

## ΜΑΘΗΜΑ: «Πληροφορική με Εφαρμογές Στατιστικής» Περιγραφική Στατιστική με την R

Για τις παρακάτω ασκήσεις να χρησιμοποιήσετε την R

### ΑΣΚΗΣΕΙΣ

#### Κατασκευή Πλαισίου Δεδομένων στην R

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

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

1. 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 κώδικάς 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)
```

2. 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.

3. 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.

4. 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;

5. Χρησιμοποιήστε την εντολή write.table για να αποθηκεύσετε το data frame που δημιουργήσατε ως ένα αρχείο txt με όνομα dataLab10.txt (βεβαιωθείτε ότι γνωρίζετε που ακριβώς έχει αποθηκευτεί).

6. Κλείστε την R, αφού έχετε αποθηκεύσει το data frame που φτιάξατε. Στη συνέχεια, ανοίξτε την R και φορτώστε το (με κατάλληλο τρόπο) από το αρχείο txt. **Υπόδειξη:** Χρησιμοποιήστε την εντολή read.table και εισάγετε τα δεδομένα με τον τίτλο datafile1. Στη συνέχεια με την εντολή attach φορτώστε τα δεδομένα στην R ώστε να είναι έτοιμα για επεξεργασία

7. Από τις μεταβλητές στο πλαίσιο δεδομένων, ποιες είναι ποσοτικές και ποιες ποιοτικές;

8. Να συμπληρώσετε τον παρακάτω πίνακα

Στατιστικό Περιγραφικό Μέτρο	Χοληστερίνη	Βάρος	Πίεση
Μέση τιμή			
Τυπική απόκλιση			
1ο τεταρτημόριο			
Διάμεσος			
3ο τεταρτημόριο			
Εύρος			
Ενδοτεταρτημοριακό εύρος			
Περιοκομμένος μέσος 10%			
80% ποσοστιαίο σημείο			
35% ποσοστιαίο σημείο			
Συντελεστής γραμμικής συσχέτισης			

9. Εξηγήστε το αποτέλεσμα που σας δίνει η εντολή  
> summary(datafile1)

10. Τι ποσοστό των ατόμων του δείγματος, έχει βάρος μικρότερο από 70kg; Τι ποσοστό έχει ύψος μεγαλύτερο από 175cm; Σε κάθε μια από τις παραπάνω υποομάδες, πόσοι είναι άνδρες, πόσοι γυναίκες;  
**Υπόδειξη:** Να χρησιμοποιήσετε τις εντολές which και length.