



Οικονομικό Πανεπιστήμιο Αθηνών, Τμήμα Πληροφορικής
Μάθημα: Στατιστική στην Πληροφορική
Ακαδημαϊκό έτος: 2022–23
Υπεύθυνη φοιτήτρια: Μαρία Κονταράτου (3200078)

3^η Σειρά Ασκήσεων

Άσκηση 1:

- (a) Έχουμε $x = 29$ κορώνες και $y = 50 - 29 = 21$ γράμματα. Τα δεδομένα μας αρχικά χαρακτηρίζονται κατάλληλα αφού $x, y > 15$

Άρα έχουμε $p = x/n = 29/50 = 0.58$

$$\sigma = \sqrt{p(1-p)/n} = \sqrt{0.58 \cdot 0.42/50} = 0.069799$$

άρα για 95%, $z = 1.96$ άρα

$$[p - z \cdot \sigma, p + z \cdot \sigma] = [0.4431, 0.7168]$$

- (b) Αφού ασχολούμαστε με δίκαιο ζάρι, η συχνότητα εμφάνισης κορώνας θα είναι 0.5
Άρα δίπλευρος έλεγχος $H_0: p=0.5$

$$\text{Στατικός έλεγχος } z = (0.58 - p) / \sqrt{(p \cdot (1-p)/50)} = 1.131370$$

Άρα $p \text{ value} = 2\Phi(-|z|) = 0.25$, που είναι αρκετά μικρή τιμή για να απορριφθεί η μηδενική υπόθεση

- (c) Πρέπει $N \geq z^2 / 4m^2 = 1.96^2 / 4 \cdot 0.01^2 = 9604$ ρίψεις

Άσκηση 2:

Το δείγμα μας είναι ανεξάρτητο από το μέγεθος του πληθυσμού.

Αυτό ισχύει λόγω του τύπου $N \geq z^2 / 4m^2$ αφού εξαρτάται από το περιθώριο λάθους και το επίπεδο εμπιστοσύνης. Άρα ο αριθμός 1100 αρκεί για τις δημοσκοπήσεις

Άσκηση 3:

- (a) Έστω p_1 : ποσοστό καπνιστών που είναι άνδρες και $n_1 = 30$ άντρες
και p_2 : ποσοστό καπνιστών που είναι γυναίκες και $n_2 = 30$ γυναίκες
Δίπλευρος έλεγχος $H_0: p_1 = p_2$

$$\bar{p}_1 = 12/30 = 0.4 \text{ ποσοστό αντρών που καπνίζουν}$$

$$\bar{p}_2 = 14/30 = 0.46 \text{ ποσοστό γυναικών που καπνίζουν}$$

$$\bar{p} = (12+14)/60 = 0.43$$

Στατιστικός έλεγχος $z = -0.521$ άρα p value: 0.6 περίπου άρα δεν μπορούμε να απορρίψουμε την αρχική υπόθεση οπότε υπάρχει σχέση μεταξύ φύλου και καπνίσματος

(b) Το διάστημα εμπιστοσύνης είναι σύμφωνα με τον τύπο

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

Και το διάστημα τελικά είναι $[-0,194, 0,061]$

(c) Έστω πραγματοποιούμε έλεγχο

H_0 : Το φύλο δεν έχει σχέση με το κάπνισμα

H_a : Το φύλο έχει σχέση με το κάπνισμα

Άρα έχουμε τον παρακάτω πίνακα συνάφειας

ΦΥΛΟ	ΚΑΠΝΙΣΤΗΣ	ΜΗ ΚΑΠΝΙΣΤΗΣ	
ΑΝΤΡΕΣ	12	18	30
ΓΥΝΑΙΚΕΣ	14	16	30
	26	34	60

(d)

ΦΥΛΟ	ΚΑΠΝΙΣΤΗΣ	ΜΗ ΚΑΠΝΙΣΤΗΣ	
ΑΝΤΡΕΣ	13	17	30
ΓΥΝΑΙΚΕΣ	13	17	30
	26	34	60

Στατιστικός έλεγχος $\chi^2 = (12-13)^2/13 + (18-17)^2/17 + (14-13)^2/13 + (16-17)^2/17 \cong 0.27149$

P value = 0.60233 (από πίνακα χ^2)

Το p value είναι ίδιο με εκείνο του υποερωτήματος (α) και συνεπώς οι έλεγχοι είναι ισοδύναμοι.

