Εκφώνηση:

Παρασκευάζονται περισσότερα κόκκινα smarties (χρωματιστά σοκολατένια κουφετάκια) από ότι μπλε; Αγοράζετε μια συσκευασία από το περίπτερο όπου βρίσκετε 22 καφέ, 19 κόκκινα, 16 κίτρινα, 15 μπλε και 8 πράσινα κουφέτα.

a. Απαντήστε στο παραπάνω ερώτημα εφαρμόζοντας ένα έλεγχο σημαντικότητας.
b. Το 2009 είχε μετρηθεί με μεγάλη ακρίβεια το ποσοστό εμφάνισης των χρωμάτων καφέ, κόκκινο, κίτρινο, μπλε και πράσινο, το οποίο βρέθηκε ότι ήταν 19.8%, 17.8%, 17.6%, 19.6%, 25.2% αντίστοιχα. Έχει αλλάξει η κατανομή αυτή από τότε;
c. Η αναλογία χρωμάτων στα smarties είναι ίδια με αυτή στα M&Ms (άλλο προϊόν χρωματιστών σοκολατένιων κουφέτων); Ανοίγοντας μια συσκευασία M&Ms βρίσκετε 10 καφέ, 12 κόκκινα, 20 κίτρινα, 9 μπλε και 5 πράσινα

A)

Θεωρώντας πως η συσκευασία από το περίπτερο αποτελεί ένα απλό τυχαίο δείγμα δηλαδή πως τα χρώματα αναμιγνύονται με τυχαίο τρόπο κατά τη συσκευασία τους κάνουμε την υπόθεση, πως η συσκευασία περιέχει κόκκινα με p πιθανότητα και μπλε με p-1 αφού μας ενδιαφέρουν μόνο αυτά τα 2 χρώματα σε αυτή την περίπτωση.

Άρα για H0 p = 1/2

Και για Ha p > 1/2

Από την εκφώνηση έχουμε n = 19+15 = 34

και $\mathrm{p} = 19/34 = 0.5588$

Ο έλεγχος σημαντικότητας $z = \dfrac{p-p'}{\sqrt{\dfrac{(p' * (1-p')}{n}}}$ δίνει σαν αποτέλεσμα το z= 0.686

Στην R αυτό μας δίνει ως αποτέλεσμα του pvalue παρακάτω:

```
> 2*pnorm(-0.686)
[1] 0.4927131
```

Έτσι συμπεραίνουμε πως η διαφορά στον αριθμό των κόκκινων και μπλέ smarties δεν είναι σημαντική.

B)

Από τα δεδομένα προκύπτει πως αν ίσχυε η αναλογία του 2009 στη συσκευασία ΄θα έπρεπε να υπάρχουν:

Χρώμα	Δεδομένα	Αναμενόμενες τιμές
Καφέ	22	80*0.198 = 15.84
Κόκκινο	19	80*0.178 = 14.24
Κίτρινο	16	80*0.176 = 14.08
Μπλέ	15	80*0.196 = 15.68
Πράσινο	8	80*0.252 = 20.16

Έτσι καταλήγουμε να κάνουμε 2 υποθέσεις έστω H0 πως η κατανομή είναι ίδια και Ha πως η κατανομή έχει αλλάξει. Για τον έλεγχο των υποθέσεων, την R και την χρήση του chisq.test για την κατανομή του 2009 παίρνουμε:

```
> c (22,19,16,15,8)->x1
> x1
[1] 22 19 16 15  8
> chisq.test(x1, p = c(0.198, 0.178, 0.176,0.196,0.252))

        Chi-squared test for given probabilities

data:  x1
X-squared = 11.613, df = 4, p-value = 0.02048
```

Άρα για p value 0.02 θεωρούμε πως η υπόθεση H0, μάλλον δεν ισχύει και καταλήγουμε πως η κατανομή μάλλον έχει αλλάξει.

Γ)

Θεωρώντας και πάλι πως η συσκευασία των smarties είναι ένα τυχαίο δείγμα θα εκτελέσουμε έλεγχο ομοιογένειας με την χρήση της R και της chisq.test για να καταλήξουμε ποια από τις παρακάτω υποθέσεις είναι αληθής.

Έστω H0: οι 2 κατανομές των χρωμάτων είναι ομοιογενείς

και Ha: οι 2 κατανομές των χρωμάτων είναι ετερογενείς

```
> matrix=matrix(c(22,10,19,12,16,20,15,9,8,5),nrow=2)
> prop.table(matrix,2)
       [,1]      [,2]      [,3]  [,4]      [,5]
[1,] 0.6875 0.6129032 0.4444444 0.625 0.6153846
[2,] 0.3125 0.3870968 0.5555556 0.375 0.3846154
> addmargins(matrix)
                    Sum
    22 19 16 15  8  80
    10 12 20  9  5  56
Sum 32 31 36 24 13 136
> chisq.test(matrix)

        Pearson's Chi-squared test

data:  matrix
X-squared = 4.6262, df = 4, p-value = 0.3278
```

Βάση του αποτελέσματος

p value=0.3278 και x^2=4.6262 καταλήγουμε πως το pvalue είναι αρκετά μεγάλο για να απορρίψουμε την H0, άρα θεωρούμε πως οι κατανομές είναι μάλλον ομοιογενείς.