

ΜΑΘΗΜΑ: "Εισαγωγή στον Προγραμματισμό"

ΕΡΓΑΣΤΗΡΙΟ 5: Πλαίσια Δεδομένων - Λίστες

Πλαίσια Δεδομένων

Άσκηση 1. Υποθέστε ότι στα πλαίσια μιας έρευνας, ρωτήθηκε ένας συγκεκριμένος αριθμός ατόμων από τον πληθυσμό και καταγράφηκαν οι απαντήσεις στις παρακάτω μεταβλητές:

1. Gender: Φύλλο (Male = Άνδρας, Female = Γυναίκα)
2. Smoking: Καπνιστική συνήθεια (Yes = καπνιστής, No = μη καπνιστής)
3. Cholesterol: Χοληστερίνη (σε mg/dL)
4. Height: Ύψος (σε cm)
5. Weight: Βάρος (σε kg)

(i). Να δημιουργήσετε ένα πλαίσιο δεδομένων (data frame) για τα παρακάτω δεδομένα, όπου τη μεταβλητή Gender να την εισάγετε ως παράγοντα (factor). Το πλαίσιο δεδομένων να το ονομάσετε df1.

AA	Gender	Smoking	Cholesterol	Height	Weight
1	Male	Yes	200	170	85
2	Male	Yes	220	175	92
3	Male	No	180	182	78
4	Female	Yes	172	165	61
5	Female	No	184	166	60
6	Female	Yes	218	170	62
7	Male	Yes	205	179	90
8	Female	No	196	159	55
9	Female	No	199	161	58
10	Female	No	202	165	60
11	Male	Yes	217	188	97
12	Male	No	178	190	85

Για διευκόλυνσή σας, σας δίνεται ο κώδικας για την κατασκευή του data frame.

```
> gender<-c("Male","Male","Male","Female","Female","Female","Male","Female","Female","Female","Male","Male")
> gender<-factor(gender)
> smoking<-c("Y","Y","N","Y","N","Y","Y","N","N","N","Y","N")
> cholesterol<-c(200,220,180,172,184,218,205,196,199,202,217,178)
> height<-c(170,175,182,165,166,170,179,159,161,165,188,190)
> weight<-c(85,92,78,61,60,62,90,55,58,60,97,85)
> df1<-data.frame(gender,smoking,cholesterol,height,weight)
```

(ii). Να εισάγετε στα δεδομένα μια επιπλέον στήλη στην οποία θα δίνεται η τιμή του δείκτη μάζας σώματος (body mass index, $bmi = weight/height^2$). Να ονομάσετε τη στήλη αυτή bmi. Χρησιμοποιήστε την εντολή transform.

(iii). Έστω ότι για κάθε μέλος του δείγματος, έχετε στη διάθεσή σας και την αντίστοιχη μέτρηση για την τιμή της συστολικής πίεσης (σε mm Hg):

133 145 110 117 120 138 136 128 130 146 140 115

Να προσθέσετε τις παραπάνω τιμές στο data frame ως μια νέα στήλη με όνομα Sys. Χρησιμοποιήστε την εντολή cbind.

(iv). Με ποια εντολή θα εμφανίσετε τα στοιχεία του πλαισίου δεδομένων για τους καπνιστές; Με ποια εντολή θα εμφανίσετε τα στοιχεία του πλαισίου δεδομένων για τους άνδρες;

(vi). Χρησιμοποιήστε την εντολή `write.table` για να αποθηκεύσετε το data frame που δημιουργήσατε ως ένα αρχείο txt με όνομα `dataLab5.txt` (βεβαιωθείτε ότι γνωρίζετε που ακριβώς έχει αποθηκευτεί).

Άσκηση 2 (Homework): Ακολουθήστε τις παρακάτω εντολές σχετικά με τη δημιουργία και τη διαχείριση πλαισίων δεδομένων.

(i) Δημιουργήστε το ακόλουθο data frame με τον editor και ονομάστε το `MYDF`. Δείτε επίσης Ντζούφρας & Καρλής (2016), ενότητα 4.1.1 Αρχικά, δώστε την εντολή `MYDF<-data.frame()`. Στη συνέχεια δώστε την εντολή `MYDF<-edit(MYDF)`. Θα ανοίξει ένα φύλλο εργασίας (όπως π.χ. στο `CALC`) οποίο μπορείτε να γράψετε τα δεδομένα, να δώσετε τα ονόματα των μεταβλητών κλπ. Αφού περάσετε τα δεδομένα, στη συνέχεια κλείνετε το παράθυρο του editor και τα δεδομένα θα εμφανιστούν στο workspace της R.

	AA	Calculus	Informati>	Algebra	Gender	Smoker	Year
1	ID001	5	6	3	F	Yes	First
2	ID002	5.5	5	3.5	M	No	First
3	ID003	6	7	2	M	Yes	Second
4	ID004	3	3.5	5.5	F	No	Third
5	ID005	2.5	1	2	M	No	Second
6	ID006	2	4	6	M	No	First
7	ID007	1	5.5	6	M	No	Fourth
8	ID008	7	8.5	6.5	F	No	Fourth
9	ID009	8.5	9	7	F	No	Second
10	ID010	6	7	7.5	F	Yes	First
11	ID011	5.5	6	6.5	M	No	Third
12	ID012	6	5	7	F	No	First

(ii) Δημιουργήστε το ίδιο data frame με εντολές και αποθηκεύστε το με το όνομα `MYDF2`.

