

Άσκηση 10. Ο ακόλουθος πίνακας βασίζεται σε στοιχεία που δόθηκαν από το τμήμα ασφαλείας αυτοκινητοδρόμων της πολιτείας της Florida για το έτος 1988. Οι επιβάτες που συμμετείχαν σε ατυχήματα έχουν ταξινομηθεί ανάλογα με το αν φορούσαν ζώνη ασφαλείας, αν εκτινάχθηκαν εκτός αυτοκινήτου, και αν επέζησαν.

	Εκτινάχθηκαν	Τραυματισμός	
		Όχι Μοιραίος	Μοιραίος
Με ζώνη ασφαλείας	NAI	1105	14
	OXI	411111	483
Χωρίς ζώνη ασφαλείας	NAI	4624	497
	OXI	157342	1008

1. Αν η χρήση ζώνης ασφαλείας ήταν ανεξάρτητη από το αν κάποιος εκτινάχθηκε και από το αν κάποιος επέζησε, ποιες θα έπρεπε να είναι οι (εκτιμημένες) αναμενόμενες συχνότητες; Επίσης, σε αυτή την περίπτωση να δώσετε τις παρατηρούμενες και εκτιμημένες πιθανότητες στα 8 κελιά του πίνακα συνάφειας.
2. Τίθεται ο ισχυρισμός ότι από τη στιγμή που γνωρίζουμε αν κάποιος εκτινάχθηκε από το αυτοκίνητο ή όχι, το αν θα επιζήσει δεν εξαρτάται από το αν φορούσε ζώνη. Είναι σωστός αυτός ο ισχυρισμός ($\alpha = 5\%$); Δώστε πιθανή εξήγηση.
3. Εξετάστε ποιο loglinear μοντέλο ταιριάζει καλύτερα στα δεδομένα. Ερμηνεύστε τα αποτελέσματα.

Απαντήσεις Άσκησης 10. Θέτουμε

X = Εκτίναξη (1 = NAI, 2 = OXI),

Y = Τραυματισμός (1 = Όχι Μοιραίος, 2 = Μοιραίος),

Z = Ζώνη Ασφαλείας (1 = Με ζώνη, 2 = Χωρίς ζώνη).

Περνάμε τα δεδομένα στον SPSS Data Editor ως εξής:

n	x	y	z
1105	1	1	1
411111	2	1	1
4624	1	1	2
157342	2	1	2
14	1	2	1
483	2	2	1
497	1	2	2
1008	2	2	2

Παρατηρούμε ότι καμία γραμμή ή στήλη του πίνακα συνάφειας δεν είναι προκαθορισμένη (δηλ. Y, X, Z : μεταβλητές απόκρισης).

(1). Αν η χρήση ζώνης ασφαλείας (Z) είναι ανεξάρτητη από το αν κάποιος εκτινάχθηκε (X) και από το αν κάποιος επέζησε (Y), τότε σύμφωνα με τη θεωρία θα πρέπει να ισχύει το loglinear μοντέλο (Z, XY) ($\log m_{ijk} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY}$). Εφαρμόζουμε το συγκεκριμένο μοντέλο στο SPSS ως εξής:

/Data/Weight cases/Weight cases by: n

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

από όπου λαμβάνουμε

Τραυματισμός				
	Όχι Μοιραίος		Μοιραίος	
	Observed	Expected	Observed	Expected
Με ζώνη ασφαλείας				
Εκτινάχθηκαν	1105 (0,19%)	4103,61 (0,71%)	14 (0%)	366,02 (0,06%)
Δεν Εκτινάχθηκαν	411111 (71,35%)	407175,4 (70,67%)	483 (0,08%)	1067,98 (0,19%)
Χωρίς ζώνη ασφαλείας				
Εκτινάχθηκαν	4624 (0,80%)	1625,39 (0,28%)	497 (0,09%)	144,98 (0,03%)
Δεν Εκτινάχθηκαν	157342 (27,31%)	161277,61 (27,99%)	1008 (0,17%)	423,02 (0,07%)
Goodness-of-fit Statistics				
	Chi-Square	DF	Sig.	
Likelihood Ratio	9021,2887	3	,0000	
Pearson	10179,9180	3	,0000	

