
M.Sc. in Data Science

Course: Probability and Statistics for Data Analysis

Instructor: Ioannis Vrontos (vrontos@aueb.gr)

Grader: Constandina Koki (kokiconst@aueb.gr)

Assignment 2 Deadline: 24 December 2017

Άσκηση 1

Έστω τυχαίο δείγμα X_1, X_2, \dots, X_n (ανεξάρτητες τυχαίες μεταβλητές) από διωνυμική κατανομή

$\text{Bin}(2, \theta)$, με συνάρτηση πιθανότητας $f(x) = \binom{2}{x} \theta^x \cdot (1-\theta)^{2-x}$, $0 < \theta < 1$.

- 1) Να βρείτε την εκτιμήτρια μεγίστης πιθανοφάνειας της παραμέτρου θ .
- 2) Να δείξετε αν ο εκτιμητής μεγίστης πιθανοφάνειας του θ είναι αμερόληπτος εκτιμητής.
- 3) Να βρείτε έναν εκτιμητή για την $P(X=2)$.

Άσκηση 2

Έστω τυχαίο δείγμα X_1, X_2, \dots, X_n (ανεξάρτητες τυχαίες μεταβλητές) από κανονική κατανομή $N(\mu, \theta)$, όπου ο μέσος μ είναι γνωστός και η διακύμανση θ είναι άγνωστη.

- 1) Να βρείτε την εκτιμήτρια μεγίστης πιθανοφάνειας της παραμέτρου θ (της διακύμανσης).
- 2) Να δείξετε αν η εκτιμήτρια μεγίστης πιθανοφάνειας της διακύμανσης θ είναι αμερόληπτη εκτιμήτρια.
- 3) Να βρείτε την διακύμανση της εκτιμήτριας της διακύμανσης, δηλαδή να βρείτε την $V(\hat{\theta})$

[Να χρησιμοποιήσετε ότι

$$x \sim N(\mu, \theta) \Rightarrow \frac{x - \mu}{\sqrt{\theta}} \sim N(0, 1) \Rightarrow \left(\frac{x - \mu}{\sqrt{\theta}} \right)^2 \sim \chi_1^2$$

και στη συνέχεια την διακύμανση της Chi-Square κατανομής]

- 4) Να βρείτε το κάτω φράγμα Cramer-Rao, και να εξετάσετε αν η εκτιμήτρια της διακύμανσης είναι αποτελεσματική.

Άσκηση 3

Έστω τυχαίο δείγμα X_1, X_2, \dots, X_n (ανεξάρτητες τυχαίες μεταβλητές) από κανονική κατανομή $N(\mu, \theta)$, όπου ο μέσος μ και η διακύμανση θ είναι άγνωστες παράμετροι.

- 1) Να βρείτε την εκτιμήτρια μεγίστης πιθανοφάνειας του $\mu^2 + \theta$
- 2) Να βρείτε την εκτιμήτρια μεγίστης πιθανοφάνειας της $P(X \leq x_o)$

Άσκηση 4

Φάρμακο χορηγείται σε ασθενείς που εμφανίζουν υψηλή συστολική πίεση. Οι υπεύθυνοι της φαρμακευτικής εταιρίας που το παράγει ενδιαφέρονται να μελετήσουν την αποτελεσματικότητα του φαρμάκου. Για το σκοπό αυτό επιλέχθηκαν με τυχαίο τρόπο 9 ασθενείς και μετρήθηκε η συστολική πίεση αμέσως πριν και 2 ώρες μετά τη λήψη 25 mg του φαρμάκου. Οι μετρήσεις συνοψίζονται στον Πίνακα 1.

Πίνακας 1: Συστολική πίεση (mm Hg) πριν και μετά την χορήγηση του φαρμάκου

Αύξων αριθμός Ασθενή	Συστολική Πίεση	
	Πριν	Μετά
1	210	201
2	180	175
3	187	184
4	175	170
5	175	168
6	176	169
7	190	183
8	206	199
9	173	168

- α) Να κατασκευάσετε ένα 95% διάστημα εμπιστοσύνης για την μέση συστολική πίεση των ασθενών πριν τη χορήγηση του φαρμάκου.
β) Να κατασκευάσετε ένα 90% διάστημα εμπιστοσύνης για την μέση συστολική πίεση των ασθενών μετά τη χορήγηση του φαρμάκου.
γ) Να κατασκευάσετε ένα 95% διάστημα εμπιστοσύνης για την μέση διαφορά της συστολικής πίεσης των ασθενών πριν και μετά τη χορήγηση του φαρμάκου.
δ) Να διεξάγετε τον ακόλουθο έλεγχο υποθέσεων

$$H_0 : \mu_1 = \mu_2$$

$$H_0 : \mu_1 > \mu_2$$

- όπου μ_1 είναι η μέση συστολική πίεση των ασθενών στον πληθυσμό πριν τη χορήγηση του φαρμάκου, και μ_2 είναι η μέση συστολική πίεση των ασθενών στον πληθυσμό μετά τη χορήγηση του φαρμάκου. Να χρησιμοποιήσετε επίπεδο σημαντικότητας $\alpha=0.05$. Σε ποιο συμπέρασμα καταλήγετε όσον αφορά την αποτελεσματικότητα του φαρμάκου;
ε) Να απαντήσετε τα παραπάνω ερωτήματα με τη χρήση του R. Να δώσετε τον κωδικα R, που γράψατε προκειμένου να απαντήσετε στα ερωτήματα (α)-(ε).