Άσκηση 4:

(a) Έστω p : ποσοστό κόκκινων smarties και
 $p-1$: ποσοστό μπλε smarties

Θέλω να δω αν τα κόκκινα smarties είναι περισσότερα από τα μπλε άρα κάνουμε έλεγχο υπόθεσης $H_0: p = 0.5$

$H_a: p > 0.5$

Από υπόθεση, $N = 19 + 15 = 34$

$X = 19$

$$\bar{p} = X/N = 19/34 = 0.5588$$

Έστω ότι τα smarties αποτελούν τυχαίο δείγμα μιας συσκευασίας
Τότε θα έχουμε στατιστικό έλεγχο $z = 0.686$ μέσω του κάτω τύπου

$$z = \frac{\bar{p} - 0.5}{\sqrt{\frac{0.5(1-0.5)}{N}}}$$

Και $p \text{ value} = 2\Phi(-|z|) = 0.4902$

Όμως η τιμή του $p \text{ value}$ είναι μικρή άρα η μηδενική υπόθεση αποδεκτή.
Συνεπώς δεν υπάρχει μεγάλη διαφορά στον αριθμό κόκκινων και μπλε smarties.

(b) Εφαρμόζουμε χ^2 έλεγχο προσαρμογής

H_0 : κατανομή των καφέ, κόκκινο, κίτρινο, πράσινο και μπλε (κατανομές 19.8%, 17.8%, 19.6%, 25.2%, 19.6%)

H_a : κατανομή χρωμάτων είναι άλλη

Συνεπώς προκύπτει ο πίνακας:

ΧΡΩΜΑ	ΔΕΔΟΜΕΝΑ	ΤΙΜΕΣ
ΚΑΦΕ	22	15.84
ΚΟΚΚΙΝΟ	19	14.24
ΚΙΤΡΙΝΟ	16	14.08
ΠΡΑΣΙΝΟ	8	20.16
ΜΠΛΕ	15	15.68

Και υπολογίζουμε την $p \text{ value} = 0.02$, που χαρακτηρίζεται μικρή τιμή άρα δεν ισχύει η μηδενική υπόθεση. Συνεπώς η κατανομή είναι διαφορετική συγκριτικά με το 2009

(c) Έστω ότι η συσκευασία των M&Ms είναι τυχαίο δείγμα με σοκολατάκια.

Αν τα δείγματα των 2 τύπων smarties και M&Ms δημιουργήθηκαν από πληθυσμούς με την ίδια κατανομή χρωμάτων.

Χρησιμοποιώ χ^2 έλεγχο για ομοιογένεια.

H_0 : Οι πληθυσμοί είναι ομοιογενείς

H_a : Οι πληθυσμοί δεν είναι ομοιογενείς

Συνεπώς προκύπτει ο πίνακας:

ΧΡΩΜΑ	SMARTIES	M&Ms	
ΚΑΦΕ	22	10	32
ΚΟΚΚΙΝΟ	19	12	31
ΚΙΤΡΙΝΟ	16	20	36
ΜΠΛΕ	15	9	24
ΠΡΑΣΙΝΟ	8	5	13

	80	56	136
--	----	----	-----

Υπολογίζουμε μέσω $\chi^2 = 4.626$, βαθμοί ελευθερίας = 4 και p value: 0.3278
 Το p value είναι μεγάλο για να μην απορριφθεί η μηδενική υπόθεση άρα οι δύο κατανομές είναι ίδιες.