3η Σειρά Ασκήσεων

Ομάδα:

Άγγελος Τσελές (Α.Μ: 3170160)

Ανδρέας Πολυχρονάκης (Α.Μ: 3170140)

<u>Άσκηση 1</u>

a)

Για να είναι ακριβές το διάστημα εμπιστοσύνης θα πρέπει:

1) Τα δεδομένα να προέρχονται από απλό τυχαίο δείγμα
2) X=29>=15 και n-X=50-29=21>15 όπου X=πλήθος κορωνών και n=πλήθος συνολικών ρίψεων

Άρα τα δεδομένα είναι κατάλληλα για τον υπολογισμό του διαστήματος εμπιστοσύνης.

Για c=95% z=abs(qnorm(0.025))=1.96

Δειγματικό ποσοστό p= X/n = 29/50 = 0.58

Έστω m = z * sqrt(p*(1-p)/n)= 1.96 * sqrt(0.58*0.42/50)=0.1368

max = p+m = 0.58+0.1368 = 0.7168

min = p-m = 0.58-0.1368= 0.4432

Διάστημα Εμπιστοσύνης 95 % = [min,max] = [44.32% , 71.68%]

b)

Για να είναι ακριβής ο έλεγχος σημαντικότητας θα πρέπει:

1) Τα δεδομένα να προέρχονται από απλό τυχαίο δείγμα
2) N*po >=10 <=> 50*0.5 >= 10 ΙΣΧΥΕΙ

   και N*(1-po) >=10 <=> 50*0.5 >= 10 ΙΣΧΥΕΙ

Άρα τα δεδομένα είναι κατάλληλα για τον έλεγχο σημαντικότητας.

<u>Μηδενική Υπόθεση Ho</u> : p = 50%

<u>Εναλλακτική Υπόθεση Ha</u> : p != 50%

Στατιστικό ελέγχου z = (p-p0)/sqrt((po*(1-po)/n)) = (0.58-0.5)/sqrt((0.5*0.5/50))= 1.131

3η Σειρά Ασκήσεων

p value = 2 * Φ(-|z|) = 2 * pnorm(-abs(z))= 0.2578

Παρατηρούμε ότι p value > a=0.05 οπότε δεν μπορούμε να απορρίψουμε την μηδενική υπόθεση σε επίπεδο σημαντικότητας 5%.

c)

Ψάχνω το n για m=0.01

Έχω ότι m = z * sqrt(p*(1-p)/n)

Άρα n = p*(1-p) * (z/m)^2

Για p=0.58 : n = 0.58*0.42 * (1.96/0.01)^2 = 9357,1376 άρα n = 9358 ρίψεις

__Άσκηση 2__

Για Ελλάδα: n = 10000000 , X=1100 , C=95% , m = 0.03%

Για ΗΠΑ: n = 300000000 , X=; , C=95% , m = 0.03 %

Για C=95% : z=abs(qnorm(0.025))=1.96

Με δεδομένα ότι z = 1.96 και m = 0.03% αρκεί ότι:

n >= (z^2) / (4*(m^2)) <=> n >= 1067.11

Άρα απαιτούνται τουλάχιστον 1068 άτομα για την πραγματοποίηση αντίστοιχων δημοσκοπήσεων στις Η.Π.Α. Παρατηρούμε ότι το μέγεθος δείγματος δεν εξαρτάται από τον πληθυσμό της χώρας που διεξάγεται η δημοσκόπηση.