Άσκηση 5

Ο πρόεδρος κάποιου τμήματος οικονομικής σχολής ενδιαφέρεται να εξετάσει την απόδοση των φοιτητών σε δύο από τα βασικά μαθήματα του τμήματος (Εισαγωγή στο Μάρκετινγκ, Λογιστική), καθώς και τον βαθμό ικανοποίησης των φοιτητών από το πρόγραμμα σπουδών του τμήματος. Επίσης, ενδιαφέρεται να μελετήσει αν η απόδοση και ο βαθμός ικανοποίησης των αγοριών διαφέρουν σε σχέση με αυτή των κοριτσιών. Για το λόγο αυτό επέλεξε με τυχαίο τρόπο 30 φοιτητές και φοιτήτριες του τμήματος που εξετάστηκαν κατά το τρέχον έτος και στα δύο αυτά μαθήματα, κατέγραψε την

απόδοση τους και την απάντησή τους στο ερώτημα «Είστε ικανοποιημένοι από το πρόγραμμα σπουδών του τμήματος;». Τα αποτελέσματα παρουσιάζονται στον πίνακα 2:

Πίνακας 2: Φύλο (1:αγόρι, 2: κορίτσι), απόδοση και ικανοποίηση (1: ναι, 0: όχι)

Φύλο	Εισαγωγή στο Μάρκετινγκ	Λογιστική	Ικανοποίηση
1	99	89	1
2	89	92	1
1	32	37	0
1	44	51	1
2	45	74	0
2	25	57	0
2	47	30	1
1	86	55	1
1	97	49	0
1	94	45	1
1	55	68	1
1	61	47	1
2	94	87	1
2	55	34	1
1	66	36	1
2	60	59	0
1	83	94	1
1	67	42	1
1	46	28	1
1	52	39	1
1	29	61	0
1	54	56	1
1	25	49	0
1	90	80	1
2	33	29	1
1	77	55	1
1	79	46	1
2	97	93	1
2	100	92	1
1	25	64	1

Να απαντήσετε στα παρακάτω ερωτήματα αναλυτικά. Λόγω του πλήθους των δεδομένων να χρησιμοποιήσετε στους υπολογισμούς σας το πακέτο R.

α) Να βρεθεί το ποσοστό των φοιτητών/τριών που είναι ικανοποιημένοι από το πρόγραμμα σπουδών για το σύνολο του δείγματος. Να κατασκευάσετε ένα 95% διάστημα εμπιστοσύνης για το ποσοστό p στον πληθυσμό.

β) Να κατασκευάσετε ένα 99% διάστημα εμπιστοσύνης για τη διαφορά των ποσοστών $p_1 - p_2$ στον πληθυσμό, όπου p_1 είναι το ποσοστό των φοιτητών που είναι ικανοποιημένοι από το πρόγραμμα

σπουδών και p_2 είναι το ποσοστό των φοιτητριών που είναι ικανοποιημένες από το πρόγραμμα σπουδών.

γ) Να εξετάσετε αν υπάρχει στατιστικά σημαντική διαφορά στο ποσοστό των φοιτητών και φοιτητριών που είναι ικανοποιημένοι/ες από το πρόγραμμα σπουδών [Να χρησιμοποιήσετε επίπεδο σημαντικότητας $\alpha=0.05$].

δ) Να κατασκευάσετε ένα 95% διάστημα εμπιστοσύνης για τη διαφορά των μέσων βαθμών $\mu_1 - \mu_2$ στον πληθυσμό, όπου μ_1 είναι ο μέσος βαθμός των φοιτητών στο μαρκετινγκ και μ_2 είναι ο μέσος βαθμός των φοιτητριών στο μαρκετινγκ.

ε) Να εξετάσετε αν υπάρχει διαφορά στο μέσο βαθμό μ_1 των φοιτητών στο μαρκετινγκ και στον μέσο βαθμό μ_2 των φοιτητριών στο μαρκετινγκ [Να χρησιμοποιήσετε επίπεδο σημαντικότητας $\alpha=0.05$].

στ) Την προηγούμενη εξεταστική περίοδο το ποσοστό των φοιτητών/τριών που έγραψαν άριστα στη λογιστική ήταν 9.5%. Νομίζετε ότι κατά το τρέχον έτος το ποσοστό αυτό αυξήθηκε; [Να χρησιμοποιήσετε επίπεδο σημαντικότητας $\alpha=0.05$].

ζ) Να απαντήσετε τα παραπάνω ερωτήματα με τη χρήση του R. Να δώσετε τον κωδικά R, που γράψατε προκειμένου να απαντήσετε στα ερωτήματα (α)-(ζ).