

Base R

Σύντομος Οδηγός

Βοήθεια

Πρόσβαση στη Βοήθεια

?mean

Εύρεση βοήθειας για μια συγκεκριμένη συνάρτηση.

help.search('weighted mean')

Εύρεση βοήθειας για μια λέξη ή φράση.

help(package = 'dplyr')

Εύρεση βοήθειας για μια βιβλιοθήκη/ πακέτο.

Πληροφορίες για αντικείμενα

str(iris)

Συνοπτικά η δομή ενός αντικειμένου.

class(iris)

Εύρεση κλάσης αντικειμένου.

Βιβλιοθήκες/πακέτα

install.packages('dplyr')

Λήψη και εγκατάσταση βιβλιοθηκών από το CRAN.

library(dplyr)

Φόρτωση βιβλιοθήκης/πακέτου, για να χρησιμοποιήσετε όλες τις συναρτήσεις της.

dplyr::select

Κλήση συγκεκριμένης συνάρτησης από μια βιβλιοθήκη.

data(iris)

Εισαγωγή ενός να ενσωματωμένου συνόλου δεδομένων στο περιβάλλον.

Κατάλογος Εργασίας - Working Directory

getwd()

Εύρεση του τρέχοντος καταλόγου εργασίας (ένας χώρος για την εισαγωγή αρχείων ή/και την αποθήκευσή τους).

setwd('C://file/path')

Αλλαγή του τρέχοντος καταλόγου εργασίας.

Μπορείτε να χρησιμοποιήσετε τα “projects” του RStudio για να ορίσετε τον κατάλογο εργασίας στον φάκελο που εργάζεστε.

Διανύσματα

Δημιουργία Διανυσμάτων

| | | |
|-------------------|-------------|--|
| c(2, 4, 6) | 2 4 6 | Παράθεση τιμών/ δημιουργία διανύσματος |
| 2:6 | 2 3 4 5 6 | Ακέραια ακολουθία αριθμών |
| seq(2, 3, by=0.5) | 2.0 2.5 3.0 | Ακολουθία αριθμών με βήμα (by=0.5) |
| rep(1:2, times=3) | 1 2 1 2 1 2 | Επανάληψη διανύσματος |
| rep(1:2, each=3) | 1 1 1 2 2 2 | Επανάληψη κάθε στοιχείου του διανύσματος |

Συναρτήσεις Διανυσμάτων

sort(x)

Ταξινόμηση x.

table(x)

Πίνακας συχνοτήτων.

rev(x)

Αντιστροφή στοιχείων του x.

unique(x)

Μοναδικές τιμές του x.

Επιλογή στοιχείων διανύσματος

Ορίζοντας τη θέση

| | |
|------------|---|
| x[4] | Το τέταρτο στοιχείο |
| x[-4] | Όλα εκτός από το τέταρτο στοιχείο |
| x[2:4] | Από το δεύτερο έως το τέταρτο στοιχείο |
| x[-(2:4)] | Όλα τα στοιχεία εκτός από το δεύτερο έως το τέταρτο |
| x[c(1, 5)] | Πρώτο και πέμπτο στοιχείο |

Ορίζοντας συνθήκη

| | |
|----------------------|--|
| x[x == 10] | Τα στοιχεία του x που είναι ίσα με 10 |
| x[x < 0] | Τα στοιχεία του x μικρότερα του μηδενός |
| x[x %in% c(1, 2, 5)] | Τα στοιχεία εντός του συνόλου {1, 2, 5}. |

Διανύσματα με όνομα

| | |
|------------|-------------------------------|
| x['apple'] | Στοιχείο με το όνομα 'apple'. |
|------------|-------------------------------|

Προγραμματισμός στην R

Εντολή for

```
for (μεταβλητή in διάστημα){  
  Εντολή/ές  
}
```

Παράδειγμα

```
for (i in 1:4){  
  j <- i + 10  
  print(j)  
}
```

Εντολή while

```
while (συνθήκη){  
  Εντολή/ές  
}
```

Παράδειγμα

```
while (i < 5){  
  print(i)  
  i <- i + 1  
}
```

Έλεγχος Ροής if

```
if (συνθήκη){  
  Εντολή/ές (αν πληρείται η συνθήκη)  
} else {  
  Εντολή/ές (για την εναλλακτική περίπτωση)  
}
```

Παράδειγμα

```
if (i > 3){  
  print('Yes')  
} else {  
  print('No')  
}
```

Συναρτήσεις

```
function_name <- function(var){  
  Εντολή/ές  
  return(new_variable)  
}
```

Παράδειγμα

```
square <- function(x){  
  squared <- x*x  
  return(squared)  
}
```

