



2^η Άσκηση: παράδοση Τρίτη 22/12

1. Μας ενδιαφέρει ο μέσος χρόνος διεκπεραίωσης μιας αίτησης (query) σε μια βάση δεδομένων. Από τις αιτήσεις μιας ημέρας, επιλέχθηκαν 20 τυχαίες και μετρήθηκαν οι ακόλουθοι χρόνοι διεκπεραίωσης σε milliseconds:

82	55	58	94	86	45	42	36	41	130
284	96	39	107	52	54	45	81	83	38

- a. Τα δεδομένα είναι κατάλληλα για τις μεθόδους συμπερασματολογίας που γνωρίζουμε; Εξηγήστε.
- b. Δώστε ένα 95% διάστημα εμπιστοσύνης για τη μέση τιμή του χρόνου διεκπεραίωσης.
2. Παρακάτω δίνονται διάφορες περιπτώσεις λάθος εφαρμογής ελέγχων σημαντικότητας. Εξηγήστε για κάθε περίπτωση ποιο είναι το λάθος και γιατί είναι λάθος.
- a. Λαμβάνεται ένα τυχαίο δείγμα μεγέθους 20 από πληθυσμό με τυπική απόκλιση 12. Η τυπική απόκλιση του δειγματικού μέσου είναι $12/20$.
- b. Ένας ερευνητής χρησιμοποιεί σε έναν έλεγχο σημαντικότητας τη μηδενική υπόθεση $H_0: \bar{x} = 10$.
- c. Σε μια στατιστική έρευνα με $\bar{x} = 45$ απορρίπτεται η μηδενική υπόθεση $H_0: \mu = 54$ όταν η εναλλακτική είναι $H_a: \mu > 54$.
- d. Σε μια στατιστική έρευνα όπου $p \text{ value} = 0.52$ απορρίπτεται η μηδενική υπόθεση.
3. Σε έλεγχο σημαντικότητας με $H_0: \mu = \mu_0$ η τιμή του στατιστικού ελέγχου z είναι 1.34.
- a. Ποιο είναι το $p \text{ value}$ για την εναλλακτική υπόθεση $H_a: \mu > \mu_0$;
- b. Ποιο είναι το $p \text{ value}$ για την εναλλακτική υπόθεση $H_a: \mu < \mu_0$;
- c. Ποιο είναι το $p \text{ value}$ για την εναλλακτική υπόθεση $H_a: \mu \neq \mu_0$;

4. Το p value για ένα δίπλευρο έλεγχο με μηδενική υπόθεση $H_0: \mu = 30$ είναι 0.04.
- a. Η τιμή 30 περιέχεται στο 95% διάστημα εμπιστοσύνης για τη μέση τιμή μ ; Γιατί;
- b. Η τιμή 30 περιέχεται στο 90% διάστημα; Γιατί;
5. Θεωρήστε τα δεδομένα που δίδονται στον Πίνακα 1 παρακάτω, τα οποία προήλθαν από απλή τυχαία δειγματοληψία (simple random sampling – SRS) 25 ενηλίκων κατοίκων Αθήνας.

Πίνακας 1

A/A	ΦΥΛΟ	ΚΑΠΝΙΣΤΗΣ	ΒΑΡΟΣ
1	A	ΝΑΙ	80
2	A	ΌΧΙ	81
3	A	ΌΧΙ	75
4	A	ΝΑΙ	83
5	Γ	ΝΑΙ	71
6	A	ΝΑΙ	73
7	Γ	ΝΑΙ	65
8	Γ	ΌΧΙ	67
9	Γ	ΌΧΙ	54
10	A	ΝΑΙ	77
11	Γ	ΌΧΙ	55
12	Γ	ΌΧΙ	83
13	A	ΌΧΙ	91
14	Γ	ΌΧΙ	6
15	A	ΝΑΙ	92
16	A	ΝΑΙ	86
17	Γ	ΌΧΙ	73
18	Γ	ΝΑΙ	82
19	Γ	ΌΧΙ	69
20	A	ΌΧΙ	73
21	Γ	ΌΧΙ	70
22	Γ	ΝΑΙ	59
23	A	ΌΧΙ	68
24	A	ΌΧΙ	72
25	A	ΌΧΙ	72

Στις απαντήσεις των παρακάτω ερωτημάτων, εξετάστε επίσης την καταλληλότητα των εκάστοτε δεδομένων.

- a. Δώστε ένα 95% διάστημα εμπιστοσύνης για το μέσο βάρος των ενηλίκων κατοίκων Αθήνας.