__Άσκηση 3__

Τα δεδομένα είναι κατάλληλα για τους παρακάτω ελέγχους καθώς:

1) Προέρχονται από απλό τυχαίο δείγμα

2) Ο αριθμός των επιτυχιών και των αποτυχιών είναι μεγαλύτεροι από 15,όπως προκύπτει εύκολα παρατηρώντας τον πίνακα

3η Σειρά Ασκήσεων

a)

<u>Μηδενική Υπόθεση Ho</u> : p1 = p2

<u>Εναλλακτική Υπόθεση Ha</u> : p1 != p2

Έστω p1 = ποσοστό ανδρών που καπνίζουν και p2 = ποσοστό γυναικών που καπνίζουν

p1 = X1 / n1 = sum(sex=="M" & smoker=="YES") /sum(sex=="M") = 12/30 = 0.4

p2 = X2 / n2 = sum(sex=="F" & smoker=="YES") /sum(sex=="F") = 14/30 = 0.4666

Έχουμε ότι p = (X1+X2) / (n1+n2) = 26/60 = 0.433333

Στατιστικό z = (p1-p2) / sqrt(p*(1-p) / (1/n1 + 1/n2)) = -0.52105

Επιπλέον έχουμε ότι p value = 2 * Φ(-|z|) = 2 * pnorm(-abs(z)) = 0.60233

Παρατηρούμε ότι το p value είναι πολύ μεγάλο οπότε δεν απορρίπτουμε την μηδενική υπόθεση.

b)

Για C=95% διάστημα εμπιστοσύνης έχουμε ότι z=abs(qnorm(0.025))=1.96

Έχουμε ότι m = z * sqrt(p1*(1-p1)/n1 + p2*(1-p2)/n2) = 0.25

min = p2-p1-m = 0.4666-0.4-0.25 = -0.1834

max = p2-p1+m = 0.4666-0.4+0.25 = 0.3166

Άρα Διάστημα Εμπιστοσύνης 95% = [-0.1834,0.3166]

c)

<u>Μηδενική Υπόθεση Ho</u>: Τα ποσοστά του αρχικού δείγματος ταυτίζονται με του πίνακα Eij

<u>Εναλλακτική Υπόθεση Ha</u>: Τα ποσοστά του αρχικού δείγματος δεν ταυτίζονται με του πίνακα Eij

Από τον z έλεγχο στο αρχικό δείγμα (0ij) έχουμε:

p1 = ποσοστό ανδρών που καπνίζουν και p2 = ποσοστό γυναικών που καπνίζουν

p1 = 12 / 30 = 0.4

p2 = 14 / 30 = 0.4666

3η Σειρά Ασκήσεων

```
> table(smoker,sex)
      sex
smoker  F  M
   NO  16 18
   YES 14 12
```

Εικόνα 1: Πίνακας Oij

Για το χ^2 έλεγχο (Eij) προκύπτουν τα εξής:

Eij= Oi*Oj / n

p1' = (30 * (12+14) / 60)/30 = 13/30 = 0.43333

p2' = (30 * (12+14) / 60)/30 = 13/30 = 0.43333

Εκτελούμε chi-squared test

```
> chisq.test(table(smoker,sex),correct=FALSE)

        Pearson's Chi-squared test

data:  table(smoker, sex)
X-squared = 0.27149, df = 1, p-value = 0.6023
```

Άρα προκύπτει ότι X-squared=0.27149,df=1 και p value = 0.6023

Παρατηρούμε ότι το p value είναι πολύ μεγάλο οπότε δεν απορρίπτουμε την μηδενική υπόθεση

d)

Προκύπτει ότι p value z ελέγχου = 0.60233 και p value χ^2 ελέγχου = 0.6023

Τα δύο p value έχουν σχεδόν ακριβώς την ίδια τιμή διότι ο πίνακας των δεδομένων είναι 2x2 και συνεπώς οι βαθμοί ελευθερίας του χ^2 ελέγχου είναι df = (2-1)(2-1) = 1, δηλαδή ίδιοι με την κανονική κατανομή.

