Εργασία 3η

ΕΡΓΑΣΙΑ 1

α) Εφόσον έχουμε 50 ρίψεις το δείγμα μας είναι απλό τυχαίο δείγμα του συνόλου απείρων ρίψεων. Έχουμε:

X=29 κορώνες από τις n=50 ρίψεις, όπου 29>15 και n-X=21 >15
$\hat{p}$ = X/n = 0.58 και z_* = 1.96 Άρα μέσω του τύπου έχουμε διάστημα εμπιστοσύνης:
[44.3, 71.6]

β) Έχουμε np_0=25 ≥10 και $n(1-p_0)$=25(1-0.5) ≥10 άρα:

Έστω η μηδενική υπόθεση H_0 (το νόμισμα είναι δίκαιο) p = $\frac{1}{2}$. Έστω ακόμα H_a (το νόμισμα είναι άδικο) p > $\frac{1}{2}$. Για διάστημα εμπιστοσύνης 95% έχουμε επίπεδο σημαντικότητας α=5%. Συνεπώς z= $\frac{\hat{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$≈1.13.

Τελικά, p-value= 1-Φ(z)=1-Φ(1.13)=0.12924 >α=0.05. Η μηδενική υπόθεση H_0, ότι το νόμισμα είναι δίκαιο, δεν μπορεί να απορριφθεί.

γ) Θέλουμε περιθώριο λάθους μικρότερο του 1%, άρα:

$z_* \sqrt{\frac{p(1-p)}{n}} \leq m = 0.01 \Leftrightarrow n \geq \frac{z_*^2}{4\times 0.01^2} = 9603.647$ εφόσον $p(1-p) \leq \frac{1}{4}$

Επομένως θέλουμε n ≥ 9604, πάνω από 9604 ρίψεις.

ΕΡΓΑΣΙΑ 2

Το μέγεθος του δείγματος δεν εξαρτάται από το μέγεθος του πληθυσμού, αλλά μόνο από το περιθώριο λάθους m και το επίπεδο εμπιστοσύνης. Αφού για την κατασκευή 95% διαστημάτων εμπιστοσύνης με περιθώριο σφάλματος 3% προκύπτει ο τύπος:

$z_* \sqrt{\frac{p(1-p)}{n}} \leq m = 0.03 \Leftrightarrow n \geq z_*^2 \frac{p(1-p)}{m^2}$

Αρκεί να ισχύει ο ακολουθος τύπος $n \geq \frac{z_*^2}{4m^2}$ καθώς (p-1)p $\leq \frac{1}{4}$.

ΙΣΧΎΕΙ ότι: Z_* =1.96 και m = 0.03 ⇒ n≥ 3.8416/0.0036 = 1067,11111 ≤1068

Συμπερασματικά 1100 άτομα αρκούν για τις δημοσκοπήσεις για τον πληθυσμό στις Η.Π.Α.. αφού 1068<1100 όπου 1068 τα άτομα που χρειαζόμαστε τουλάχιστον για τις αντίστοιχες δημοσκοπήσεις στις ΗΠΑ.

ΕΡΓΑΣΙΑ 3

α)

Στη συγκεκριμένη περίπτωση έχουμε ένα απλό τυχαίο δείγμα και τα δεδομένα μας είναι ανεξάρτητα οπότε θα είναι ακριβής ο έλεγχος σημαντικότητας.

Έχουμε p_1 (ποσοστό καπνιστών στον υποπληθυσμό ανδρών)

και p_2 (ποσοστό καπνιστών στον υποπληθυσμό των γυναικών)

Θα θεωρήσουμε τον έλεγχο H_0: $p_1 = p_2$

Με βάση τον πίνακα έχουμε :

n_1= 30 X_1 = 12

n_2= 30 X_2 = 14

```
> prop.test(x=c(14,12),n=c(30,30),alternative="less")

        2-sample test for equality of proportions with continuity correction

data:  c(14, 12) out of c(30, 30)
X-squared = 0.067873, df = 1, p-value = 0.6028
alternative hypothesis: less
95 percent confidence interval:
 -1.000000  0.309977
sample estimates:
   prop 1    prop 2
0.4666667 0.4000000
```

Τα δεδομένα μπορούν να θεωρηθούν ως δύο απλά τυχαία δείγματα από κάθε υποπληθυσμό (άνδρες και γυναίκες)
Το p value που υπολογίζει ο z έλεγχος είναι ακριβές αφού $X_1 = 12 \geq 5$