Οι (εκτιμημένες) αναμενόμενες συχνότητες δίνονται στον παραπάνω πίνακα μαζί με τις παρατηρούμενες και τα αντίστοιχα ποσοστά. Είναι φανερό ότι υπάρχει μεγάλη διαφορά μεταξύ των παρατηρούμενων και των εκτιμώμενων συχνοτήτων (π.χ. αν το Z ήταν ανεξάρτητο από τα X, Y , τότε θα περιμέναμε περίπου 366 θανατηφόρα περιστατικά ατόμων που φορούσαν ζώνη και εκτινάχθηκαν, ενώ παρατηρούμε μόνο 14!). Το γεγονός αυτό αποτελεί μία ένδειξη για το ότι δεν πρέπει να είναι σωστό το μοντέλο. Αυτό φαίνεται καθαρά από το αντίστοιχο p-value (πιθανότητα να εμφανιστεί ένα τέτοιο δείγμα όταν ισχύει το μοντέλο) το οποίο είναι σχεδόν 0. Επομένως απορρίπτουμε ότι το μοντέλο (Z, XY) είναι σωστό, απορρίπτουμε δηλαδή την υπόθεση ότι η Z είναι ανεξάρτητη από τις X, Y .

(2). Ο ισχυρισμός αυτός, είναι ισοδύναμος με το ότι οι Z, Y είναι ανεξάρτητες δεδομένης της X . Επομένως αν ισχύει αυτός ο ισχυρισμός θα πρέπει (σύμφωνα με τη θεωρία) να ισχύει το loglinear μοντέλο (XZ, XY) ($\log m_{ijk} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ}$). Εφαρμόζουμε το συγκεκριμένο μοντέλο στο SPSS ως εξής:

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

από όπου λαμβάνουμε:

Τραυματισμός				
	Όχι Μοιραίος		Μοιραίος	
	Observed	Expected	Observed	Expected
Με ζώνη ασφαλείας				
Εκτινάχθηκαν	1105 (0,19%)	1027,36 (0,18%)	14 (0%)	91,64 (0,02%)
Δεν Εκτινάχθηκαν	411111 (71,35%)	410517,25 (71,25%)	483 (0,08%)	1076,75 (0,19%)
Χωρίς ζώνη ασφαλείας				
Εκτινάχθηκαν	4624 (0,80%)	4701,64 (0,82%)	497 (0,09%)	419,36 (0,07%)
Δεν Εκτινάχθηκαν	157342 (27,31%)	157935,75 (27,41%)	1008 (0,17%)	414,25 (0,07%)
Goodness-of-fit Statistics				
	Chi-Square	DF	Sig.	
Likelihood Ratio	1144,6363	2	3, -249	
Pearson	1268,8218	2	3, -276	

Επειδή το αντίστοιχο p-value είναι πολύ μικρό, απορρίπτουμε ότι ισχύει το συγκεκριμένο μοντέλο, δηλαδή απορρίπτουμε τον ισχυρισμό αυτό. Εδώ ίσως θα περίμενε κανείς ότι από τη στιγμή που κάποιος εκτινάχθηκε, η σοβαρότητα του τραυματισμού του δεν επηρεάζεται από το αν φορούσε ή όχι ζώνη. Σύμφωνα όμως με τα δεδομένα κάτι τέτοιο δεν φαίνεται να συμβαίνει. Σχετικά μπορεί να δοθεί η εξής πιθανή εξήγηση: κάποιος που εκτινάσσεται από το αυτοκίνητο ενώ φορούσε ζώνη, θα εκτιναχθεί με πολύ μικρότερη ορμή (σε σχέση με

την περίπτωση που δεν θα φορούσε ζώνη) διότι η ζώνη απορροφά ένα μέρος της ορμής αυτής. Επομένως, και ο τραυματισμός του ως συνέπεια της εκτίναξής του από το αυτοκίνητο είναι πιθανότερο να μην είναι θανάσιμος. Επίσης, κάποιος που δεν εκτινάσσεται από το αυτοκίνητο ενώ φορούσε ζώνη, είναι πιθανότερο να μην χτυπήσει με μεγάλη ορμή στο εσωτερικό του αυτοκινήτου και άρα ο τραυματισμός του είναι πιθανότερο να μην είναι θανάσιμος.

