



## Πανεπιστήμιο Πατρών Τμήμα Οικονομικών Επιστημών

Ακαδημαϊκό έτος 2025 - 2026

### 2<sup>η</sup> Εργασία μαθήματος «Εισαγωγή στην Επιστήμη Δεδομένων»: Στατιστική Επεξεργασία δεδομένων με τη χρήση της γλώσσας προγραμματισμού R

#### Σκοπός

Στην 2<sup>η</sup> εργασία καλείστε να εκτελέσετε δύο εργασίες: να δώσετε μία περίληψη ομιλίας και να αναλύσετε δεδομένα με χρήση της γλώσσας R.

#### I. Περίληψη ομιλίας (20%)

Την Παρασκευή 19 Δεκεμβρίου το Τμήμα Οικονομικών Επιστημών θα τιμήσει με τον τίτλο του επίτιμου διδάκτορα τον Καθηγητή στο London Business School (LBS) Ηλία Παπαϊωάννου (<https://www.london.edu/faculty-and-research/faculty-profiles/p/papaioannou-e>). Η εκδήλωση θα γίνει στις 12:00 στην αίθουσα τελετών του Πανεπιστημίου Πατρών που βρίσκεται στο κτίριο Α' της Πρυτανείας. Ο κος Παπαϊωάννου στα πλαίσια της επιτιμοποίησης θα κάνει μία ομιλία πάνω σε θέμα που άπτεται των οικονομικών.

Παρακολουθήστε την ομιλία του κ. Παπαϊωάννου και παραθέστε μία περίληψη της ομιλίας σε όχι λιγότερο από 750 λέξεις.

#### II. Ανάλυση δεδομένων με R (80%)

Στο ζητούμενο αυτό, θα φτιάξετε ένα πρόγραμμα στη γλώσσα προγραμματισμού R, το οποίο επεξεργάζεται στατιστικά τα δεδομένα της απογραφής του 2021, τα οποία δίνονται μαζί με την εκφώνηση της εργασίας. **Για να απαντήσετε στις ερωτήσεις της εργασίας αυτής κάνετε αυστηρά χρήση του αρχείου .csv που σας δίνεται μαζί με την εκφώνηση. Για την 2<sup>η</sup> εργασία, μην κάνετε χρήση του αρχείου .xlsx που δόθηκε μαζί με την 1<sup>η</sup> εργασία.**

Το αρχείο GrCensus2021.csv που δίνεται συνημμένα μαζί με την εκφώνηση της εργασίας, παρέχει πληροφορίες για τον πληθυσμό της Ελλάδας, όπως αυτές προέκυψαν κατά την απογραφή του 2021 και δημοσιεύτηκαν από την Ελληνική Στατιστική Αρχή (<http://www.statistics.gr/>) σε μορφή που λέγεται CSV (Comma-Separated values).

Το αρχείο GrCensus2021.csv περιέχει ακριβώς τα ίδια δεδομένα απογραφής που περιέχει και το αρχείο GrCensus2021.xls της 1<sup>ης</sup> Εργασίας. Ωστόσο, το αρχείο που δίνεται έχει μία ελαφρώς διαφορετική δομή, προκειμένου να διευκολυνθείτε στην συγγραφή των προγραμμάτων που απαντούν στα ερωτήματα της εργασίας αυτής. Κάντε αυστηρά χρήση του αρχείου GrCensus2021.csv που δίνεται με την εκφώνηση.

Ειδικότερα, το αρχείο GrCensus2021.csv που θα επεξεργαστείτε στα πλαίσια της 2<sup>ης</sup> εργασίας περιέχει τα δεδομένα σε μορφή που καλείται csv<sup>1</sup> (Comma-Separated Values). Τα αρχεία csv είναι αρχεία απλού κειμένου, τα οποία περιέχουν γραμμές, όπου σε κάθε γραμμή οι τιμές διαχωρίζονται με έναν ειδικό χαρακτήρα. Ο χαρακτήρας διαχωρισμού (separator) μπορεί να είναι είτε το κόμμα (,), ο χαρακτήρας tab, το semicolon (;) ή άλλος χαρακτήρας. Στο αρχείο GrCensus2021.csv που σας δίνεται μαζί με την εκφώνηση της 2<sup>ης</sup> εργασίας, ο χαρακτήρας διαχωρισμού των τιμών κάθε γραμμής είναι ο χαρακτήρας semicolon (;).

Το αρχείο GrCensus2021.csv περιέχει τις ίδιες γραμμές και τις ίδιες στήλες με το αρχείο GrCensus2021.xlsx της 1<sup>ης</sup> εργασίας.

