

9ο Φροντιστήριο – Στατιστική Ι

Κεφ. 9

Σύγκριση Συχνοτήτων: Το στατιστικό κριτήριο χ^2

Ασκήσεις Κατανόησης

Άσκηση 1

Μία ερευνήτρια θέλει να μελετήσει τη σχέση ανάμεσα στο σωματικό βάρος και στη χώρα προέλευσης. Για να μετρήσει το βάρος των συμμετεχόντων χρησιμοποιεί το βάρος σε κιλά. Μπορεί η ερευνήτρια να χρησιμοποιήσει το κριτήριο χ^2 για να ελέγξει την υπόθεσή της;

Απάντηση:

Όχι δεν μπορεί, γιατί η μία μεταβλητή έχει μετρηθεί σε αναλογική κλίμακα, ενώ για να χρησιμοποιηθεί το κριτήριο χ^2 θα πρέπει οπωσδήποτε και οι δύο μεταβλητές να είναι κατηγορικές.

Άσκηση 2

Το μέγεθος της οικογένειας (3μελής, 4μελής, 5μελής, 6μελής) επηρεάζει την ικανότητα ανάγνωσης (υψηλή, μέτρια, χαμηλή). Είναι αυτή μία ορθή εναλλακτική υπόθεση για να εξετασθεί με τη χρήση του κριτηρίου χ^2 ;

Απάντηση:

Δεν είναι σωστή, γιατί το κριτήριο χ^2 δεν μπορεί να χρησιμοποιηθεί για τον έλεγχο αιτιότητας μεταξύ δύο μεταβλητών (η λέξη “επηρεάζει” δείχνει αιτιότητα).

Άσκηση 3

Ποια η διαφορά ανάμεσα στις πραγματικές και τις αναμενόμενες συχνότητες; Γιατί είναι σημαντικές για τον υπολογισμό του κριτηρίου χ^2 ;

Απάντηση:

Οι πραγματικές συχνότητες είναι οι συχνότητες που προκύπτουν κατά τη διεξαγωγή της έρευνας, ενώ οι αναμενόμενες είναι οι συχνότητες που περιμένουμε να εμφανιστούν αν ίσχυε η μηδενική υπόθεση. Είναι πολύ σημαντικές στον υπολογισμό του χ^2 γιατί το αποτέλεσμα δείχνει στην ουσία το μέγεθος των διαφορών των δύο συχνοτήτων.

Θεωρία:

$$\chi^2 = \sum \frac{(\Pi - A)^2}{A}$$

Π : η πραγματική συχνότητα κάθε κατηγορίας

A : η αναμενόμενη συχνότητα κάθε κατηγορίας

Άσκηση 4

Σύμφωνα με τη θεωρία, αν οι πραγματικές συχνότητες μιας μεταβλητής είναι τυχαίες, ποια θα πρέπει να είναι η σχέση τους με τις αναμενόμενες συχνότητες;

Απάντηση:

Αν οι πραγματικές συχνότητες είναι τυχαίες τότε θα πρέπει να είναι σχεδόν όμοιες με τις αναμενόμενες. Στην ουσία αυτή είναι και η μηδενική μας υπόθεση.

Άσκηση 5

Αν θέλουμε να διαπιστώσουμε κατά πόσο η κατανομή που έχει το δείγμα μας πλησιάζει την κατανομή που θα αναμέναμε να είχε σύμφωνα με τη θεωρία μας, ποιον τύπο κριτηρίου χ^2 θα πρέπει να χρησιμοποιήσουμε;

Απάντηση:

Εφόσον θέλουμε να ελέγξουμε κατά πόσο η κατανομή που έχει το δείγμα μας πλησιάζει την κατανομή που αναμένουμε, θα πρέπει να χρησιμοποιήσουμε το χ^2 ως κριτήριο καλής εφαρμογής.