Εισαγωγή και Εξαγωγή Δεδομένων

Επιπλέον δείτε το πακέτο **readr**.

| Εισαγωγή | Εξαγωγή | Περιγραφή |
|------------------------------|-------------------------------|---|
| df <- read.table('file.txt') | write.table(df, 'file.txt') | Εισαγωγή και εξαγωγή οριοθετημένου αρχείου κειμένου |
| df <- read.csv('file.csv') | write.csv(df, 'file.csv') | Εισαγωγή και εξαγωγή αρχείων που περιέχουν τιμές διαχωρισμένες με κόμματα |
| load('file.RData') | save(df, file = 'file.Rdata') | Εισαγωγή και εξαγωγή αρχείων δεδομένων R, ένας ειδικός τύπος αρχείων της R. |

| Συνθήκες | a == b | Ισότητα | a > b | Μεγαλύτερο από | a >= b | Μεγαλύτερο ή ίσο από | is.na(a) | Έλεγχος απούσας τιμής |
|----------|--------|-----------|-------|----------------|--------|----------------------|------------|-----------------------|
| | a != b | Ανισότητα | a < b | Μικρότερο από | a <= b | Μικρότερο ή ίσο από | is.null(a) | Έλεγχος έγκυρης τιμής |

Τύποι μεταβλητών

Μετατροπή ενός τύπου δεδομένων σε άλλον.

| | | |
|--------------|------------------------------------|--|
| as.logical | TRUE, FALSE, TRUE | Λογική Τιμή (Boolean-TRUE/FALSE) |
| as.numeric | 1, 0.1 | Ακέραιοι/ δεκαδικοί αριθμοί |
| as.character | '1', '0', '1' | Ακολουθία χαρακτήρων. |
| as.factor | '1', '0', '1', levels: '1', '0' | Ακολουθία χαρακτήρων με προκαθορισμένα επίπεδα. Απαραίτητοι για κάποια στατιστικά μοντέλα. |

Μαθηματικές Συναρτήσεις

| | | | |
|--------------|---------------------------------------|-------------|------------------|
| log(x) | Φυσικός λογάριθμος. | sum(x) | Άθροισμα. |
| exp(x) | Εκθετική συνάρτηση. | mean(x) | Μέση τιμή. |
| max(x) | Μέγιστη τιμή. | median(x) | Διάμεσος. |
| min(x) | Ελάχιστη τιμή. | quantile(x) | Ποσοστημόρια. |
| round(x, n) | Στρογγυλοποίηση σε n δεκαδικά ψηφία. | rank(x) | Τάξη στοιχείων. |
| signif(x, n) | Στρογγυλοποίηση σε n σημαντικά ψηφία. | var(x) | Διακύμανση. |
| cor(x, y) | Συσχέτιση. | sd(x) | Τυπική Απόκλιση. |

Ανάθεση Τιμής

```
> a <- 'apple'
> a
[1] 'apple'
```




Περιβάλλον

| | |
|-----------------|---|
| ls() | Κατάλογος των αποθηκευμένων μεταβλητών στο περιβάλλον |
| rm(x) | Διαγραφή της μεταβλητής x από το περιβάλλον |
| rm(list = ls()) | Διαγραφή όλων των αποθηκευμένων μεταβλητών από το περιβάλλον. |

Μπορείτε να χρησιμοποιήσετε το παράθυρο με το περιβάλλον (environment) του Rstudio για να δείτε τις αποθηκευμένες τις μεταβλητές.

Πίνακες

```
m <- matrix(x, nrow = 3, ncol = 3)
Δημιουργία πίνακα.
```

| | |
|---|--|
|  <code>m[2,]</code> – Επιλογή γραμμής | <code>t(m)</code> Ανάστροφος |
|  <code>m[, 1]</code> – Επιλογή στήλης | <code>m %*% n</code> Πολλαπλασιασμός πινάκων |
|  <code>m[2, 3]</code> – Επιλογή στοιχείου | <code>solve(m, n)</code> Επίλυση ως προς x: $m \cdot x = n$ |

Λίστες

```
l <- list(x = 1:5, y = c('a', 'b'))
Η λίστα είναι μια συλλογή αντικειμένων διαφορετικού τύπου.
```

| | | | |
|---|---|--|--|
| <code>l[[2]]</code> Το 2 ^ο στοιχείο της λίστας l. | <code>l[1]</code> Νέα λίστα μόνο με το 1 ^ο στοιχείο της λίστας l. | <code>l\$x</code> Επιλογή στοιχείου με όνομα x. | <code>l['y']</code> Νέα λίστα μόνο με το στοιχείο 'y' της λίστας l. |
|---|---|--|--|




Επιπλέον δείτε το πακέτο **dplyr**.