$n_1 - X_1$ =30-12=18≥ 5 , $X_2 = 14 \geq 5, n_2 - X_2$ =30-14=16≥ 5

Έχουμε

$$\hat{p} = \frac{X_1 + X_2}{n_1 + n_2} = \frac{12+14}{30+30} = \frac{26}{60} \approx 0.4333$$

$$\hat{p_1} = \frac{12}{30} = 0.4 \text{ (male)}$$

$$\hat{p_2} = \frac{14}{30} \approx 0.4666 \text{ (female)}$$

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p_1}(1-\hat{p_1})}{n_1} + \frac{\hat{p_2}(1-\hat{p_2})}{n_2}}} = \frac{0.4 - 0.4666}{\sqrt{\frac{0.4(1-0.4)}{30} + \frac{0.4666(1-0.4666)}{30}}} = \frac{-0{,}0666}{\sqrt{\frac{0{,}4 * 0{,}6}{30} + \frac{o{,}24888}{30}}} = \frac{0{,}666}{\sqrt{0{,}01629}} = \frac{0{,}0666}{0127632284} = -0.52181155$$

Έχουμε p – value≈ 0.6 όπως διαπιστώσαμε.

Για διάστημα εμπιστοσύνης 95% έχουμε ότι το επίπεδο σημαντικότητας α = 5%.
Άρα, p-value > α δηλαδή δεν μπορούμε να απορρίψουμε τη μηδενική υπόθεση,
δηλαδή ότι οι γυναίκες καπνίζουν στην ίδια συχνότητα με τους άντρες.
Αυτό σημαίνει ότι μπορούμε να υποθέσουμε ότι το κάπνισμα δεν έχει σχέση με το φύλο.

β)Το επίπεδο εμπιστοσύνης του διαστήματος που θα βρούμε είναι ακριβές (95%) αφού ισχύει ότι $X_1 = 10$ $n_1 - X_1 = 30-12=18 \geq 10$, $X_2 = 14 \geq 10, n_2 - X_2 = 30-14=16 \geq 10$

$$\widehat{p}_1 - \widehat{p}_2 \pm Z_* \sqrt{\frac{\widehat{p}_1(1-\widehat{p}_1)}{n_1} + \frac{\widehat{p}_2(1-\widehat{p}_2)}{n_2}}$$

$$= 0,4 - 0,4666 \pm 0,5218 * 0,127632284 = -0,0666 \pm 0,66599 = [-0,73259, 0,59939]$$

Υπενθυμίζουμε ότι $z_* = 1.96$ Διαπιστώνουμε πως το ποσοστό των γυναικών καπνιστριων είναι μεγαλύτερο από το αντίστοιχο των ανδρών.Υποδεικνύεται μια ενδεχόμενη σχέση μεταξύ φύλου και καπνίσματος που σε ένα μεγαλύτερο δείγμα θα την επιβεβαιώναμε ή θα την διαψεύδαμε καθώς σε αυτό το πλήθος η διαφορά δεν είναι σημαντική.

c+d)Για να είναι ακριβής ο έλεγχος σημαντικότητας θα πρέπει τα δεδομένα μας να προέρχονται από απλό τυχαίο δείγμα (SRS). Στη συγκεκριμένη περίπτωση έχουμε ένα απλό τυχαίο δείγμα και τα δεδομένα μας είναι ανεξάρτητα. Οπότε μπορούμε να προχωρήσουμε σε έλεγχο σημαντικότητας. Έστω οι κατηγορικές μεταβλητές Φ (για το φύλο) και Κ (για το κάπνισμα). Έστω η μηδενική υπόθεση : Ηο : το φύλο δεν σχετίζεται με το κάπνισμα και η εναλλακτική υπόθεση H_α:το φύλο σχετίζεται με το κάπνισμα

Ο πίνακας συνάφειας για τα δεδομένα του Πίνακα 1 είναι:

	Smokers	Non-Smokers	
Males	12	18	30
Females	14	16	30
	26	34	60

$$\text{Αναμενόμενο Πλήθος} = \frac{\text{Άθροισμα Γραμμής*Άθροισμα Στήλης}}{\text{Συνολικό Μέγεθος Δείγματος}}$$

Αναμενομενος πίνακας:

	Smokers	Non-Smokers	
Males	13	17	30
Females	13	17	30
	26	34	60

Η τιμή του στατιστικού ελέγχου είναι $x^2 = \frac{(12-13)^2}{13} + \frac{(18-17)^2}{17} + \frac{(14-13)^2}{13} + \frac{(16-17)^2}{13} \approx 0.271$

Άρα έχουμε df = (rows − 1) × (columns − 1) = 1 .

