ΙΑΤΡΙΚΗ ΣΤΑΤΙΣΤΙΚΗ

Ακαδ. Έτος 2024-2025

Δοκιμασία X^2 κατά ζεύγη (McNemar's test)

- Εφαρμόζεται όταν υπάρχει αντιστοιχία των παρατηρήσεων ανά ζεύγη
  - πχ πριν-μετά στα ίδια άτομα
  - πχ εξομοιωμένα ζεύγη ατόμων

- Στην περίπτωση που πληρούνται οι προϋποθέσεις για να γίνει X^2 κατά ζεύγη, μπορεί να γίνει και απλό?

- Ποιο προτιμούμε και γιατί?

- Το X^2 κατά ζεύγη γιατί είναι πιο ισχυρό, εφόσον βέβαια έχουμε την αντιστοιχία σε ζεύγη.

Δοκιμασία X^2 κατά ζεύγη (McNemar's test)

- Θέλουμε να διερευνήσουμε αν τα προβλήματα αϋπνίας σε εργαζόμενους εμφανίζονται συχνότερα τις καθημερινές ή τα Σαββατοκύριακα. Ρωτάμε 100 άτομα.

Αϋπνία τις καθημερινές	Αϋπνία τα Σαββατοκύριακα	Συχνότητα
+	+	30
-	-	50
+	-	15=ε
-	+	5=ζ

Τύπος για τον υπολογισμό του X^2 κατά ζεύγη:

$$X^2 = \frac{(\varepsilon - \zeta)^2}{\varepsilon + \zeta}$$

ή με τη διόρθωση κατά Yates:

$$X^2 = \frac{(|\varepsilon - \zeta| - 1)^2}{\varepsilon + \zeta}$$

Εφαρμογή του τύπου, ΒΕ, αξιολόγηση

$$X^2 = \frac{(|\varepsilon - \zeta| - 1)^2}{\varepsilon + \zeta} = \frac{(|5 - 15| - 1)^2}{5 + 15} = 4,05$$

ΒΕ πάντα 1 στο X^2 κατά ζεύγη.

ΒΕ	10%	5%	1%	0,1%
1	2,71	3,84	6,64	10,83

3,84 < 4,05 < 6,64
5% > P > 1%

Συμπέρασμα: υπάρχει στατιστικά σημαντική διαφορά στη συχνότητα αϋπνίας μεταξύ καθημερινών και Σαββατοκύριακων
Ερμηνεία: η αϋπνία είναι συχνότερη τις καθημερινές απ'ότι τα Σ/Κ

Παρόλο που όταν μπορούμε να κάνουμε X^2 κατά ζεύγη, το προτιμούμε γιατί είναι πιο ισχυρό, ας δούμε πώς θα ήταν το απλό X^2 στο παράδειγμα μας.

Αϋπνία τις καθημερινές	Αϋπνία τα Σαββατοκύριακα	Συχνότητα
+	+	30
-	-	50
+	-	15=ε
-	+	5=ζ

Προβλήματα αϋπνίας τις καθημερινές και τα Σαββατοκύριακα.

	Αϋπνία		
	Ναι	Όχι	Σύνολο
Καθημερινές	45	55	100
Σαββατοκύριακα	35	65	100
Σύνολο	80	120	200

Τι παρατηρούμε σ' αυτόν τον Πίνακα?

Ο συνολικός αριθμός των παρατηρήσεων είναι διπλάσιος. Γιατί?

Ποιες αναλογίες συγκρίνουμε με αυτόν τον τετράπτυχο Πίνακα?

Προβλήματα αϋπνίας τις καθημερινές και τα Σαββατοκύριακα.			
	Αϋπνία		
	Ναι	Όχι	Σύνολο
Καθημερινές	45	55	100
Σαββατοκύριακα	35	65	100
Σύνολο	80	120	200

Αν η αναλογία αυτών που έχουν αϋπνία τις καθημερινές, δηλαδή 45/100, διαφέρει από την αναλογία αυτών που έχουν αϋπνία το Σαββατοκύριακο, 35/100, ή μήπως οι διαφορές αυτές οφείλονται στην τυχαία δειγματοληπτική διακύμανση?

Αν εφαρμόσουμε τον αντίστοιχο τύπο:

$$X^2 = \frac{(|a \cdot d - b \cdot c| - n/2)^2 \cdot n}{(a+c)(b+d)(a+b)(c+d)}$$

$X^2 = [|(45*65)-(35*55)|-100]^2 *200/(100*100*80*120)= 1,69$

BE=1

BE	10%	5%	1%	0,1%
1	2,71	3,84	6,64	10,83

1,69 < 2,71
P > 10% , Μη Στατιστικά Σημαντικό

Αν από τα ίδια δεδομένα θελήσουμε να διερευνήσουμε ένα άλλο πρόβλημα: Τα άτομα που έχουν αϋπνία τις καθημερινές, έχουν μεγαλύτερη πιθανότητα να έχουν αϋπνία και τα Σαββατοκύριακα σε σχέση με αυτούς που δεν έχουν αϋπνία τις καθημερινές?