- b. Δώστε ένα 80% διάστημα εμπιστοσύνης για τη διαφορά του μέσου βάρους μεταξύ ανδρών και γυναικών (ενηλίκους κατοίκους Αθηνών).
- c. Το κάπνισμα έχει σχέση με το βάρος; Διατυπώστε έναν κατάλληλο έλεγχο σημαντικότητας και σχολιάστε τα ευρήματά σας.
6. (Άσκηση 7.24 από “Introduction to the Practice of Statistics”, Moore, McCabe, Craig, WH Freeman, 6th edition (IPS)) Αυτοκίνητα που έχουν κατασκευαστεί στις τελευταίες δεκαετίες είναι εξοπλισμένα με υπολογιστές που εκτιμούν την απόδοση των καυσίμων σε λίτρα καυσίμου που καταναλώθηκαν ανά 100 χιλιόμετρα απόστασης. Σε ένα τέτοιο αυτοκίνητο, σημειώνονταν η ένδειξη του υπολογιστή απόδοσης κάθε φορά που γίνονταν ανεφοδιασμός και ακολούθως μηδενίζονταν ο μετρητής (ώστε η ένδειξη να αφορά τη διαδρομή μεταξύ διαδοχικών ανεφοδιασμών). Δίνεται ένα τυχαίο δείγμα 20 τέτοιων ενδείξεων:

5.7	4.6	6.4	6.3	6.9	5.2	4.9	5.4	4.9	5.6
5.4	5.3	4.9	5.1	5.0	6.0	6.3	5.4	5.3	5.4

- a. Τα δεδομένα είναι κατάλληλα για τις μεθόδους συμπερασματολογίας που γνωρίζουμε; Εξηγήστε.
- b. Βρείτε τη μέση τιμή και τυπική απόκλιση για τα δεδομένα αυτά.
- c. Εκτιμήστε τη μέση τιμή μ της απόδοσης του αυτοκινήτου, με ένα 95% διάστημα εμπιστοσύνης χρησιμοποιώντας τα παραπάνω δεδομένα.
7. (Άσκηση 7.34 από IPS) Μια ασφαλιστική εταιρία αυτοκινήτων συνεργάζεται με ένα συνεργείο αυτοκινήτων για εκτίμηση ζημιών όπου όμως υπάρχει η ανησυχία ότι το συνεργείο υπερεκτιμά τις ζημιές. Για να δει η εταιρία αν όντως συμβαίνει αυτό, έλαβε τις εκτιμήσεις ζημιών για 10 αυτοκίνητα από το εν λόγω συνεργείο και τις συνέκρινε με τις εκτιμήσεις ενός άλλου εμπειρογνώμονα. Ακολουθούν τα αποτελέσματα:

Αυτοκίνητο	1	2	3	4	5	6	7	8	9	10
Συνεργείο	500	1550	1250	1300	750	1000	1250	1300	800	2500
Εμπειρογνώμονας	400	1500	1300	1300	800	800	1000	1100	650	2200

Ελέγξτε τη μηδενική υπόθεση ότι δεν υπάρχει υπερεκτίμηση ζημιών από το συνεργείο. Τι συμπέρασμα βγάζετε;

8. Χρησιμοποιήστε τα δεδομένα με τις [απαντήσεις του ερωτηματολογίου 2020](#) για να διατυπώσετε ελέγχους σημαντικότητας για τα ακόλουθα ερωτήματα:

- a. Βρείτε ένα 95% διάστημα εμπιστοσύνης για τη διαφορά του μέσου ύψους μεταξύ ανδρών και γυναικών φοιτητών πληροφορικής του ΟΠΑ.
- b. Οι άνδρες φοιτητές πληροφορικής -που έχουν πάρει ή θα έπαιρναν το μάθημα «Στατιστική στην Πληροφορική»-, επιτυγχάνουν μεγαλύτερο μέσο βαθμό στο μάθημα των Πιθανοτήτων από τον αντίστοιχο πληθυσμό γυναικών; Απαντήστε σε επίπεδο σημαντικότητας 5%.
- c. Ο μέσος βαθμός στα Μαθηματικά 1 διαφέρει από το μέσο βαθμό στις Πιθανότητες -μεταξύ των φοιτητών που έχουν πάρει ή θα έπαιρναν το μάθημα «Στατιστική στην Πληροφορική»-;

Σχόλια/οδηγίες:

1. *Πριν την εξαγωγή συμπερασμάτων (διαστήματα εμπιστοσύνης ή ελέγχους σημαντικότητας), είναι πολύ σημαντικό να κάνετε διερευνητική ανάλυση δεδομένων για να διαπιστώνετε αν τα δεδομένα είναι κατάλληλα για την ακρίβεια των εκάστοτε μεθόδων, σύμφωνα με τα κριτήρια που έχουν ειπωθεί στις διαλέξεις.*
2. *Μπορείτε να χρησιμοποιήσετε όποιο λογισμικό στατιστικών υπολογισμών θέλετε.*
3. *Μπορείτε να κάνετε τις ασκήσεις ομαδικά σε ομάδες έως 2 το πολύ ατόμων. Εργασίες ομάδων με δύο μέλη θα παραδίνονται μόνο από το ένα μέλος (μην ξεχάσετε να αναφέρετε τα ονόματα/ΑΜ όλων των μελών!)*