Κεφάλαιο 3

Μαθηματικοί Υπολογισμοί στην R

Ένα μεγάλο μέρος της ανάλυσης δεδομένων απαιτεί διάφορους μαθηματικούς υπολογισμούς. Αυτό το κεφάλαιο εισαγάγει τον αναγνώστη στις διάφορες δυνατότητες που έχει η R για να γίνουν τέτοιοι υπολογισμοί. Οι υπολογιστικές δυνατότητες της R αρχίζουν από απλές πράξεις μέχρι και πολύπλοκους μαθηματικούς υπολογισμούς, όπως π.χ. η μεγιστοποίηση συναρτήσεων.

3.1 Αριθμητικές πράξεις και απλές συναρτήσεις

Οι βασικές αριθμητικές πράξεις γίνονται με τη βοήθεια των συμβόλων που βρίσκονται στον ακόλουθο πίνακα.

Σύμβολο	**Πράξη**
+	Πρόσθεση
-	Αφαίρεση
*	Πολλαπλασιασμός
/	Διαίρεση
^	Ύψωση σε δύναμη
%/%	Ακέραια διαίρεση
%%	Υπόλοιπο διαίρεσης

Πίνακας 3.1: Βασικά αριθμητικά σύμβολα.

Ακολουθούν μερικά παραδείγματα τα οποία καταδεικνύουν πώς χρησιμοποιούνται τα σύμβολα των βασικών αριθμητικών πράξεων.

```
> 7+3
[1] 10
> 15-19
[1] -4
> 4*67
[1] 268
> 56/9
[1] 6.222222
> 2^6
[1] 64
> 27%/%3.4
[1] 7
> 27%%3.4
[1] 3.2
> 7*3.4+3.2
[1] 27
```

Το σύμβολο ^ είναι χρήσιμο όχι μόνο για ύψωση σε δύναμη αλλά και υπολογισμό ριζών.

```
> 16^(1/2)
[1] 4
> 2^(1/3)
[1] 1.259921
```

Αυτές οι εντολές χρησιμοποιούνται όχι μόνο με αριθμούς αλλά και με διανύσματα και πίνακες. Το επόμενο παράδειγμα δείχνει πως λειτουργούν σε αυτές τις περιπτώσεις.

```
> x <- c(1,4,7)
> y <- c(2,4,6,4,6,10)
> A <- matrix(c(2,3,4,5,6,7,1,2,3), nrow=3)
> A
     [,1] [,2] [,3]
[1,]    2    5    1
[2,]    3    6    2
```

```
[3,]    4    7    3
> B <- rbind(c(0,0,1), c(2,4,5), c(1,4,2))
> B
     [,1] [,2] [,3]
[1,]    0    0    1
[2,]    2    4    5
[3,]    1    4    2
> A*B
     [,1] [,2] [,3]
[1,]    0    0    1
[2,]    6   24   10
[3,]    4   28    6
> x+y
[1]  3  8 13  5 10 17
> A/y
          [,1] [,2] [,3]
[1,] 1.0000000 1.25  0.5
[2,] 0.7500000 1.00  0.5
[3,] 0.6666667 0.70  0.5
Warning messages:
  Length of longer object is not a multiple
  of the length of the shorter object in: A/y
```

Οι περισσότεροι υπολογισμοί με διανύσματα και πίνακες γίνονται *κατά στοιχείο*, υποθέτωντας ότι οι πίνακες έχουν τις ίδιες διαστάσεις. Στις πράξεις με διανύσματα, αν το ένα διάνυσμα είναι μικρότερης διάστασης από το άλλο, τότε τα στοιχεία του μικρότερου διανύσματος επαναλαμβάνονται κυκλικά έτσι ώστε τα δύο διανύσματα να έχουν στο τέλος ίσες διαστάσεις. Μαθηματικοί υπολογισμοί μεταξύ διανυσμάτων και πινάκων δεν έχουν συνήθως τα αναμενόμενα αποτελέσματα και θα πρέπει να χρησιμοποιούνται με μεγάλη προσοχή.