- Ποιες αναλογίες θα συγκρίνουμε?

Αϋπνία τις καθημερινές	Αϋπνία τα Σαββατοκύριακα	Συχνότητα
+	+	30
-	-	50
+	-	15=ε
-	+	5=ζ

- Όλα τα άτομα που έχουν αϋπνία τις καθημερινές είναι 45. Απ' αυτούς οι 30 έχουν αϋπνία και τα Σαββατοκύριακα. Άρα η μία αναλογία είναι 30/45

- Τα άτομα που δεν έχουν αϋπνία τις καθημερινές είναι συνολικά 55. Απ' αυτούς έχουν αϋπνία το Σαββατοκύριακο οι 5. Άρα η άλλη αναλογία είναι 5/55.

- Μπορούμε να φτιάξουμε έναν τετράπτυχο Πίνακα για τη σύγκριση αυτών των 2 αναλογιών με το X^2

Έχει σχέση η αϋπνία τις καθημερινές με την αϋπνία το Σαββατοκύριακο?			
	Αϋπνία τις καθημερινές		
Αϋπνία το Σαββατοκύριακο	Ναι	Όχι	Σύνολο
Ναι	30	5	35
Όχι	15	50	65
Σύνολο	45	55	100

$$X^2 = \frac{(|a \cdot d - b \cdot c| - n/2)^2 \cdot n}{(a+c)(b+d)(a+b)(c+d)}$$

X^2 = [|(30*50)-(5*15)|-50]^2 *100/(35*65*45*55)=33,58

BE=1

BE	10%	5%	1%	0,1%
1	2,71	3,84	6,64	10,83

10,83 < 33,58
0,1% > P Στατιστικά Πολύ Σημαντικό

Συμπέρασμα: Υπάρχει πολύ στατιστικά σημαντική σχέση ανάμεσα στην αϋπνία τις καθημερινές και τα Σαββατοκύριακα
Ερμηνεία: Τα άτομα που έχουν αϋπνία τις καθημερινές, έχουν μεγαλύτερη πιθανότητα να έχουν αϋπνία και το Σαββατοκύριακο σε σχέση με αυτούς που δεν έχουν αϋπνία τις καθημερινές

Δοκιμασία X^2 ως κριτήριο καλής εφαρμογής

- Κατανομή των εισαγωγών ασθενών με οξεία λευχαιμία στο νοσοκομείο μιας μεγάλης πόλης κατά εποχή κατά την διάρκεια ενός έτους

	Χειμώνας	Άνοιξη	Καλοκαίρι	Φθινόπωρο	Σύνολο
Εισαγωγές	35	22	20	23	100

ΜΗΔΕΝΙΚΗ-ΕΝΑΛΛΑΚΤΙΚΗ ΥΠΟΘΕΣΗ στο X^2 καλής εφαρμογής

- Ho: Η κατανομή της ποιοτικής μεταβλητής ακολουθεί δεδομένη μορφή (π.χ. ομοιόμορφη, κανονική κτλ)

- H1: Η κατανομή της ποιοτικής μεταβλητής δεν ακολουθεί δεδομένη μορφή (π.χ. ομοιόμορφη, κανονική κτλ)

Δοκιμασία X^2 ως κριτήριο καλής εφαρμογής

- Κατανομή των εισαγωγών ασθενών με οξεία λευχαιμία στο νοσοκομείο μιας μεγάλης πόλης κατά εποχή κατά την διάρκεια ενός έτους

	Χειμώνας	Άνοιξη	Καλοκαίρι	Φθινόπωρο	Σύνολο
Παρατηρηθείσες	35	22	20	23	100
Αναμενόμενες	25	25	25	25	100

$$X^2 = \sum_{i=1}^{4} \frac{(O_i - E_i)^2}{E_i}$$

$X^2 = [(35-25)^2/25] + [(22-25)^2/25] + [(20-25)^2/25] + [(23-25)^2/25] = 5{,}52$

Βαθμοί ελευθερίας στο X^2 καλής εφαρμογής

- Είναι ο αριθμός των κελιών (π.χ. 4 στο παράδειγμα) μείον τον αριθμό των δεδομένων που χρησιμοποιήσαμε για να υπολογίσουμε τις αναμενόμενες συχνότητες

- Πόσα δεδομένα χρησιμοποιήσαμε για να υπολογίσουμε τις αναμενόμενες συχνότητες????

