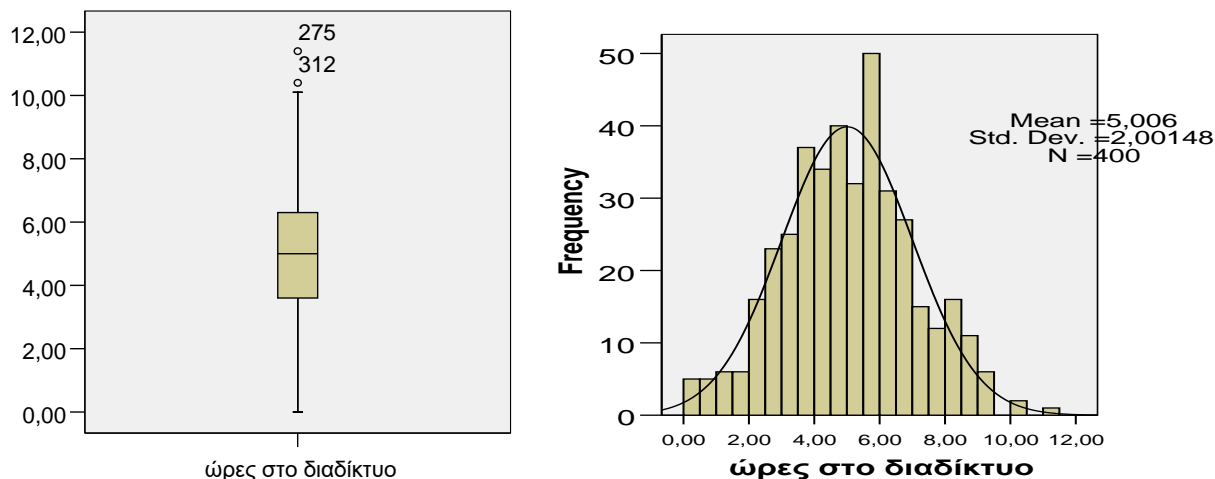


## Επαναληπτικές Ασκήσεις 2018

### Άσκηση 1

(α) Σε έρευνα για τη χρήση του διαδικτύου από φοιτητές που έγινε το 2006 σε ένα δείγμα φοιτητών, σε ερώτημα για τις ώρες που ανά εβδομάδα χρησιμοποιούν το διαδίκτυο, πήραμε τις απαντήσεις και σχημάτισαμε τα παρακάτω γραφήματα και τα περιγραφικά μέτρα για το δείγμα (δες ιστόγραμμα). [Χρησιμοποιείστε τις τιμές των μέτρων σε ακρίβεια πρώτου δεκαδικού ψηφίου].



(i) Με βάση το θηκόγραμμα και το ιστόγραμμα από το δείγμα, σχολιάστε την κατανομή του χρόνου που ανά εβδομάδα χρησιμοποιούν οι φοιτητές το διαδίκτυο. Σχολιάστε τις τιμές 275 και 312 στο θηκόγραμμα.

(ii) Χρησιμοποιώντας διάστημα εμπιστοσύνης σε επίπεδο 99% εξετάστε αν ο μέσος χρόνος που ανά εβδομάδα χρησιμοποιούν οι φοιτητές το διαδίκτυο μπορεί να υπερβαίνει τις 6 ώρες.

(iii) Η ίδια έρευνα επαναλήφθηκε το 2009 σε δείγμα 450 φοιτητών και υπολογίστηκαν η δειγματική μέση τιμή ως 5.5 ώρες τη βδομάδα και η δειγματική τυπική απόκλιση ως 1.9 ώρες τη βδομάδα. Με βάση τα δύο δείγματα, του 2006 και του 2009 και χρησιμοποιώντας κατάλληλο διάστημα εμπιστοσύνης σε επίπεδο 95% εξετάστε αν ο μέσος χρόνος που ανά εβδομάδα χρησιμοποιούν οι φοιτητές το διαδίκτυο διαφέρει και πόσο στα δύο έτη.

(γ) Στα πλαίσια έρευνας για την εξάρτηση της χρήσης του διαδικτύου από την ηλικία, μετρήθηκε η ηλικία (σε έτη) και ο εβδομαδιαίος χρόνος χρήσης του διαδικτύου σε 7 φοιτητές και τα αποτελέσματα δίνονται στον παρακάτω πίνακα.

Ηλικία	18	19	19	21	22	23	25
Χρόνος διαδικτύου	6	7	5	4	6	8	7

Κάνετε κατάλληλο διάγραμμα διασποράς και υπολογίστε το συντελεστή συσχέτισης για το παραπάνω δείγμα. Με βάση αυτά τα αποτελέσματα, εξηγήστε αν φαίνεται να υπάρχει εξάρτηση του χρόνου χρήσης διαδικτύου από την ηλικία του φοιτητή και αν ναι τι είδους είναι αυτή. Επίσης εξηγήστε αν μπορούμε να κάνουμε ακριβείς προβλέψεις (εκτιμώντας την ευθεία ελαχίστων τετραγώνων) του χρόνου χρήσης του διαδικτύου από την ηλικία του φοιτητή.