Ακολουθούν μερίκες γνωστές συναρτήσεις οι οποίες είναι εγκατεστημένες μέσα στην R και μπορούν να κληθούν ανά πάσα στιγμή από τον χρήστη. Οι συναρτήσεις αυτές υπολογίζονται κατά στοιχείο.

Συνάρτηση	**Πράξη**
sqrt()	Τετραγωνική ρίζα
abs()	Απόλυτη τιμή
floor()	Προηγούμενος ακέραιος
ceiling()	Επόμενος ακέραιος
sin()	Ημίτονο
cos()	Συνημίτονο
tan()	Εφαπτωμένη
asin()	Τόξο ημιτόνου
acos()	Τόξο συνημιτόνου
atan()	Τόξο εφαπτωμένης
exp()	Εκθετική συνάρτηση
log()	Λογάριθμος
log10()	Λογάριθμος με βάση το 10
gamma()	Συνάρτηση Γάμμα
lgamma()	Φυσικός λογάριθμος της απόλυτης τιμής της συνάρτηση Γάμμα

Πίνακας 3.2: Αριθμητικές συναρτήσεις.

Τα επόμενα παραδείγματα χρησιμοποιούν μερικές από αυτές τις συναρτήσεις.

```
> abs(-10.56)
[1] 10.56
> floor(5.6)
[1] 5
> ceiling(5.6)
[1] 6
> log(x)
[1] 0.000000 1.386294 1.945910
> log(x, base=2)   #logaritm to base 2
[1] 0.000000 2.000000 2.807355
> cos(A)
           [,1]      [,2]       [,3]
[1,] -0.4161468 0.2836622  0.5403023
[2,] -0.9899925 0.9601703 -0.4161468
[3,] -0.6536436 0.7539023 -0.9899925
> atan(A)
```

```
         [,1]     [,2]      [,3]
[1,] 1.107149 1.373401 0.7853982
[2,] 1.249046 1.405648 1.1071487
[3,] 1.325818 1.428899 1.2490458
> exp(y)
[1]     7.389056    54.598150   403.428793    54.598150   403.428793 22026.465795
```

3.2 Πράξεις Διανυσμάτων και Πινάκων

Όπως αναφέρθηκε πιο πάνω, στις περιπτώσεις των διανυσμάτων οι διάφορες αριθμητικές πράξεις εφαρμόζονται σε κάθε στοιχείο τους. Σε αυτό το σημείο θα γίνει αναφορά στο πώς εκτελούνται διάφοροι υπολογισμοί με διανύσματα ή πίνακες. Ο επόμενος πίνακας δίνει σύμβολα και συναρτήσεις για αυτές τις πράξεις.

Σύμβολα - Συνάρτηση	**Πράξη**
%*%	Εσωτερικό γινόμενο διανυσμάτων ή πολλαπλασιασμός πινάκων
t()	Ανάστροφος πίνακα
solve()	Αντίστροφος πίνακα (αν υπάρχει)
diag()	Εξαγωγή της διαγωνίου αλλά και κατασκευή διαγώνιου πίνακα
eigen()	Ιδιοτιμές και ιδιοδιανύσματα πίνακα

Πίνακας 3.3: Πράξεις διανυσμάτων και πινάκων.