ΒΕ = 4-1 = 3

ΒΕ	10%	5%	1%	0,1%
3	6,25	7,82	11,35	16,27

5,52 < 6,25
P > 10%

Συμπέρασμα: Δεν υπάρχει στατιστικά σημαντική εποχικότητα στις εισαγωγές ασθενών με οξεία λευχαιμία

Αξιολόγηση του βαθμού συσχέτισης

- Ένα στατιστικά σημαντικό εύρημα δεν είναι αναγκαστικά και βιολογικά σημαντικό. Η στατιστική σημαντικότητα συνδέεται με τη βεβαιότητα μας ότι μια διαφορά ή σχέση υπάρχει στον αντίστοιχο πληθυσμό και όχι αναγκαστικά με τη «δύναμη της σχέσης».

- Αν θέλουμε να συγκρίνουμε τα αποτελέσματα 2 δοκιμασιών X^2 ως προς τη δύναμη της σχέσης, αν μεν έχουν τον ίδιο αριθμό ΒΕ τότε συγκρίνουμε τα πηλίκα X^2/n. Αλλιώς $X^2/(n*ΒΕ)$.

Ειδικότερα στους τετράπτυχους Πίνακες υπολογίζουμε τον Σχετικό λόγο

Σχετικός λόγος: πηλίκο διαγωνίων γινομένων

ΣΛ= (a*d)/(b*c)

	Νόσημα	
Παράγοντας έκθεσης	+	-
+	a	b
-	c	d

Μας λέει πόσες φορές μεγαλύτερη πιθανότητα έχουν να νοσήσουν οι εκτεθειμένοι (κατηγορία «+», στον παράγοντα έκθεσης) σε σχέση με τους μη εκτεθειμένους (κατηγορία «-», στον παράγοντα έκθεσης)

Στο παράδειγμα

Κατανομή 2239 γυναικών καρκινοπαθών του μαστού και 1370 μη καρκινοπαθών κατά αναμνηστικό προκλητής ή φυσιολογικής εμμηνόπαυσης			
Τρόπος εμμηνόπαυσης	Καρκινοπαθείς	Μη-καρκινοπαθείς	Σύνολο
Προκλητή εμμ.	a= 469 (21%)	b= 473 (35%)	a+b= 942
Φυσιολογική εμμ.	c= 1770 (79%)	d= 897 (65%)	c+d= 2667
Σύνολο	a+c=2239	b+d= 1370	n= 3609

ΣΛ= (469*897)/(473*1770) =0,50

Ποια είναι η σημασία του ΣΛ σ' αυτό το παράδειγμα?
Αν αλλάζαμε τις κατηγορίες που φαίνονται στις σειρές του Πίνακα (δηλαδή βάζαμε πάνω την φυσιολογική και κάτω την προκλητή εμμηνόπαυση) τι ΣΛ θα υπολογίζαμε?

ΣΛ= (473*1770) /(469*897) =1/0,50=2

Σύγκριση κινδύνων

- Πολλές φορές στην ιατρική μελετάται η σχέση ανάμεσα σε ένα παράγοντα (έκθεση) και ένα αποτέλεσμα (π.χ. νόσημα). Η σχέση αυτή εκφράζεται με τους δείκτες συσχέτισης ή αιτιότητας.

- Ο τρόπος που υπολογίζεται ο κίνδυνος για μια νόσο αφενός και ο δείκτης αιτιότητας αφετέρου εξαρτάται από το σχεδιασμό έρευνας. Θα αναφερθούμε σε δύο βασικούς σχεδιασμούς: την προοπτική έρευνα, ή έρευνα κοορτής και την έρευνα ασθενών- μαρτύρων.

Προοπτική έρευνα

- Σε μια προοπτική έρευνα επιλέγονται εκτεθειμένοι και μη-εκτεθειμένοι (στον παράγοντα που μελετάμε) και παρακολουθούνται διαχρονικά. Στο τέλος του χρόνου παρατήρησης έχουμε τα παρακάτω αποτελέσματα:

	Νόσημα		
Παράγοντας	Ναι	Όχι	Σύνολο
Ναι	a	b	a+b
‘Οχι	c	d	c+d
Σύνολο	a+c	b+d	n

Σύμφωνα με τα παραπάνω επιλέγονται επομένως τα a+b και c+d

Έρευνες ασθενών - μαρτύρων

- Στις έρευνες ασθενών-μαρτύρων επιλέγονται ασθενείς και μάρτυρες (υγιείς ως προς την νόσο) και αναζητούμε στον παρελθόν τους την συχνότητα έκθεσης στον παράγοντα κινδύνου ∞

	Νόσημα		
Παράγοντας	Ναι	Όχι	Σύνολο
Ναι	a	b	a+b
‘Οχι	c	d	c+d
Σύνολο	a+c	b+d	n

Σύμφωνα με τα παραπάνω επιλέγονται επομένως τα a+c και b+d

- Και στα 2 είδη ερευνών χρειάζεται να γίνουν συγκρίσεις μεταξύ ομάδων με διαφορετικά χαρακτηριστικά