(iii) Και στα δύο data frames να ορίσετε τα ονόματα των παρατηρήσεων: Jane, Michael, John, Mary, Jose, Scottie, Nick, Helen, Sarah, Wendy, James, Carol.

(iv) Ορίστε τη μεταβλητή `Smoker` ως κατηγορική μεταβλητή (παράγοντα) ενώ τη μεταβλητή `Year`, να την ορίσετε διατάξιμη κατηγορική.

(v) Δημιουργήστε ένα νέο data frame με τα ακόλουθα στοιχεία και το όνομα `MYDFtemp`

Working

Jane **No**
Michael **No**
John **Yes**
Mary **No**
Jose **No**
Scottie **No**
Nick **Yes**

Helen	Yes
Sarah	No
Wendy	No
James	Yes
Carol	No

(vi) Προσθέστε τη μεταβλητή Working στο MYDF2.

(vii) Ποιες είναι οι διαστάσεις του MYDF2;

(viii) Ερμηνεύστε το αποτέλεσμα της εντολής str(MYDF2).

(ix) Να φτιάξετε το πλαίσιο δεδομένων MYDF2new, στο οποίο θα έχετε τα δεδομένα σας σε αλφαβητική σειρά (ως προς το όνομα κάθε φοιτητή/τριας).

Άσκηση 3 (Εισαγωγή – Εξαγωγή Δεδομένων).

(i) Τοποθετήστε το αρχείο EXAMPLElab.txt στον κατάλογο C:\LABexample (αν δεν υπάρχει, να τον δημιουργήσετε) και στη συνέχεια διαβάστε το με την εντολή read.table (προσοχή, η 1^η γραμμή του έχει τους τίτλους, headers, των μεταβλητών στο dataset).

(ii) Αποθηκεύστε το με την κατάλληλη εντολή σε μορφή csv και ανοίξτε το με το CALC.

(iii) Αποθηκεύστε το αρχείο με 3 διαφορετικούς τρόπους με τις εντολές write, dput και dump. Ποιες οι διαφορές μεταξύ των 3 αρχείων;

(iv) Διαβάστε ξανά στην R τα αρχεία του ερωτήματος (iii).

(v) Δημιουργήστε μια λίστα με το data frame και το κείμενο “This is the course for R programming”. Με ποιες εντολές μπορείτε να αποθηκεύσετε αυτό το αντικείμενο;

(vi) Αποθηκεύστε όλα τα αντικείμενα του RData χρησιμοποιώντας την εντολή dump και ξαναδιαβάστε τα σε έναν άδειο χώρο εργασίας.

Άσκηση 4 (Λίστες): Εκτελέστε τις παρακάτω εντολές, σχετικές με την οργάνωση και διαχείριση Λιστών στην R.

(i) Έστω τα διανύσματα

```
Year1<-c(1980:2020)
```

```
Month1<-1:12
```

```
Day1<-1:31
```

Να φτιάξετε μια λίστα που θα έχει τα παραπάνω διανύσματα ως στοιχεία με την ονομασία Date1.

(ii) Αντικαταστήστε το αντικείμενο Year1 με τις χρονιές 1976-2020

(iii) Να ορίσετε στην R τον πίνακα $\mathbf{A}_1 = \begin{bmatrix} 6 & 11 & 0.75 & 1 \\ 3 & 17 & -0.2 & -1 \\ -5 & 23 & 0.15 & 5 \\ 0 & 34 & 1.33 & -9 \end{bmatrix}$. Στη συνέχεια να φτιάξετε μια

λίστα με όνομα L2, η οποία θα περιέχει τα αποτελέσματα των πράξεων $3 \cdot A_1$, $A_1/4$, $\exp(A_1)$, με ονόματα 'triaA1', 'A1tetarta', 'ekthA1' καθώς και τα αντικείμενα της λίστας Date1.

(iv) Από τη λίστα Date1 αφαιρέστε την τιμή 6 από το διάνυσμα Month1.

(v) Από τη λίστα L2 αφαιρέστε το διάνυσμα Month1, αφού πρώτα το κρατήσετε ως ένα ξεχωριστό διάνυσμα.

(vi) Φτιάξτε μια λίστα με όνομα L3 η οποία θα περιέχει τις λίστες Date1 και L2.

(vii) Στη λίστα L3 προσθέστε ως 3^ο στοιχείο τον πίνακα A_1^2 , όπου ο A_1 έχει οριστεί στο (iii).

(viii) Να ορίσετε τη νέα λίστα ως L3 και να επιβεβαιώσετε ότι γνωρίζετε τι θα σας δώσουν οι εντολές

```
> L3[[3]]
```

```
>L3[[2]][[1]]
```

```
>L3[[2]][[2]]
```

```
>L3[[2]][[3]]
```

```
> array(c(L3[[3]],L3[[2]][[1]],L3[[2]][[2]],L3[[2]][[3]]),dim=c(4,4,4))
```

(ix) Φτιάξτε ένα διάνυσμα που θα έχει τα στοιχεία 2-10 του Year1 και τα στοιχεία 4, 6, 8 του πίνακα $\exp(A_1)$, όπως έχουν αποθηκευτεί στη λίστα L2.