## Ζητούμενα

Συγγράψτε ένα πρόγραμμα με τη γλώσσα προγραμματισμού R, το οποίο θα διαβάξει το αρχείο GrCensus2021.csv, θα το επεξεργάζεται καταλλήλως και εμφανίζει στην οθόνη τα παρακάτω:

- 1) Το όνομα και τον συνολικό πληθυσμό εκείνων των περιφερειών της Ελλάδας, με τον μικρότερο συνολικό πληθυσμό γυναικών ανεξαρτήτου ηλικίας.
- 2) Τον μέσο όρο του πληθυσμού των γυναικών ανεξαρτήτου ηλικίας όλων των δημοτικών ενοτήτων.
- 3) Το ποσοστό των δήμων όπου ο πληθυσμός των αντρών ηλικίας μεταξύ 25 ( $\geq 25$ ) και 39 ετών ( $< 39$ ) είναι μεγαλύτερος από 15000 ( $\geq 15000$ ).
- 4) Τη διάμεσο του συνολικού πληθυσμού ανεξαρτήτου ηλικίας και φύλου όλων των δημοτικών ενοτήτων.
- 5) Τη διακύμανση και τη τυπική απόκλιση του πληθυσμού των ανδρών με ηλικία μεγαλύτερη ή ίση από 45 ετών ( $\geq 45$ ) όλων των δημοτικών ενοτήτων.
- 6) Απεικονίστε με τη μορφή ραβδογράμματος (μορφή ράβδων) τους συνολικούς πληθυσμούς ανεξαρτήτου φύλου και ηλικίας όλων των περιφερειών της Ελλάδας. Στο ραβδόγραμμα θα πρέπει:
  - a. Να εμφανίζεται κάθε ράβδος περιφέρειας με διαφορετικό χρώμα
  - b. Τα ονόματα των περιφερειών να εμφανίζονται ολόκληρα
- 7) Πόσες ακραίες τιμές έχει ο συνολικός πληθυσμός ανεξαρτήτου φύλου και ηλικίας των περιφερειακών ενοτήτων της Ελλάδας; Δημιουργήστε το κατάλληλό διάγραμμα όπου να φαίνεται το πλήθος των ακραίων τιμών.

## III.Ομάδες εργασίας

Η εργασία θα εκπονηθεί ομαδικά. Δεν θα χρειαστεί, ωστόσο, να δημιουργηθούν νέες ομάδες: η 2<sup>η</sup> εργασία θα εκπονηθεί από τις ίδιες ομάδες που εκπονούν την 1<sup>η</sup> εργασία.

## Τρόπος και ημερομηνία παράδοσης

Για την παράδοση της εργασίας θα πρέπει να σταλούν στο eclass όλα τα αρχεία που περιέχουν τον πηγαίο κώδικα σε R που έχετε δημιουργήσει και ο οποίος, εάν εκτελεστεί, θα απαντά στα ζητούμενα της εργασίας.

Ειδικότερα το (ή τα) αρχεία που περιέχουν τον πηγαίο κώδικα σε R θα πρέπει να συμπιεστούν σε ένα αρχείο μορφής .zip ή .rar και το οποίο θα πρέπει να αναρτηθεί, από ένα μέλος της

---

<sup>1</sup> [https://en.wikipedia.org/wiki/Comma-separated\\_values](https://en.wikipedia.org/wiki/Comma-separated_values)

ομάδας, στο χώρο της κάθε ομάδας που έχει δημιουργηθεί στο [eclass.upatras.gr](https://eclass.upatras.gr) . Η διαδικασία παράδοσης της εργασίας μπορεί να γίνει από οποιοδήποτε μέλος της ομάδας.

**Η καταληκτική ημερομηνία παράδοσης της 2<sup>ης</sup> εργασίας είναι η Παρασκευή, 16 Ιανουαρίου 2026 00:00:00 (GMT+2).**

#### **IV.Ερωτήσεις/Απορίες**

Για οποιαδήποτε ερώτηση ή απορία σχετικά με την εργασία μπορείτε να στείλετε email στη διεύθυνση [tzagara@upatras.gr](mailto:tzagara@upatras.gr) ή [daskalou@upatras.gr](mailto:daskalou@upatras.gr) . Απορίες μπορούν επίσης **(και συστήνεται!)** να συζητηθούν κατά τη διάρκεια του μαθήματος.

#### **V.Βαρύτητα της εργασίας**

Η εργασία είναι υποχρεωτική για τους φοιτητές του πρώτου έτους και συνεισφέρει το 15% του τελικού τους βαθμού. Η εργασία δεν είναι υποχρεωτική για τους φοιτητές μεγαλύτερων ετών.

Καλή επιτυχία!