



Πανεπιστήμιο Πατρών Τμήμα Οικονομικών Επιστημών

Ακαδημαϊκό έτος 2023 - 2024

2^η Εργασία μαθήματος «Εισαγωγή στους Η/Υ και Εφαρμογές»: Στατιστική Επεξεργασία δεδομένων με τη χρήση της γλώσσας προγραμματισμού R

<https://www.youtube.com/watch?v=EcOi6nplsVc>

-Τα ρομπότ δεν θα μας πάρουν μόνο τις δουλειές. Αλλά και τις γυναίκες

<https://www.youtube.com/watch?v=oY2nVQNIUB8>

-Γνωρίζετε τον Scott Sterling; Τι, όχι;

Σκοπός

Στόχος της 2^{ης} εργασίας είναι να έρθει ο φοιτητής σε επαφή με τον προγραμματισμό υπολογιστών ώστε να δει ο ίδιος πόσο εύκολο είναι σήμερα να φτιαχτούν προγράμματα που επεξεργάζονται δεδομένα μεγάλου όγκου.

Στα πλαίσια της 2^{ης} εργασίας, θα φτιάξετε ένα πρόγραμμα στη γλώσσα προγραμματισμού R, το οποίο επεξεργάζεται στατιστικά τα δεδομένα της απογραφής του 2011, 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^η εργασία, μην κάνετε χρήση του αρχείου .xlsx που δόθηκε μαζί με την 1^η εργασία.**

Εισαγωγή

Το αρχείο GrCensus2011.csv που δίνεται συνημμένα μαζί με την εκφώνηση της εργασίας, παρέχει πληροφορίες για τον πληθυσμό της Ελλάδας, όπως αυτές προέκυψαν κατά την απογραφή του 2011 και δημοσιεύτηκαν από την Ελληνική Στατιστική Αρχή (<http://www.statistics.gr/>)¹ σε μορφή που λέγεται CSV (Comma-Separated values).

Το αρχείο GrCensus2011.csv περιέχει ακριβώς τα ίδια δεδομένα απογραφής που περιέχει και το αρχείο GrCensus2011.xls της 1^η 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 GrCensus2011.csv που δίνεται με την εκφώνηση.

Ειδικότερα, το αρχείο GrCensus2011.csv που θα επεξεργαστείτε στα πλαίσια της 2^{ης} εργασίας περιέχει τα δεδομένα σε μορφή που καλείται csv² (Comma-Separated Values). Τα αρχεία csv

¹ Τα στοιχεία ελήφθησαν από τη σελίδα <http://www.statistics.gr/portal/page/portal/ESYE/PAGE-census2011tables> (Πίνακες Απογραφής 201, Πίνακας 4α: Απογραφή Πληθυσμού 2011. Μόνιμος Πληθυσμός κατά φύλο και ομάδες ηλικιών Μεγάλες Γεωγραφικές Ενότητες (NUTS 1), Αποκεντρωμένες Διοικήσεις, Περιφέρειες (NUTS 2), Περιφερειακές Ενότητες, Δήμοι, Δημοτικές Ενότητες)

² https://en.wikipedia.org/wiki/Comma-separated_values

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 GrCensus2011.csv που σας δίνεται μαζί με την εκφώνηση της 2^{ης} εργασίας, ο χαρακτήρας διαχωρισμού των τιμών κάθε γραμμής είναι ο χαρακτήρας semicolon (;).

Το αρχείο GrCensus2011.csv περιέχει τις ίδιες γραμμές και τις ίδιες στήλες με το αρχείο GrCensus2011.xlsx της 1^{ης} εργασίας. Μπορείτε να χρησιμοποιήσετε το αρχείο GrCensus2011.xlsx για να προσανατολίξετε σχετικά με τις στήλες που πρέπει να προσπελάσετε/επεξεργαστείτε στο αρχείο GrCensus2011.csv.

Ζητούμενα της εργασίας

Συγγράψτε ένα πρόγραμμα με τη γλώσσα προγραμματισμού R, το οποίο θα διαβάσει³ το αρχείο GrCensus2011.csv, θα το επεξεργάζεται καταλλήλως και εμφανίζει στην οθόνη τα παρακάτω:

- 1) Το όνομα και τον συνολικό πληθυσμό εκείνων των δήμων, με τον μεγαλύτερο συνολικό πληθυσμό γυναικών ανεξαρτήτου ηλικίας.
- 2) Τον μέσο όρο του πληθυσμού των αντρών ανεξαρτήτου ηλικίας όλων των δημοτικών ενοτήτων.
- 3) Το πλήθος των δημοτικών ενοτήτων όπου ο πληθυσμός των αντρών ηλικίας μεταξύ 0 (≥ 0) και 20 ετών (< 20) είναι μεγαλύτερος από 5000.
- 4) Τη διάμεσο του συνολικού πληθυσμού ανεξαρτήτου ηλικίας και φύλου όλων των δημοτικών ενοτήτων.
- 5) Τη διακύμανση και τη τυπική απόκλιση του πληθυσμού των γυναικών με ηλικία μεγαλύτερη ή ίση από 45 ετών (≥ 45) όλων των δημοτικών ενοτήτων.

Ομάδες εργασίας

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

Τρόπος και ημερομηνία παράδοσης

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

Ειδικότερα το (ή τα) αρχεία που περιέχουν τον πηγαίο κώδικα σε R θα πρέπει να συμπεριστούν σε ένα αρχείο μορφής .zip ή .rar και το οποίο θα πρέπει να αναρτηθεί, από ένα μέλος της ομάδας, στο εργαλείο Εργασίες. Η διαδικασία παράδοσης της εργασίας μπορεί να γίνει από οποιοδήποτε μέλος της ομάδας.

Η καταληκτική ημερομηνία παράδοσης της 2^{ης} εργασίας είναι η Δευτέρα, 11 Δεκεμβρίου 2023 23:55 (GMT+2).

³ **ΣΗΜΑΝΤΙΚΟ:** Στην εντολή ανάγνωσης του αρχείου GrCensus.csv, θα πρέπει απαραίτητως να προσθέσετε το όρισμα `fileEncoding = "utf8"` προκειμένου να διαβαστεί σωστά το αρχείο (και ειδικά οι ελληνικοί χαρακτήρες). Εάν δεν εκτελέσετε την εντολή ανάγνωσης με αυτό το όρισμα, ενδέχεται να μην εμφανιστούν σωστά οι ελληνικοί χαρακτήρες.

Ερωτήσεις/Απορίες

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

Βαρύτητα της εργασίας

Η εργασία είναι υποχρεωτική για τους φοιτητές του πρώτου έτους και συνεισφέρει το 10% του τελικού τους βαθμού. Η εργασία δεν είναι υποχρεωτική για τους φοιτητές μεγαλύτερων ετών.

Καλή επιτυχία!