Άσκηση 6

Πότε χρησιμοποιούμε το κριτήριο χ^2 ως κριτήριο ανεξαρτησίας;

Απάντηση:

Όταν θέλουμε να διαπιστώσουμε τη σχέση μεταξύ δύο μεταβλητών οι οποίες έχουν διαφορετικά επίπεδα μεταξύ τους.

Άσκηση 7

Ένας ερευνητής θέλει να μελετήσει τη σχέση ανάμεσα στην επαγγελματική καθοδήγηση (υπάρχει – δεν υπάρχει) και στη σιγουριά που νιώθει το άτομο για την επαγγελματική του επιλογή (έχει – δεν έχει). Η ανάλυση έδειξε στατιστικώς σημαντικό αποτέλεσμα. Τί συμπεραίνετε για τη σχέση μεταξύ των δύο αυτών μεταβλητών;

Απάντηση:

Συμπεραίνουμε ότι το αν ένα άτομο νιώθει σιγουριά για την επαγγελματική του επιλογή έχει σχέση με το αν έχει δεχθεί επαγγελματική καθοδήγηση ή όχι.

Άσκηση 8

Μία ερευνήτρια θέλησε να μελετήσει τη σχέση μεταξύ νέων πτυχιούχων και ανεργίας. Για τον σκοπό αυτό ζήτησε τα δεδομένα από την Ελληνική Στατιστική Αρχή. Στον παρακάτω πίνακα παρουσιάζονται τα στοιχεία που της έδωσαν. Μπορεί να χρησιμοποιήσει το κριτήριο χ^2 ;

Άνεργοι	Εργαζόμενοι
38%	62%

Απάντηση:

Μία από τις βασικές προϋποθέσεις του κριτηρίου χ^2 είναι τα δεδομένα να είναι εκφρασμένα σε συχνότητες. Εφόσον δίνονται ποσοστά θα μπορούσε να χρησιμοποιηθεί αν ήταν γνωστό το μέγεθος του συνολικού δείγματος, με τη βοήθεια του οποίου θα μετέτρεπε η ερευνήτρια τα ποσοστά σε συχνότητες.

Άσκηση 9

Ποιο πρόβλημα υπάρχει όταν θέλουμε να χρησιμοποιήσουμε το κριτήριο χ^2 για τη σύγκριση ενός πίνακα διπλής εισόδου 2×2 ; Πώς θα αντιμετωπίζουμε ένα τέτοιο πρόβλημα;

Απάντηση:

Πολλοί πιστεύουν ότι σε μία τέτοια περίπτωση υπάρχει αυξημένη πιθανότητα σφάλματος Τύπου I. Αυτό οφείλεται στο γεγονός ότι το χ^2 δεν είναι ασφαλές στην περίπτωση που έχουμε ένα βαθμό ελευθερίας. Για το λόγο αυτό προτείνεται η διόρθωση Yates.

$$[df=(\Sigma-1)(\Gamma-1)=(2-1)(2-1)=1]$$

Άσκηση 10

Ποιο είναι το πιο ισχυρό στατιστικό κριτήριο για τη μελέτη της επίδρασης μιας ανεξάρτητης μεταβλητής στην εξαρτημένη: το κριτήριο χ^2 ή το κριτήριο t (του προηγούμενου κεφαλαίου);

Απάντηση:

Πιο ισχυρό κριτήριο είναι το κριτήριο t , αφενός γιατί είναι ένα παραμετρικό κριτήριο, σε αντίθεση με το χ^2 που είναι μη παραμετρικό (δεν απαιτεί καμία υπόθεση για την ακριβή μορφή της κατανομής του πληθυσμού), αφετέρου γιατί με το κριτήριο t ουσιαστικά μπορούμε να ελέγξουμε την επίδραση της ανεξάρτητης μεταβλητής πάνω στην εξαρτημένη, ενώ με το χ^2 δεν μπορούμε.

ΤΕΛΟΣ