
Introduction to R for Social Sciences

Έλεγχος χ^2

Αναστάσιος Εμβαλωτής & Αικατερίνη Σαργιώτη

Έλεγχος χ^2

- Διερεύνηση της ενδεχόμενης **σχέσης** μεταξύ δύο **ονομαστικών ή κατηγορικών μεταβλητών**
- Αρχική απεικόνιση σχέσης μέσα από **πίνακα συνάφειας** (πίνακας διπλής εισόδου)
- Έλεγχος υποθέσεων:
 - H_0 : Δεν υπάρχει σχέση μεταξύ των μεταβλητών (οι μεταβλητές είναι ανεξάρτητες μεταξύ τους).
 - H_1 : Η μία μεταβλητή εξαρτάται από την άλλη.
- Αν $p\text{-value} \leq \alpha$ (επίπεδο στατιστικής σημαντικότητας) \rightarrow Απορρίπτεται η H_0 (άρα υπάρχει σχέση μεταξύ των δύο μεταβλητών)
- Αν $p\text{-value} > \alpha$ (επίπεδο στατιστικής σημαντικότητας) \rightarrow **ΔΕΝ** απορρίπτεται η H_0 (άρα **ΔΕΝ** υπάρχει σχέση μεταξύ των δύο μεταβλητών)

Προϋποθέσεις εφαρμογής ελέγχου χ^2

1. Το δείγμα δεν θα πρέπει να είναι ούτε πολύ μεγάλο ούτε πολύ μικρό. Προτείνεται ο αριθμός των περιπτώσεων που ελέγχονται να είναι μεταξύ 25 και 250 περιπτώσεων.
2. Η επιλογή του δείγματος θα πρέπει να έχει γίνει με τυχαίο τρόπο και οι παρατηρήσεις να είναι ανεξάρτητες μεταξύ τους.
3. Πρέπει να υπάρχει ικανοποιητικός αριθμός παρατηρήσεων σε κάθε κελί. Αυτό σημαίνει ότι το 75% των κελιών θα πρέπει να έχει τουλάχιστον πέντε παρατηρήσεις σε κάθε κελί ή οι αναμενόμενες συχνότητες κάθε κελιού να είναι τουλάχιστον πέντε.

Παράδειγμα

- Εξετάζεται κατά πόσο η υποστήριξη που δείχνουν οι γονείς στις εκπαιδευτικές προσπάθειες και τα επιτεύγματα των παιδιών τους εξαρτάται από το επίπεδο εκπαίδευσης της μητέρας.
- Υποθέσεις:
 - H_0 : Δεν υπάρχει σχέση μεταξύ του επιπέδου εκπαίδευσης της μητέρας και της υποστήριξης που δείχνουν οι γονείς στις εκπαιδευτικές προσπάθειες και τα επιτεύγματα των παιδιών τους (το επίπεδο εκπαίδευσης της μητέρας είναι ανεξάρτητο από την υποστήριξη που δείχνουν οι γονείς στις εκπαιδευτικές προσπάθειες και τα επιτεύγματα των παιδιών τους).
 - H_1 : Υπάρχει σχέση μεταξύ του επιπέδου εκπαίδευσης της μητέρας και της υποστήριξης που δείχνουν οι γονείς στις εκπαιδευτικές προσπάθειες και τα επιτεύγματα των παιδιών τους (η υποστήριξη που δείχνουν οι γονείς στις εκπαιδευτικές προσπάθειες και τα επιτεύγματα των παιδιών τους εξαρτάται από το επίπεδο εκπαίδευσης της μητέρας).

Εφαρμογή στην R

- Δημιουργία πίνακα συνάφειας με τις μεταβλητές που εξετάζονται

```
Table_x2 = table(Misced, Support)
```

```
Table_x2
```

- Εγκατάσταση πακέτου (“MASS”) για την εφαρμογή του ελέγχου χ^2

```
install.packages(“MASS”)
```

```
library(MASS)
```

- Εφαρμογή του ελέγχου χ^2 με την εντολή *chisq.test(Table_x2)*

```
x2 = chisq.test(Table_x2)
```

```
x2
```

Παραβίαση προϋποθέσεων – Επανακωδικοποίηση

- Επανακωδικοποίηση γραμμών ή στηλών του πίνακα συνάφειας
- Επανακωδικοποίηση γραμμών με την εντολή `rbind()`

```
rtable=rbind(table ["1η τιμή που θέλουμε να επανακωδικοποιήσουμε",] + table ["2η τιμή που θέλουμε να επανακωδικοποιήσουμε",] + table ["3η τιμή που θέλουμε να επανακωδικοποιήσουμε",] + table ["4η τιμή που θέλουμε να επανακωδικοποιήσουμε",] + table ["5η τιμή που θέλουμε να επανακωδικοποιήσουμε", ], table ["6η τιμή που θέλουμε να επανακωδικοποιήσουμε", ] + table ["7η τιμή που θέλουμε να επανακωδικοποιήσουμε", ])
```

- Επανακωδικοποίηση στηλών με την εντολή `cbind()`

```
ctable=cbind(table [, "1η τιμή που θέλουμε να κωδικοποιήσουμε"] + table [, "2η τιμή που θέλουμε να κωδικοποιήσουμε"], table [, "3η τιμή που θέλουμε να κωδικοποιήσουμε"] + table [, "4η τιμή που θέλουμε να κωδικοποιήσουμε"])
```

Νέα εφαρμογή ελέγχου χ^2

- Χρήση πίνακα συνάφειας μετά την επανακωδικοποίηση των γραμμών ή στηλών του προηγούμενου πίνακα συνάφειας
- Εφαρμογή του ελέγχου χ^2 με την εντολή `chisq.test(ctable_x2)` ή `chisq.test(rtable_x2)`

```
x2_M = chisq.test(ctable_x2)
```

```
x2_M
```

ή

```
x2_M = chisq.test(rtable_x2)
```

```
x2_M
```

Παραβίαση προϋποθέσεων – Μέγεθος δείγματος

- Μεγάλο δείγμα (> 250 περιπτώσεις) → τυχαία επιλογή ενός μέρους του αρχικού με την εντολή

New_sample = dataset[sample(1:nrow(dataset), a, replace=FALSE),]