3η Σειρά Ασκήσεων

<u>Άσκηση 4</u>

Τα δεδομένα είναι κατάλληλα για τους παρακάτω ελέγχους διότι:

1) Προέρχονται από ένα απλό τυχαίο δείγμα(συσκευασία περιπτέρου)

2) Ο αριθμός επιτυχιών και αποτυχιών είναι μεγαλύτερος από 15

Από τα δεδομένα έχουμε:

X_Brown = 22

X_Red = 19

X_Yellow = 16

X_Blue = 15

X_Green = 8

Sum = 80

Άρα τα ποσοστά εμφάνισης κάθε χρώματος είναι:

p_brown = 22/80 = 0.275

p_red = 19/80 = 0.2375

p_yellow = 16/80 = 0.20

p_blue = 15/80 = 0.1875

p_green = 8/80 = 0.10

a)

Τα δεδομένα είναι κατάλληλα για τους παρακάτω ελέγχους διότι:

1) Προέρχονται από ένα απλό τυχαίο δείγμα(συσκευασία περιπτέρου)

2) n*po >= 10 και n*(1-po) >= 10

<u>Μηδενική Υπόθεση Ho</u>: p_red = p_blue

<u>Εναλλακτική Υπόθεση Hα</u>: p_red > p_blue

3η Σειρά Ασκήσεων

Κάνουμε z έλεγχο:

z = (p_red - p_blue)/sqrt(p_blue*(1-p_blue)/n)

p value = 1- Φ(z) = 0.1259423

Το p value είναι σχετικά μεγάλο οπότε η μηδενική υπόθεση δεν θα απορριφθεί.

b)

p_old_brown = 0.198 => X_old_Brown = 0.198 * 80 = 15.84

p_old_red = 0.178 => X_old_red = 0.178 * 80 = 14.24

p_old_yellow = 0.176 => X_old_yellow = 0.176 * 80 = 14.08

p_old_blue = 0.196 => X_old_blue = 0.196 * 80 = 15.68

p_old_green = 0.252 => X_old_green = 0.252 * 80 = 20.16

<u>Μηδενική Υπόθεση Ho</u>: Δεν έχει αλλάξει η κατανομή

<u>Εναλλακτική Υπόθεση Ha</u>: Έχει αλλάξει η κατανομή

Ισχύει ότι x^2 = (Oij - Eij)^2 / Eij

x1=(X_Brown - X_old_Brown)^2 / X_old_Brown = (22-15.84)^2 / 15.84 = 2.3955

x2=(X_Red - X_old_Red)^2 / X_old_Red = (19-14.24)^2 / 14.24 = 1.5911

x3=(X_Yellow - X_old_Yellow)^2 / X_old_Yellow = (16-14.08)^2 / 14.08 = 0.2618

x4=(X_Blue - X_old_Blue)^2 / X_old_Blue = (15-15.68)^2 / 15.68 = 0.0294

x5=(X_Green - X_old_Green)^2 / X_old_Green = (8-20.16)^2 / 20.16 = 7.334

Άρα χ^2 = x1 + x2 + x3 + x4 + x5 = 11.6118

df = 5-1 =4

Άρα p value = 0.02 ,οπότε απορρίπτουμε την μηδενική υπόθεση. Η κατανομή έχει αλλάξει

3η Σειρά Ασκήσεων

c)

Μηδενική Υπόθεση Ho: Η αναλογία χρωμάτων στα smarties είναι ίδια με αυτή στα M&Ms

Εναλλακτική Υπόθεση Ha: Η αναλογία χρωμάτων στα smarties δεν είναι ίδια με αυτή στα M&Ms

Από τα δεδομένα έχουμε:

X_MM_Brown = 10

X_MM_Red = 12

X_MM_Yellow = 20

X_MM_Blue = 9

X_MM_Green = 5

Sum = 56

Εκτελούμε chi-squared test με t1 τον πίνακα για τα M&Ms και t2 τον πίνακα για τα smarties.

```
> t1

  Blue  Brown  Green    Red Yellow
     9     10      5     12     20
> t2

  Blue  Brown  Green    Red Yellow
    15     22      8     19     16
> prop2 <- prop.table(t1)
> chisq.test(t2, p = c(propt2[1], propt2[2], propt2[3], propt2[4], propt2[5]))

        Chi-squared test for given probabilities

data:  t2
X-squared = 10.358, df = 4, p-value = 0.03481
```

Το p value = 0.03481 είναι πολύ μικρό, οπότε μπορούμε να απορρίψουμε την μηδενική υπόθεση