## Άσκηση 2

Στη διπλωματική του εργασία ένας φοιτητής ανέπτυξε ένα νέο αλγόριθμο A και θέλει να τον συγκρίνει ως προς την ταχύτητα επίλυσης προβλημάτων με ένα άλλο γνωστό αλγόριθμο B. Για τον αλγόριθμο B έχει από τη βιβλιογραφία πως σε 16 προβλήματα ενός συγκεκριμένου τύπου που δοκιμάστηκε είχε μέσο χρόνο εκτέλεσης 12 ms (χρόνος της κεντρικής μονάδας επεξεργασίας σε χιλιοστά δευτερολέπτου) και τυπική απόκλιση 3 ms. Ο φοιτητής δοκίμασε τον αλγόριθμο του A σε 12 προβλήματα ίδιου τύπου και μέτρησε τους χρόνους εκτέλεσης που δίνονται παρακάτω (σε ms)

11.2	12.3	13.4	14.7	11.0	17.1	14.4	10.7	11.4	14.1	15.5	14.8
------	------	------	------	------	------	------	------	------	------	------	------

(α) Κάνετε το θηκόγραμμα για το δείγμα των 12 χρόνων εκτέλεσης του αλγόριθμου A και σχολιάστε με βάση το θηκόγραμμα αν μπορεί η κατανομή του χρόνου εκτέλεσης του αλγορίθμου A στο συγκεκριμένο τύπο προβλημάτων να είναι κανονική.

(β) Εξετάστε υπολογίζοντας κατάλληλο 95% διάστημα εμπιστοσύνης αν μπορεί να θεωρηθεί πως η τυπική απόκλιση του χρόνου εκτέλεσης του αλγορίθμου A στο συγκεκριμένο τύπο προβλημάτων μπορεί να είναι αυτή του αλγορίθμου B που βρέθηκε σε 16 παρόμοια προβλήματα (3 ms).

(γ) Θεωρείστε πως ο χρόνος εκτέλεσης αλγορίθμου (είτε του A ή του B) ακολουθεί κανονική κατανομή με την ίδια διασπορά ανεξάρτητα του αλγορίθμου. Για το συγκεκριμένο τύπο προβλημάτων εξετάστε υπολογίζοντας κατάλληλο 90% διάστημα εμπιστοσύνης αν μπορεί να θεωρηθεί πως οι δύο αλγόριθμοι έχουν τον ίδιο μέσο χρόνο εκτέλεσης και αν αυτός διαφέρει μεταξύ των δύο αλγορίθμων πόση είναι η διαφορά.

Ο αλγόριθμος A δοκιμάστηκε επίσης και σε προβλήματα με διαφορετικό βαθμό δυσκολίας που δίνεται από το μέγεθος των δεδομένων του προβλήματος και βρέθηκαν τα παρακάτω αποτελέσματα.

Πλήθος δεδομένων (σε χιλιάδες)	7	9	11	13	15
Χρόνος εκτέλεσης (σε φυσικό λογάριθμο ms)	5.1	9.9	13.3	15.5	18.1

όπου τα δεδομένα δίνονται σε κλίμακα χιλιάδων (δηλαδή αν το πλήθος δεδομένων είναι  $n$ , στον παραπάνω πίνακα η τιμή είναι  $n/1000$ ) και επίσης ο χρόνος εκτέλεσης είναι σε κλίμακα φυσικού λογαρίθμου (δηλαδή αν ο χρόνος εκτέλεσης είναι  $t$ , στον παραπάνω πίνακα η τιμή είναι  $\log_e(t)$ ).

(δ) Σχηματίστε το διάγραμμα διασποράς και με βάση αυτό το διάγραμμα σχολιάστε αν φαίνεται να εξαρτάται, και πως, ο χρόνος εκτέλεσης αλγορίθμου A (σε κλίμακα φυσικού λογαρίθμου) από το πλήθος των δεδομένων (σε κλίμακα χιλιάδων).

(ε) Υπολογίστε τις εκτιμήσεις των παραμέτρων της γραμμικής παλινδρόμησης του χρόνου εκτέλεσης του αλγορίθμου A (σε κλίμακα φυσικού λογαρίθμου) ως προς το βαθμό δυσκολίας (πλήθος δεδομένων προβλήματος σε χιλιάδες) με τη μέθοδο ελαχίστων τετραγώνων.

(στ) Για τον αλγόριθμο B είναι γνωστό από τη βιβλιογραφία πως αν  $n$  είναι το πλήθος των δεδομένων στο πρόβλημα, ο χρόνος εκτέλεσης  $t$  αυξάνει ως  $t \propto e^{3n}$ , όπου  $\propto$  σημαίνει «ανάλογα με», δηλαδή  $t = a \cdot e^{3n}$  για κάποιο  $a$ . Με βάση την ευθεία ελαχίστων τετραγώνων που βρήκατε παραπάνω για τον αλγόριθμο A, σχολιάστε αν ο χρόνος εκτέλεσης του αλγορίθμου A αυξάνει επίσης με τον ίδιο ρυθμό, δηλαδή αυτόν του αλγορίθμου B.