



Πανεπιστήμιο Αιγαίου

Τμήμα Επιστημών της Θάλασσας-Σχολή Περιβάλλοντος

Ανοικτό ακαδημαϊκό μάθημα

Εισαγωγή στην Πληροφορική και Προγραμματισμός

Διδάσκοντες:

Γ. Τσιρτσής, Καθηγητής
Δρ Β. Κολοβογιάννης, ΕΔΙΠ



Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο



ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



Άδειες Χρήσης

- Το παρόν εκπαιδευτικό υλικό υπόκειται σε άδειες χρήσης Creative Commons.
- Για εκπαιδευτικό υλικό, όπως εικόνες, που υπόκειται σε άλλου τύπου άδειας χρήσης, η άδεια χρήσης αναφέρεται ρητώς.



Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στα πλαίσια του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «**Ανοικτά Ακαδημαϊκά Μαθήματα στο Πανεπιστήμιο Αιγαίου**» έχει χρηματοδοτήσει μόνο τη αναδιαμόρφωση του εκπαιδευτικού υλικού.
- Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Δια Βίου Μάθηση» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.



Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο



ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ



7η Διάλεξη

Αρχεία εισόδου και εξόδου

Η έννοια του αρχείου δεδομένων (Data file)-Δημιουργία αρχείου δεδομένων

Το αρχείο δεδομένων είναι ένα μόνιμα αποθηκευμένο αρχείο που περιέχει δεδομένα αριθμητικά ή αλφαριθμητικά (λέξεις, κείμενο).

Ο τύπος αρχείου που είναι ευκολότερο να διαβάσει και να γράψει η R είναι τα αρχεία csv (comma separated values ή κείμενο διαχωρισμένο με κόμματα). Τα αρχεία αυτά κατασκευάζονται σε λογισμικό φύλλου εργασίας (π.χ. LibreOffice Calc ή MS Excel) συμπληρώνοντας τιμές στις γραμμές και στήλες ενός φύλλου εργασίας. Στη συνέχεια το αρχείο αποθηκεύεται ως αρχείο τύπου csv. Αν ο υπολογιστής έχει ελληνική ρύθμιση, ως δεκαδικό σύμβολο χρησιμοποιείται το κόμμα, οπότε επιλέγουμε ως διαχωριστικό των τιμών των κελιών της ίδιας γραμμής το ελληνικό ερωτηματικό (;). Αν ο υπολογιστής έχει αγγλική ρύθμιση, ως δεκαδικό σύμβολο χρησιμοποιείται η τελεία, οπότε επιλέγουμε ως διαχωριστικό των τιμών των κελιών της ίδιας γραμμής το κόμμα (,).

Αρχείο csv όπως εμφανίζεται σε έναν κειμενογράφο.

Στη πρώτη γραμμή εμφανίζονται οι επικεφαλίδες των 2 στηλών και στις επόμενες γραμμές οι τιμές σε 3 γραμμές και 2 στήλες.

Οι τιμές μεταξύ τους διαχωρίζονται με κόμμα και δεκαδικό σύμβολο είναι η τελεία.

Το αρχείο κατασκευάστηκε με λογισμικό LibreOffice Calc

```
GDM_bcc.csv x
Geographic, sal_A
0.943620206198332, 0
0.00587294631791026, 0.00488115374189212
0, 0.213216571261492
```



7η Διάλεξη

Αρχεία εισόδου και εξόδου

Εντολές ανάγνωσης δεδομένων από αρχείο

Για την επικοινωνία της R με αρχείο αυτό πρέπει να βρίσκεται στον τρέχοντα φάκελο εργασίας (working directory) που η διαδικασία ορισμού του έχει περιγραφεί σε προηγούμενη διάλεξη (Διάλεξη 3). Για να διαβάσουμε τα περιεχόμενα ενός αρχείου και να τα καταχωρίσουμε στη μεταβλητή με δείκτη με όνομα π.χ. x χρησιμοποιείται η εντολή:

```
x=read.csv(file="test.csv", sep=";", dec=".", header=FALSE) όπου "test.csv" παράδειγμα ονόματος αρχείου δεδομένων.
```

Το ερωτηματικό στο `sep=";"` είναι το διαχωριστικό μεταξύ τιμών της ίδιας γραμμής που χρησιμοποιήσαμε κατά την αποθήκευση του αρχείου csv και το κόμμα στο `dec="."` είναι το δεκαδικό σύμβολο του αρχείου csv. Αν το αρχείο csv έχει αποθηκευθεί σε αγγλική έκδοση και έχει δεκαδικό σύμβολο την τελεία και διαχωριστικό το κόμμα, θα γράφαμε αντίστοιχα:

```
x=read.csv(file="test.csv", sep=";", dec=".", header=FALSE).
```

Η επιλογή `header=FALSE` χρησιμοποιείται αν θέλουμε να διαβάζονται δεδομένα από την πρώτη γραμμή του αρχείου. Αν το αρχείο έχει στην πρώτη γραμμή τις επικεφαλίδες των στηλών, αρχίζουμε να διαβάζουμε τιμές δεδομένων από την δεύτερη γραμμή, οπότε γράφουμε `header=TRUE`.