Ακολουθούν μερικά παραδείγματα.

```
> A%*%B   #matrix multiplication
     [,1] [,2] [,3]
[1,]   11   24   29
[2,]   14   32   37
[3,]   17   40   45
> z <- c(2,3,1)
> z%*%x   #vector dot product
     [,1]
[1,]   21
> t(A)    # transpose of a matrix
     [,1] [,2] [,3]
```

```
[1,]    2    3    4
[2,]    5    6    7
[3,]    1    2    3
> diag(A)   # extract the diagonal
[1] 2 6 3
> sum(diag(A))   # trace of a matrix
[1] 11
> X <- diag(c(1,2,3,4)) # create a diagonal matrix
> X
     [,1] [,2] [,3] [,4]
[1,]    1    0    0    0
[2,]    0    2    0    0
[3,]    0    0    3    0
[4,]    0    0    0    4
> I <-  diag(4) # create an identity matrix
> I
     [,1] [,2] [,3] [,4]
[1,]    1    0    0    0
[2,]    0    1    0    0
[3,]    0    0    1    0
[4,]    0    0    0    1
> solve(B)
      [,1]  [,2] [,3]
[1,] -3.00  1.00 -1.0
[2,]  0.25 -0.25  0.5
[3,]  1.00  0.00  0.0
> eigen(A)    # compute eigenvalues and eigenvectors of a matrix
$values:
[1]  1.072015e+001  2.798467e-001 -1.887379e-015
$vectors:
           [,1]      [,2]       [,3]
[1,] -0.4902022 -2.332769 -0.7817656
[2,] -0.6806916  0.239993  0.1954414
[3,] -0.8711809  2.812755  0.5863242
> prod(eigen(A)$values)   # determinant
[1] -5.662137e-015
```

Επίσης μπορούν να χρησιμοποιηθούν οι συναρτήσεις `kronecker` (για το Kronecker γινόμενο δύο πινάκων), `qr` (για ανάλυση QR), `svd` (για ανάλυση ιδιάζουσας τιμής) και `chol` (για ανάλυση Choleski).

3.3 Γραμμικό Σύστημα Εξισώσεων

Η εντολή `solve` δε χρησιμεύει μόνο στον υπολογισμό του αντίστροφου ενός πίνακα, αλλά και στην επίλυση ενός γραμμικού συστήματος εξισώσεων της μορφής $Ax = y$, με την προϋπόθεση ότι υπάρχει λύση. Για παράδειγμα, έστω το γραμμικό σύστημα 2 εξισώσεων και 2 αγνώστους,

$$\begin{aligned} 2x + 3y &= 13 \\ x - 2y &= -4 \end{aligned}$$

Για να λυθεί αυτό το σύστημα, πρώτα κατασκευάζεται ο πίνακας A με τους συντελεστές των αγνώστων και μετά υπολογίζεται η λύση, θέτοντας σαν δεύτερο όρισμα το διάνυσμα των σταθερών όρων, όπως φαίνεται πιο κάτω.

```
> A <- rbind( c(2,3), c(1,-2))
> A
     [,1] [,2]
[1,]    2    3
[2,]    1   -2
> solve(A, c(13,-4))
[1] 2 3
> solve(A)  # getting the inverse
          [,1]       [,2]
[1,] 0.2857143  0.4285714
[2,] 0.1428571 -0.2857143
> solve(rbind(c(1,2), c(2,4))) # getting the inverse of a singular matrix
Error in solve.qr(a): apparently singular matrix
```

Περισσότερες συναρτήσεις σε σχέση με πράξεις πινάκων βρίσκονται μέσα στη βιβλιοθήκη της R, `Matrix`, η οποία καλείται με την εντολή `library(Matrix)`.

3.4 Τυχαίοι Αριθμοί

Στην R υπάρχουν πολλές συναρτήσεις για τη γέννηση τυχαίων αριθμών και υπολογισμών πιθανοτήτων σε σχέση με τις πιο γνωστές κατανομές πιθανοτήτων. Κάθε μια από αυτές τις συναρτήσεις έχει όνομα το οποίο αρχίζει με ένα από τα ακόλουθα τέσσερα γράμματα, τα οποία καθορίζουν το είδος της συνάρτησης.

r: Γεννήτρια τυχαίων αριθμών.

p: Συνάρτηση κατανομής ($F(x) = P[X \leq x]$).

d: Συνάρτηση πιθανότητας ($f(x)$).

q: Αντίστροφη συνάρτηση κατανομής ($F^{-1}(x)$).