Οι παραπάνω επισημάνσεις γίνονται φανερές και από την εξέταση των παρακάτω δύο υποπινάκων (για $X=1$, $X=2$). Οι εκτιμημένες συχνότητες που δίνονται είναι αυτές που προβλέπονται κάτω από την υπόθεση της ανεξαρτησίας μεταξύ των Y , Z)

Υποπίνακας 1. Περιπτώσεις όπου σημειώθηκε εκτίναξη εκτός αυτοκινήτου:

	Τραυματισμός					
	Όχι Μοιραίος			Μοιραίος		
	observed	expected	dev. residual	observed	expected	dev. residual
Με ζώνη ασφαλείας	1105	1027,36	(2,39)	14	91,64	(-10,13)
Χωρίς ζώνη ασφαλείας	4624	4701,64	(-1,14)	497	419,36	(3,68)

(εκτιμημένες συχνότητες από το μοντέλο $\log m_{ijk} = \mu + \lambda_j^Y + \lambda_k^Z$, $X = 1$)

Goodness-of-fit Statistics	Chi-Square	DF	Sig.
Likelihood Ratio	123,2398	1	1,E-28
Pearson	87,2964	1	9,E-21

Παρατηρούμε ότι το μοντέλο απορρίπτεται κυρίως λόγω του μεγάλου κατάλοιπου (-10.13) που αφορά την περίπτωση μοιραίου τραυματισμού ατόμων που φορούσαν ζώνη ασφαλείας δεδομένου ότι εκτινάχθηκαν. Αν η σοβαρότητα του τραυματισμού ήταν ανεξάρτητη από τη χρήση ζώνης ασφαλείας (δεδομένης της εκτίναξης), θα αναμένονταν περίπου 91 μοιραίοι τραυματισμοί ατόμων που φορούσαν ζώνη (δεδομένου πάντα ότι εκτινάχθηκαν) ενώ παρατηρήθηκαν μόλις 14 (δηλ. η χρήση ζώνης μείωσε την πιθανότητα μοιραίου τραυματισμού ακόμη και σε αυτούς που εκτινάχθηκαν εκτός αυτοκινήτου).

Όμοια συμπεράσματα προκύπτουν και από τον ακόλουθο πίνακα:

Υποπίνακας 2. Περιπτώσεις όπου δεν σημειώθηκε εκτίναξη εκτός αυτοκινήτου:

	Τραυματισμός					
	Όχι Μοιραίος			Μοιραίος		
	observed	expected	dev. residual	observed	expected	dev. residual
Με ζώνη ασφαλείας	411111	410517,25	(0,93)	483	1076,75	(-20,32)
Χωρίς ζώνη ασφαλείας	157342	157935,75	(-1,49)	1008	414,25	(24,60)

(εκτιμημένες συχνότητες από το μοντέλο $\log m_{ijk} = \mu + \lambda_j^Y + \lambda_k^Z$, $X = 2$)

Goodness-of-fit Statistics	Chi-Square	DF	Sig.
Likelihood Ratio	1021,3966	1	4,-224
Pearson	1181,5255	1	6,-259

Παρατηρούμε ότι και αυτό το μοντέλο απορρίπτεται κυρίως λόγω των μεγάλων καταλοίπων (-20.32, 24.60). Αν η σοβαρότητα του τραυματισμού ήταν ανεξάρτητη από τη χρήση ζώνης ασφαλείας (δεδομένου ότι δεν εκτινάχθηκαν), θα αναμένονταν περίπου 1076 μοιραίοι τραυματισμοί ατόμων που φορούσαν ζώνη (δεδομένου πάντα ότι δεν εκτινάχθηκαν) ενώ παρατηρήθηκαν μόλις 483. (δηλ. η χρήση ζώνης μείωσε την πιθανότητα μοιραίου τραυματισμού και σε αυτούς που δεν εκτινάχθηκαν).

