκριμένο μοντέλο στο ερώτημα :

Έχει η Y ίδια κατανομή σε όλες τις στάθμες της X ; (δηλ. για $X=1$ και $X=2$)

Αν η Y έχει ίδια κατανομή σε όλες τις στάθμες της X τότε σύμφωνα με τη θεωρία θα πρέπει να ισχύει το loglinear μοντέλο (X, Y) ($\log m_{ij} = \mu + \lambda_i^X + \lambda_j^Y$). Εφαρμόζουμε το συγκεκριμένο μοντέλο στο SPSS ως εξής:

Θα χρησιμοποιήσουμε μόνο τον υποπίνακα για τον οποίο ισχύει ότι $Z=1$ (τύπος έλκους Α). Για να απενεργοποιήσουμε τα δεδομένα για τα οποία ισχύει $Z=2$ (τύπος έλκους Β) εισάγουμε (compute) μία νέα μεταβλητή $Z1=2-Z$ (δηλ. $Z1=1$ αν $Z=1$ και $Z1=0$ αν $Z=2$). Στον πίνακα δεδομένων επιλέγουμε μόνο τις γραμμές για τις οποίες $Z1 \neq 0$ ($Z=1$) εκτελώντας

`/Data/Select cases/Use filter variable: Z1`

Στη συνέχεια εφαρμόζουμε το μοντέλο (X, Y) εκτελώντας:

`/Data/Weight cases/Weight cases by: n`
`/Analyze/Loglinear/general/Factors: x,y (custom Model: main effects x,y, Options: Display Estimates)`

από όπου λαμβάνουμε

	Μη χρήστης Ασπιρίνης		Χρήστης Ασπιρίνης	
	Observed	Expected	Observed	Expected
<i>Cases έλκους Α</i>	39 (60.9%)	48,97 (76.5%)	25 (39.1%)	15,03 (23.5%)
<i>Controls</i>	62 (91.2%)	52,03 (76.5%)	6 (8.8%)	15,97 (23.5%)

(Εφόσον έχουμε υποθέσει ότι η Y έχει ίδια κατανομή σε όλες τις στάθμες της X οι αναμενόμενες πιθανότητες στις στήλες είναι ίσες)

Goodness-of-fit Statistics

	Chi-Square	DF	Sig.
Likelihood Ratio	17,6755	1	3,E-05
Pearson	16,7770	1	4,E-05

Το μοντέλο (X, Y) δεν εφαρμόζει καλά στα δεδομένα διότι η πιθανότητα να εμφανιστεί ένα τέτοιο δείγμα υπό την H_0 (H_0 : ισχύει το μοντέλο (X, Y)) είναι μόλις $4 \times 10^{-5} = 0.00004$ (p-value που προκύπτει από τη στατιστική συνάρτηση χι-τετράγωνο του Pearson). Επομένως απορρίπτουμε ότι το μοντέλο (X, Y) είναι σωστό, απορρίπτουμε δηλαδή ότι η Y έχει ίδια κατανομή σε όλες τις στάθμες της X . **Αρα υπάρχει σχέση μεταξύ της χρήσης ασπιρίνης και του έλκους τύπου Α** (μάλιστα είναι φανερό ότι η χρήση ασπιρίνης αυξάνει την πιθανότητα εμφάνισης έλκους τύπου Α).

Με ανάλογο τρόπο μπορούμε να απαντήσουμε στο ερώτημα (β). Σε αυτή την περίπτωση εκτελούμε

`/Compute Z2 = Z-1`

`/Data/Select cases/Use filter variable: Z2`

`/Data/Weight cases/Weight cases by: n`

`/Analyze/Loglinear/general/Factors: x,y (custom Model: main effects x,y, Options: Display Estimates)`

από όπου λαμβάνουμε:

	Μη χρήστης Ασπιρίνης		Χρήστης Ασπιρίνης	
	Observed	Expected	Observed	Expected
<i>Cases έλκους Β</i>	49 (86%)	49,27 (86.4%)	8 (14%)	7,73 (13.6%)
<i>Controls</i>	53 (86.9%)	52,73 (86.4%)	8 (13.1%)	8,27 (13.6%)

Goodness-of-fit Statistics

	Chi-Square	DF	Sig.
Likelihood Ratio	,0213	1	,8840
Pearson	,0213	1	,8840

Το μοντέλο (X, Y) εφαρμόζει αρκετά καλά στα δεδομένα διότι η πιθανότητα να εμφανιστεί ένα τέτοιο δείγμα υπό την H_0 (H_0 : ισχύει το μοντέλο (X, Y)) είναι 88,4%. Επομένως δεν μπορούμε να απορρίψουμε ότι το μοντέλο (X, Y) είναι σωστό (η Y έχει ίδια κατανομή σε όλες τις στάθμες της X). **Αρα δεν υπάρχουν αρκετά στοιχεία που να μας οδηγούν στο ότι υπάρχει σχέση μεταξύ της χρήσης ασπιρίνης και του έλκους τύπου B.**

Τέλος μπορούμε να δούμε μαζί τα ερωτήματα (α) και (β) στον τριδιάστατο πίνακα συνάφειας επαναδιατυπώνοντάς τα σε μία ερώτηση:

Έχει σχέση η χρήση ασπιρίνης (Y) με το έλκος (X);

Στον τριδιάστατο πίνακα συνάφειας τα αθροίσματα των γραμμών θεωρούνται προκαθορισμένα και επομένως το παραπάνω ερώτημα μεταφράζεται στο εξής:

Έχει η Y ίδια κατανομή σε όλες τις στάθμες της X (για δεδομένη στάθμη της Z);

Το μοντέλο του παραπάνω ερωτήματος αντιστοιχεί (βλ. θεωρία) στο loglinear μοντέλο (YZ, XZ) (από την συνθήκη Birch θα πρέπει οπωσδήποτε να συμπεριληφθούν στο loglinear μοντέλο και οι όροι $\mu + \lambda_j^X + \lambda_k^Z + \lambda_{jk}^{XZ}$ (XZ) ώστε τα αντίστοιχα εκτιμημένα αθροίσματα συχνοτήτων των γραμμών να είναι ίσα με τα προκαθορισμένα). Εφαρμόζουμε το loglinear μοντέλο (YZ, XZ) εκτελώντας

```
/Data/Select cases/all cases
/Data/Weight cases/Weight cases by: n
/Analyze/Loglinear/general/Factors: x,y,z (custom Model: main effects x, y, z ,
Interactions y*z, x*z, Options:Display Estimates)
```

από όπου λαμβάνουμε:

	Χρήση Ασπιρίνης				Σύνολο
	Μη χρήστης		Χρήστης		
	Observed	Expected	Observed	Expected	
τύπος έλκους Α					
<i>Cases</i>	39 (60.9%)	48,97 (76.5%)	25 (39.1%)	15,03 (23.5%)	64 (100%)
<i>Controls</i>	62 (91.2%)	52,03 (76.5%)	6 (8.8%)	15,97 (23.5%)	68 (100%)
τύπος έλκους Β					
<i>Cases</i>	49 (86%)	49,27 (86.4%)	8 (14%)	7,73 (13.6%)	57 (100%)
<i>Controls</i>	53 (86.9%)	52,73 (86.4%)	8 (13.1%)	8,27 (13.6%)	61 (100%)
Goodness-of-fit Statistics					
	Chi-Square	DF	Sig.		
Likelihood Ratio	17,6967	2	,0001		
Pearson	16,7983	2	,0002		

Επειδή $p\text{-value} = 0.0002$ απορρίπτουμε ότι το μοντέλο (YZ, XZ) είναι σωστό, και άρα υπάρχει σχέση μεταξύ της χρήσης ασπιρίνης (Y) και του έλκους (X).

γ) Από τον αμέσως παραπάνω έλεγχο είδαμε ότι υπάρχει σχέση μεταξύ της χρήσης ασπιρίνης (Y) και του έλκους (X). Επομένως φυσιολογικά γεννάται το ερώτημα αν η σχέση που δεχτήκαμε ότι υπάρχει (μεταξύ X και Y) είναι η ίδια σε όλες τις στάθμες της Z (το οποίο είναι ισοδύναμο με το ερώτημα (γ))

Από την ξεχωριστή εξέταση των δύο υποπινάκων παραπάνω προκύπτει ότι η σχέση μεταξύ των X, Y είναι διαφορετική για $Z=1$ και για $Z=2$ (διότι είδαμε ότι η Y εξαρτάται από τη X για $Z=1$ ενώ η Y είναι ανεξάρτητη της X για $Z=2$).

Εναλλακτικά, μπορούμε να εξετάσουμε το παραπάνω ερώτημα (η σχέση μεταξύ X και Y είναι η ίδια σε όλες τις στάθμες της Z ;) και στον τριδιάστατο πίνακα συνάφειας. Συγκεκριμένα, εφαρμόζουμε το loglinear μοντέλο (XZ, XY) το οποίο αντιστοιχεί στην υπόθεση ότι η Y έχει ίδια κατανομή σε όλες τις στάθμες της Z (για δεδομένη στάθμη της X). Όμοια με παραπάνω λαμβάνουμε ότι

	Χρήση Ασπιρίνης				Σύνολο
	Μη χρήστης		Χρήστης		
	Observed	Expected	Observed	Expected	
τύπος έλκους Α					
<i>Cases</i>	39 (60.9%)	46,55 (72.7%)	25 (39.1%)	17,45 (27.3%)	64 (100%)
<i>Controls</i>	62 (91.2%)	60,62 (89.1%)	6 (8.8%)	7,38 (10.9%)	68 (100%)
τύπος έλκους Β					
<i>Cases</i>	49 (86%)	41,45 (72.7%)	8 (14%)	15,55 (27.3%)	57 (100%)
<i>Controls</i>	53 (86.9%)	54,38 (89.1%)	8 (13.1%)	6,62 (10.9%)	61 (100%)
Goodness-of-fit Statistics					
	Chi-Square	DF	Sig.		
Likelihood Ratio	10,5384	2	,0051		
Pearson	10,1329	2	,0063		

και επομένως (όπως ήταν και αναμενόμενο) η χρήση ασπιρίνης δεν επιφέρει τα ίδια αποτελέσματα και στους δύο τύπους έλκους (p-value=0.0063).