και από την r έχουμε :

```
> y<-c(12,14,18,16)
> n<-sum(y)
> n
[1] 60
> p<-c(13,13,17,17)
> p<-p/sum(p)
> cc<-sum((y-n*p)^2/(n*p))
> cc
[1] 0.2714932
> pchisq(cc,df=1,lower.tail=FALSE)
[1] 0.6023319
```

Άρα p-value = 0.602. Για διάστημα εμπιστοσύνης 95% έχουμε α = 0.05 και άρα p-value = 0.602 > 0.05=α οπότε δε μπορούμε να απορρίψουμε την μηδενική υπόθεση. Άρα οι μεταβλητές φύλο και κάπνισμα είναι ανεξάρτητες, δηλαδή το φύλο δε σχετίζεται με το κάπνισμα.Παρατηρούμε ότι η τιμή αυτή είναι ίση με την τιμή του p-value που υπολογίσαμε στο ερώτημα α. Άρα σε κάθε περίπτωση δεν μπορούμε να απορρίψουμε τη μηδενική υπόθεση, δηλαδή αυτό σημαίνει ότι μπορούμε να υποθέσουμε ότι το κάπνισμα δεν έχει σχέση με το φύλο.

ΕΡΓΑΣΙΑ 4

α) Έχουμε απλό τυχαίο δείγμα με ανεξάρτητα δεδομένα.

Έστω μηδενική υπόθεση H_0, παρασκευή τόσων κόκκινων smarties όσων και μπλέ, $p_K = p_M$

Έστω επίσης εναλλακτική υπόθεση H_a, παρασκευή περισσότερων κόκκινων smarties από μπλε, $p_K > p_M$. Όπως βλέπουμε p-value = 0.3035

```
> prop.test(x=19,n=34,alternative="greater")

        1-sample proportions test with continuity correction

data:  19 out of 34, null probability 0.5
X-squared = 0.26471, df = 1, p-value = 0.3035
alternative hypothesis: true p is greater than 0.5
95 percent confidence interval:
 0.4056088 1.0000000
sample estimates:
        p
0.5588235
```

Έχουμε α=5%, και p-value = 0.3035 >α=0.05. Η μηδενική υπόθεση H_0 δεν μπορεί να απορριφθεί.

β) Έχουμε απλό τυχαίο δείγμα με ανεξάρτητα δεδομένα.

Έστω μηδενική υπόθεση H_0, η κατανομή των χρωμάτων καφέ, κόκκινο, κίτρινο, μπλε και πράσινο είναι 19.8%, 17.8%, 17.6%, 19.6%, 25.2% αντίστοιχα.

Έστω επίσης εναλλακτική υπόθεση H_a, η κατανομή χρωμάτων να μην είναι αυτή της μηδενικής υπόθεσης.

```
> q<- c(22,19,16,15,8)
> p<- c(0.198,0.178,0.176,0.196,0.252)
> chisq.test(q,p=p)

        Chi-squared test for given probabilities

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

Έχουμε α=5%, και p-value = 0.02048 <α=0.05. Η μηδενική υπόθεση H_0 απορρίπτεται, τελικά η κατανομή χρωμάτων έχει αλλάξει από το 2009.

γ) Έχουμε απλό τυχαίο δείγμα με ανεξάρτητα δεδομένα.

Έστω μηδενική υπόθεση H_0, Smarties και M&Ms έχουν ίδια κατανομή χρωμάτων.
Έστω επίσης εναλλακτική υπόθεση H_a, οι κατανομές χρωμάτων Smarties και M&Ms να διαφέρουν.
Έχουμε:

	Καφέ	Κόκκινα	Κίτρινα	Μπλε	Πράσινα	Total
Smarties	22	19	16	15	8	80
M&Ms	10	12	20	9	5	56
Total	32	31	36	24	13	136

```
> smarties<-c(22,19,16,15,8)
> MMs<-c(10,12,20,9,5)
> x<-rbind(smarties,MMs)
> colnames(x)<-c("brown","red","yellow","blue","green")
> x
         brown red yellow blue green
smarties    22  19     16   15     8
MMs         10  12     20    9     5
> chisq.test(x)

        Pearson's Chi-squared test

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

Έχουμε α=5%, και p-value = 0.3278 > α=0.05. Η μηδενική υπόθεση H_0, Smarties και M&Ms έχουν ίδια κατανομή χρωμάτων, δεν μπορεί να απορριφθεί.