Ο ακόλουθος πίνακας παρουσιάζει τις πιο σημαντικές συναρτήσεις κατανομών στην R και τα επόμενα παραδείγματα εξηγούν πως χρησιμοποιούνται αυτές οι συναρτήσεις.

beta	Κατανομή Βήτα
binom	Διωνυμική Κατανομή
chisq	X^2 Κατανομή
gamma	Κατανομή Γάμμα
lnorm	Κατανομή Lognormal
norm	Κανονική Κατανομή
pois	Κατανομή Poisson
t	Κατανομή t
unif	Ομοιόμορφη Κατανομή

Πίνακας 3.4: Κατανομές τυχαίων μεταβλητών.

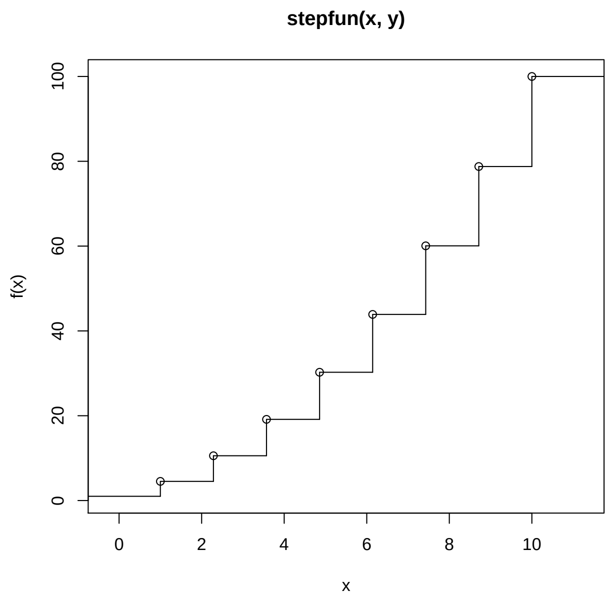
```
> x
[1] 1 4 7
> pnorm(x)
[1] 0.8413447 0.9999683 1.0000000
> pnorm(x, mean=2, sd=2)
[1] 0.3085375 0.8413447 0.9937903
> dnorm(x)
[1] 2.419707e-001 1.338302e-004 9.134720e-012
> qchisq(c(0.90,0.95,0.99), 2)
[1] 4.605170 5.991465 9.210340
> runif(30, -10, 10)
 [1]  9.213183280  8.749200171 -9.117961340  5.292370273  4.117153846
```

```
 [6]  0.071010422  8.572964445  6.805462409  0.942033436 -0.243897894
[11] -2.020305358 -4.729607552  8.518492579 -1.429708684  9.201227576
[16] -4.033246860  1.544312881 -0.227093771 -6.805263087 -6.349458788
[21] -5.736339493 -4.680284206  4.654475767  2.877361933  7.949797967
[26] -0.003705453  1.538872020  8.116273358 -9.711499065  4.931013034
```

3.5 Άλλες Χρήσιμες Συναρτήσεις

Στην R υπάρχουν και άλλες πολλές συναρτήσεις οι οποίες μπορούν να χρησιμοποιηθούν για υπολογισμούς αλλά δεν μπορούν όλες να επεξηγηθούν λεπτομερώς σε αυτό το κεφάλαιο. Αξίζει να αναφερθεί η συνάρτηση `integrate`, η οποία υπολογίζει το ολοκλήρωμα μιας πραγματικής συνάρτησης σε ένα διάστημα τιμών, η συνάρτηση `diff` η οποία επιστρέφει την n-οστή διαφορά με βήμα k για ένα σύνολο τιμών και η συνάρτηση `fft` η οποία δίνει τον γρήγορο μετασχηματισμό Fourier ενός συνόλου τιμών. Ακολουθεί ένα παράδειγμα με τη συνάρτηση `stepfun` η οποία υπολογίζει την αριστερή συνεχή συνάρτηση βήματος από σημεία (x, y).

```
> x <- seq(1,10, length=8)
> y <- seq(1,10,length=9)^2
> stepfun(x,y)
Step function
Call: stepfun(x, y)
 x[1:8] =      1, 2.2857, 3.5714,  ..., 8.7143,     10
9 plateau levels =      1, 4.5156, 10.562,  ..., 78.766,    100
> plot.stepfun(stepfun(x,y))
```



Σχήμα 3.1: Συνάρτηση βήματος