Αν το αρχείο δεδομένων περιέχει μία σειρά τιμών (ένα διάνυσμα) σε μία γραμμή αυτές θα καταχωρηθούν στις μεταβλητές `x[1]`, `x[2]`, `x[3]` κλπ. Αν περιέχει πίνακα τιμών (τιμές σε γραμμές και στήλες) αυτές θα καταχωρηθούν στις μεταβλητές `x[i,j]` όπου `i` η γραμμή και `j` η στήλη. Σημειώνεται ότι αν το αρχείο περιέχει τιμές μόνο σε μία στήλη, αυτές θα καταχωρηθούν σε πίνακα (δύο διαστάσεις) ως π.χ. `y[1,1]`, `y[2,1]`, `y[3,1]` κ.λ.π. με τη δεύτερη διάσταση να είναι 1.

Αν επιθυμούμε να ορίσει ο χρήστης κατά τη διάρκεια εκτέλεσης του προγράμματος το όνομα του αρχείου γράφουμε τις εντολές:

```
print("Please define the name of the file") ή κάτι αντίστοιχο  
filename=scan(what="")
```

```
x=read.csv(file=filename, sep=";", dec=".", header=FALSE) όπου filename τυχαίο όνομα μεταβλητής στην οποία ο χρήστης καταχωρεί το όνομα του αρχείου.
```



7η Διάλεξη

Αρχεία εισόδου και εξόδου

Εντολές εκτύπωσης αποτελεσμάτων σε αρχείο

Για να αποθηκεύσουμε αποτελέσματα που βρίσκονται σε μορφή μεταβλητής, διανύσματος ή πίνακα σε αρχείο εξόδου χρησιμοποιούμε την εντολή:

`write.table(x, file="output.csv", row.names=FALSE, col.names=FALSE)` όπου `x` μεταβλητή την τιμή της οποίας θέλουμε να αποθηκεύσουμε (μπορεί να είναι και πολλές τιμές αν πρόκειται για μεταβλητή με δείκτη), `"output.csv"` παράδειγμα ονόματος αρχείου εξόδου. Τις επιλογές `row.names=FALSE` και `col.names=FALSE` τις χρησιμοποιούμε όταν δεν θέλουμε να αποθηκευθούν στο αρχείο εξόδου οι επικεφαλίδες-ονόματα των γραμμών και στηλών τις οποίες αυτόματα δημιουργεί η R για κάθε μεταβλητή. Αν τις θέλουμε γράφουμε `TRUE` αντί `FALSE`.

Και στην περίπτωση αυτή αν επιθυμούμε, μπορεί να ορίσει ο χρήστης κατά την διάρκεια εκτέλεσης του προγράμματος το όνομα του αρχείου γράφοντας τις εντολές:

```
print("Please define the name of the output file")
filename=scan(what="")
write.table(x, file=filename, row.names=FALSE, col.names=FALSE)
```

Για να γράψουμε σε ένα αρχείο που ήδη έχουμε γράψει αποτελέσματα, χωρίς να σβήσουν τα ήδη γραμμένα, χρησιμοποιούμε την δυνατότητα `append=TRUE` και γράφουμε την δεύτερη εντολή ως:

```
write.table(y, file="output.csv", append=TRUE, row.names=FALSE, col.names=FALSE)
```

Για να γράψουμε σε αρχείο μπορεί να χρησιμοποιηθεί και η εντολή `cat`, που αναφέρθηκε για εκτύπωση στην οθόνη στη Διάλεξη 3.

Η γενική μορφή της είναι `cat(....., file=filename, sep="...", append=TRUE)`, όπου τα `.....` στην αρχή της παρένθεσης αφορούν στις μεταβλητές που θα τυπωθούν, το `filename` είναι το όνομα του αρχείου στο οποίο θα γίνει η εγγραφή, τα `....` μετά το `sep` διάφορα σύμβολα και το `append=TRUE` ή `FALSE` όπως προηγουμένως.

Με το όρισμα `sep="..."` δηλώνεται τι επιθυμούμε να διαχωρίζει τις τιμές. Χρήσιμα διαχωριστικά είναι το κόμμα για αρχεία `.csv`, δηλαδή το όρισμα `sep=","`, το `sep="\t"` που βάζει ένα διαχωριστικό `tab` ανάμεσα στις τιμές, το `sep="\n"` που αλλάζει γραμμή μετά από κάθε τιμή που εγγράφεται.