



1η Άσκηση: παράδοση Παρασκευή 1/11

1. Δίνονται τα ακόλουθα δεδομένα.

Δεδομένα I									
30.3	31.0	31.1	32.1	32.6	32.7	33.4	33.6	34.2	34.5

Δεδομένα II									
0.0	0.0	0.2	0.8	1.2	1.4	3.2	4.2	6.4	9.0

Δεδομένα III									
0	1	6	8	10	13	15	16	17	17
18	18	20	20	21	25	26	30	35	39
40	41	43	44	46	48	52	54	58	59
59	60	66	81	86	87	88	89	94	96

- a. Για κάθε ομάδα δεδομένων δώστε το stemplot και boxplot – χωρίς τη χρήση λογισμικού/υπολογιστή.
- b. Για κάθε ομάδα δεδομένων, ποια ομάδα τιμών συνοψίζει καλύτερα την κατανομή: η μέση τιμή και τυπική απόκλιση ή η σύνοψη των 5 αριθμών;
- c. Πόσο ακριβής θα ήταν η προσέγγιση της κατανομής των δεδομένων από μια καμπύλη πυκνότητας της Κανονικής κατανομής; (Δώστε συγκεκριμένα παραδείγματα εγγύτητας ή απόκλισης των ποσοστημορίων τους.)
2. Βρείτε στο διαδίκτυο δεδομένα σε μορφή πίνακα με τουλάχιστον 1 κατηγορική και 2 ποσοτικές μεταβλητές, για οποιοδήποτε θέμα σας ενδιαφέρει και μεταφορτώστε τα σε ένα αρχείο στον υπολογιστή σας, ώστε να μπορείτε να τα επεξεργαστείτε με λογισμικό στατιστικής ανάλυσης, πχ., της γλώσσας R που διατίθεται ελεύθερα στο <http://www.r-project.org/>. (Στο eclass του μαθήματος μπορείτε να βρείτε συνδέσμους προς ιστοσελίδες οργανισμών που προσφέρουν ελεύθερη πρόσβαση σε δεδομένα διαφόρων θεμάτων. Τα περισσότερα δεδομένα μπορούν να μεταφορτωθούν στον υπολογιστή σας σε μορφή Microsoft Excel, CSV ή ως αρχείο χαρακτήρων (text).)
- a. Δώστε μια σύντομη περιγραφή από που προέρχονται τα δεδομένα και πόσες περιπτώσεις περιέχονται.

- b. Ποιες είναι κατηγορικές και ποιες ποσοτικές μεταβλητές; Δώστε μια σύντομη περιγραφή κάθε μιας από αυτές (ή ορισμένων εάν είναι πάρα πολλές).
 - c. Δώστε τις κατανομές των μεταβλητών σε γραφική μορφή. Σχολιάστε τη μορφή των κατανομών, πιθανούς λόγους που έχουν αυτή τη μορφή, την ύπαρξη ατυπικών σημείων (outliers) κτλ.
 - d. Για κάθε ποσοτική μεταβλητή, υπολογίστε α) τη μέση τιμή και τυπική απόκλιση, και β) τη σύνοψη των πέντε αριθμών. Σχολιάστε την καταλληλότητα των α), β) για κάθε μεταβλητή.
 - e. Επιλέξτε δύο μεταβλητές και διερευνήστε τη σχέση τους. Εάν θεωρήσετε ότι υπάρχει σχέση, αυτή είναι αιτιατή ή όχι; Σχολιάστε αναλόγως.
3. Εδώ θα διερευνήσετε τη σχέση μεταξύ ύψους (μεταβλητή height) και μεγέθους παπουτσιού (μεταβλητή size) στα δεδομένα των απαντήσεων ερωτηματολογίου 2019 που βρίσκονται στο eclass.
- a. Δώστε το scatterplot και σχολιάστε τη *μορφή, κατεύθυνση και δύναμη* της σχέσης των δύο μεταβλητών.
 - b. Υπολογίστε τον συντελεστή συσχέτισης και εκτελέστε γραμμική παλινδρόμηση ελαχίστων τετραγώνων.

Σχόλια/οδηγίες:

1. Για τον υπολογισμό αριθμητικών περιγραφών και γραφικών παραστάσεων μπορείτε να χρησιμοποιήσετε οποιοδήποτε λογισμικό στατιστικής ανάλυσης επιθυμείτε, εκτός στο ερώτημα 1(α) το οποίο θα πρέπει να κάνετε χωρίς τη βοήθεια υπολογιστή.
2. Τη στατιστική ανάλυση θα πρέπει να την κάνετε εσείς οι ίδιοι με τη χρήση λογισμικού— δεν επιτρέπεται να χρησιμοποιήσετε «έτοιμα» γραφήματα/αποτελέσματα κτλ. των δεδομένων που επιλέξατε.
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!)
4. Για τη 2η άσκηση, ΔΕΝ μπορείτε να χρησιμοποιήσετε τα δεδομένα του ερωτηματολογίου που βρίσκονται στο eclass καθώς και δεδομένα που

υπάρχουν σε ιστότοπους για εκπαιδευτικούς σκοπούς. (Ούτε και δεδομένα που περιλαμβάνονται στην R ή άλλο λογισμικό.)

5. Τα δεδομένα που μπορείτε να χρησιμοποιήσετε θα πρέπει να τα βρείτε από ιστότοπους οργανισμών/εταιριών/ιδρυμάτων συγκεκριμένου ενδιαφέροντος, για παράδειγμα, από στατιστικές υπηρεσίες, οικονομικούς, αθλητικούς οργανισμούς, μη εκπαιδευτικά ιστολόγια. Αν δυσκολεύεστε να εντοπίσετε τέτοιους ιστότοπους, στο eclass μπορείτε να βρείτε συνδέσμους ορισμένων οργανισμών.