Πλαίσια Δεδομένων

```
df <- data.frame(x = 1:3, y = c('a', 'b', 'c'))
Τα πλαίσια δεδομένων είναι μια ειδική περίπτωση τύπου λίστας, όπου τα επιμέρους στοιχεία τους είναι ίδιου μήκους.
```

| x | y |
|---|---|
| 1 | a |
| 2 | b |
| 3 | c |

Υποσύνολο Πίνακα

| | | |
|---|---|--|
| <code>df[, 2]</code>  | <code>df[2,]</code>  | <code>df[2, 2]</code>  |
|---|---|--|

Επιλογή υποσυνόλου

| | |
|---|---|
| <code>df\$x</code>  | <code>df[[2]]</code>  |
|---|---|

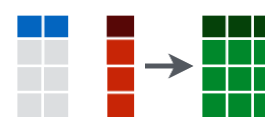
| Κατανόηση πλαισίου δεδομένων | |
|------------------------------|-----------------------------|
| <code>View(df)</code> | Προβολή πλαισίου δεδομένων. |
| <code>head(df)</code> | Προβολή 6 πρώτων γραμμών. |

`nrow(df)`
Αριθμός γραμμών.

`ncol(df)`
Αριθμός στηλών.

`dim(df)`
Αριθμός στηλών και γραμμών.

cbind – Σύνδεση κατά στήλες



rbind – Σύνδεση κατά γραμμές



Συμβολοσειρές - Strings

`paste(x, y, sep = ' ')`

`paste(x, collapse = ' ')`

`grep(pattern, x)`

`gsub(pattern, replace, x)`

`toupper(x)`

`tolower(x)`

`nchar(x)`

Ποιοτικές Μεταβλητές - Factors

`factor(x)`

Ορισμός μιας μεταβλητής σε ποιοτική κατηγορική ή σε ποιοτική διάταξη.

`cut(x, breaks = 4)`

Μετατροπή μιας αριθμητικής μεταβλητής σε factor, χωρίζοντας της σε διαστήματα/κλάσεις.

Δείτε ακόμα το πακέτο **stringr**.

Σύνδεση πολλών διανυσμάτων.

Σύνδεση των επιμέρους στοιχείων ενός διανύσματος.

Εύρεση συμβολοσειράς κειμένου σε ένα διάνυσμα x βάσει κανονικής έκφρασης (regular expression).

Αντικατάσταση μιας συμβολοσειράς κειμένου με άλλη σε ένα διάνυσμα x.

Μετατροπή κειμένου σε κεφαλαία γράμματα.

Μετατροπή κειμένου σε πεζά γράμματα.

Μετρά το πλήθος χαρακτήρων σε μια ακολουθία χαρακτήρων.

Στατιστική

`lm(y ~ x, data=df)`
Γραμμικό μοντέλο.

`glm(y ~ x, data=df)`
Γενικευμένο γραμμικό μοντέλο.

`summary`
Περισσότερες πληροφορίες για ένα μοντέλο.

`t.test(x, y)` Έλεγχος υπόθεσης για τη διαφορά $\mu_1 - \mu_2$ των μέσων τιμών δύο πληθυσμών.

`prop.test` Έλεγχος υπόθεσης για τη διαφορά $p_1 - p_2$ των αναλογιών δύο πληθυσμών.

`pairwise.t.test` Έλεγχος υπόθεσης για τη διαφορά $\mu_1 - \mu_2$ των μέσων τιμών δύο πληθυσμών για εξαρτημένα δείγματα- ζευγαρωτές παρατηρήσεις.

`aov`
Ανάλυση Διασποράς

Κατανομές

| | Τυχαίοι αριθμοί | Πυκνότητα πιθανότητας | Αθροιστική πιθανότητα | Ποσοστιαία σημεία |
|------------|-----------------|-----------------------|-----------------------|-------------------|
| Κανονική | rnorm | dnorm | pnorm | qnorm |
| Poisson | rpois | dpois | ppois | qpois |
| Διωνυμική | rbinom | dbinom | pbinom | qbinom |
| Ομοιόμορφη | runif | dunif | punif | qunif |

Γραφήματα

Δείτε επίσης το πακέτο **ggplot2**.

`plot(x)`
Γράφημα τιμών του x σε σειρά.

`plot(x, y)`
Διάγραμμα διασποράς x και y.

`hist(x)`
Ιστόγραμμα του x.

Ημερομηνίες

Δείτε περισσότερα στο πακέτο **lubridate**.