

## תרגיל 1- היכרות עם תוכנת R

לצורך התרגיל הבא, יש להוריד ולהתקין את תוכנת R מהאתר <http://cran.r-project.org> ניתן לגשת ישירות לכתובת <http://cran.r-project.org/bin/windows/base> ולקרוא את קובץ readme.r לפני ההתקנה.

מעבר לעזרה שבתוכנה עצמה, מומלץ להיעזר באתר: <http://wiki.r-project.org/rwiki/doku.php> הדרכה בעברית ניתן למצוא באתר הקורס. ניתן ורצוי להתשמש בפורום הקורס להתייעצות. יש לצרף את הקוד ששימש לפתרון. סעיפים בדרגת קושי גבוהה יותר סומנו בכוכבית(\*). כדאי להשתדל לענות עליהם אבל אין הכרח.

---

1. תאר במילים שלך (לא חייבים הרבה מילים) את סוג האובייקטים הבאים:

matrix(a)  
data frame(b)  
vector(c)  
list(d)  
function(e)  
formula(f)  
expression(g)

**הערה:** בשביל לקבל עזרה ב R ניתן להשתמש בפקודות:

```
>help('something')  
>?something  
>help.search('something')
```

2. שימרו במשתנה<sup>1</sup> ex01q2 את הנתונים שב:

<http://www-stat.stanford.edu/~tibs/ElemStatLearn/datasets/ozone.data>

צריך להיות מחובר בשעת הפעלת הפקודה היות שהיא קוראת ישירות מהאינטרנט.

המתג header של הפקודה read.table יטפל בשורת הכותרת...

להלן תיאור הנתונים:

---

1 הסבר על ייבוא קבצים ניתן למצוא כאן: <http://cran.r-project.org/doc/manuals/R-data.html#Imports>

## מבוא לסטטיסטיקה- תשס"ח

Data frame with components: ozone, radiation, temperature, and wind.  
Measurements of daily ozone concentration (ppb), wind speed (mph),  
daily maximum temperature (degrees F), and solar radiation  
(langleys) on 111 days from May to September 1973 in New York.

- (a) כמה משתנים יש בקובץ? השתמשו בפקודה `dim`  
(b) מהם שמות המשתנים בקובץ? השתמשו בפקודה `names`  
(c) מאיזה סוג כל אחד מהמשתנים? השתמשו בפקודה `class`  
שימו לב שניתן לקרוא לעמודות/משתנים ספציפיים במספר דרכים (שימושי גם לסעיפים קודמים):

```
>ex01q2['V2']  
>ex01q2$V2  
>ex01q2[,2]  
>subset(ex01q2,select='V2')
```

- (d) הסבירו את הפלט של הפקודה `summary` המופעלת על קובץ הנתונים:

```
>summary(ex01q2)
```

- (e) שרטטו תרשים של רמת האוזון כנגד רמת הקרינה. השתמשו בפקודה `plot`.  
(f) הוסיפו כותרות לתרשים ולצירים. השתמשו במתגים `ylab,xlab,main` של הפקודה `plot`.  
(g) צרו משתנה בשם `temp.celcius` ובו הטמפרטורות בצלזיוס (הן כרגע בפרנהייט).  
$$T_c = (5/9) * (T_f - 32)$$
  
`Tc` = temperature in degrees Celsius, `Tf` = temperature in degrees Fahrenheit  
(h) מהי הטמפרטורה המינימלית והמקסימלית? השתמשו בפקודות `min, max`.

3. צרו את הוקטורים הבאים:

(a) `(1,5,4,6,8,5)`. השתמשו בפקודה `c`.

(b) `(11,11,11,11,11,11,11,11,11)`. השתמשו בפקודה `rep`.

(c) `(1,3,5,7,9,11,13,15)`. השתמשו בפקודה `seq`.

4. צרו את המשתנה `x` על ידי הפקודה:

```
>x=seq(from=23,to=99.6,by=0.3)  
>y=matrix(x,ncol=16)
```

- (a) מהי הספרה במקום ה-`[12,3]` במטריצה `y`?  
(b) מהו סכום איברי הוקטור `x`? השתמשו בפקודה `sum`

## מבוא לסטטיסטיקה- תשס"ח

(c) מהו סכום איברי המטריצה  $y$ ? (האם זה מפתיע?)

(d) מהו סכום הריבועים של איברי  $x$ ?

5. ייבאו את קובץ הנתונים הבא לתוך המשתנה `ex01q5`

```
>read.table('http://www-stat.stanford.edu/~tibs/ElemStatLearn/datasets/marketing.data')
```

(a) כמה משתנים יש בקובץ? כמה תצפיות?  
השתמשו בפקודה `dim`

(b) מהי הספרה במקום ה- [7015,13] בקובץ הנתונים?

(c) כמה תאים חסרים בקובץ הנתונים? השתמשו בפקודות `table` ו-`is.na`

(d) מהו הממוצע של משתנה 4 עבור כל התצפיות בהן המשתנה השני שווה 2? השתמשו בפקודות `mean`, `subset` או לחלופין ב-`tapply`

(e) מהו סכום המשתנה השמיני עבור כל התצפיות בהן המשתנה השני שווה 1?  
השתמשו בפקודה `subset`, `sum`

6. מה מאפיין את סוגי הקבצים הבאים ומהם המתגים המתאימים של הפקודה `read.table` לייבא אותם?

- a) tab delimited values
- b) comma separated values (csv)
- c) xls