3. Εξετάζοντας όλα τα δυνατά μοντέλα προκύπτει ο παρακάτω πίνακας:

Μοντέλο	Likelihood Ratio	X ² Pearson's	d.f.	p-value
(X,Y,Z)	11444.37	47427.63	4	0.0000, 0.0000
(XY,Z)	9021.28	10179.91	3	0.0000, 0.0000
(XZ,Y)	3567.72	13988.95	3	0.0000, 0.0000

(YZ,Y)	9557.06	21648.84	3	0.0000, 0.0000
(XY,XZ)	1144.63	1268.82	2	0.0000, 0.0000
(XY,YZ)	7133.97	8056.49	2	0.0000, 0.0000
(XZ,YZ)	1680.41	4591.56	2	0.0000, 0.0000
(XY,XZ,YZ)	2.85	2.55	1	0.0911, 0.1099

Άρα μόνο το μοντέλο (XY,XZ,YZ) είναι αποδεκτό. Ο πίνακας συνάφειας με τις αναμενόμενες και παρατηρούμενες συχνότητες και ποσοστά για το μοντέλο (XY,XZ,YZ) είναι ο ακόλουθος:

	Τραυματισμός			
	Όχι Μοιραίος		Μοιραίος	
	Observed	Expected	Observed	Expected
Με ζώνη ασφαλείας				
Εκτινάχθηκαν	1105 (0,19%)	1098,13 (0,19%)	14 (0%)	20,87 (0%)
Δεν Εκτινάχθηκαν	411111 (71,35%)	411117,87 (71,35%)	483 (0,08%)	476,13 (0,08%)
Χωρίς ζώνη ασφαλείας				
Εκτινάχθηκαν	4624 (0,80%)	4630,87 (0,80%)	497 (0,09%)	490,13 (0,09%)
Δεν Εκτινάχθηκαν	157342 (27,31%)	157335,13 (27,31%)	1008 (0,17%)	1014,87 (0,18%)

Σε αυτό το μοντέλο, κανένα ζεύγος μεταβλητών δεν είναι δεσμευμένα ανεξάρτητο (δεδομένης της τρίτης μεταβλητής). Επιπλέον, επειδή για το (XY,XZ,YZ) ισχύει ότι

$$\log(m\pi_{ijk}) = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ}$$

προκύπτει εύκολα ότι,

$$\begin{aligned} \log \theta_{11(k)} &= \log \frac{\pi_{11k} \pi_{22k}}{\pi_{12k} \pi_{21k}} = \log \frac{m\pi_{11k} m\pi_{22k}}{m\pi_{12k} m\pi_{21k}} \\ &= \log m\pi_{11k} + \log m\pi_{22k} - \log m\pi_{12k} + \log m\pi_{21k} \\ &= \dots = \lambda_{11}^{XY} + \lambda_{22}^{XY} - \lambda_{12}^{XY} - \lambda_{21}^{XY} \end{aligned}$$

για $k = 1, 2$. Δηλαδή, ο δείκτης συνάφειας *log odds ratio* μεταξύ των X,Y είναι ο ίδιος σε όλες τις στάθμες της Z . Το ίδιο ισχύει και για το *log odds ratio* μεταξύ των X,Z (ίδιο σε όλες τις στάθμες της Y) και για το *log odds ratio* μεταξύ των Y,Z (ίδιο σε όλες τις στάθμες της X). Δηλαδή, σε αυτό το μοντέλο, ο δείκτης συνάφειας μεταξύ δύο μεταβλητών είναι ο ίδιος σε όλα τα επίπεδα της τρίτης μεταβλητής.