Elements of R

1 Arithmetic

The expressions +, −, *, / are used in the usual way. Exponents are indicated by expressions like 3 ^ 4, which evaluates to 81. There are various common functions that work like sqrt(9) and abs(−4).

2 Logic

Equality is expressed by ==. Lack of equality is ! =. The inequalities are <, <= and >, >=. The logical operations and, or, not are written &, |, !.

3 Vectors

A vector can be generated by c(5, 2, 4). This combines the numbers 5, 2, 4 to form a single vector. The vector 2:5 is the same as the vector c(2,3,4,5). The vector seq(2,5, 0.1) is the same as the vector 20:50/10.

4 Assignment

A variable is assigned a value by the command

variable <- expression

Thus, for instance

x <- c(5,2,4)

makes x stand for the corresponding vector. In this context we can say x “becomes” c(5,2,4).

5 Vector operations

If x is a vector, then length(x) tells how many components it has, and x[3] selects the third component.

The sum of the components is sum(x), and the mean is mean(x). This is the same as sum(x)/length(x). The variance var(x) is defined with the n − 1 factor in the denominator. The standard deviation is sd(x).
The largest and smallest elements of a vector are given by \( \text{max}(x) \) and \( \text{min}(x) \). The expression \( \text{sort}(x) \) gives a vector with the same entries, but sorted in increasing order. The expression \( \text{median}(x) \) gives the same result as \( \text{quantile}(x, 0.5) \). The quartiles can be obtained by \( \text{quantile}(x, 0.25, 0.5, 0.75) \).

With two vectors of the same length one can compute the correlation coefficient \( \text{cor}(x,y) \). The two vectors can be plotted by \( \text{plot}(x,y) \).

6 Functions

A function is denoted by giving inputs and an expression for an output. Thus
\[
(\text{function } (x) \ x \times (1 - x))
\]
denotes a function that takes input \( x \) and gives output \( x(1-x) \). If we wanted to give this function a name, such as \( h \), then we would make the assignment
\[
h \leftarrow \text{function } (x) \ x \times (1 - x).
\]
Thus \( h(2) \) would return \(-2\).

7 Probability distributions

For each probability distribution there are three functions and one random sample generator. Thus for the normal distribution these are:
\[
\text{dnorm}(x, \text{mean}, \text{sd}) \quad \text{density: computes density as a function of } x
\]
\[
\text{pnorm}(q, \text{mean}, \text{sd}) \quad \text{distribution: computes probability as a function of quantile } q
\]
\[
\text{qnorm}(p, \text{mean}, \text{sd}) \quad \text{inverse distribution: computes quantile as a function of probability } p
\]
\[
\text{rnorm}(n, \text{mean}, \text{sd}) \quad \text{generates random sample of size } n
\]
Similarly, for the binomial distribution there are the functions \( \text{dbinom}(x, \text{size}, \text{prob}) \), \( \text{pbinom}(q, \text{size}, \text{prob}) \), \( \text{qbinom}(p, \text{size}, \text{prob}) \), and \( \text{rbinom}(n, \text{size}, \text{prob}) \).

Here are some of the probability distributions that are commonly used. The following listing has the p version of the function, but the d,p,q,and r versions all exist.
\[
\text{pnorm}(q, \text{mean}, \text{sd}) \quad \text{normal distribution}
\]
\[
\text{pgamma}(q, \text{shape}, \text{rate}) \quad \text{Gamma distribution}
\]
\[
\text{pexp}(q, \text{rate}) \quad \text{exponential distribution: same as pgamma}(x,1,\text{rate})
\]
\[
\text{pchisq}(q, \text{df}) \quad \text{chi square distribution: same as pgamma}(x,\text{df}/2,1/2)
\]
\[
\text{pt}(q, \text{df}) \quad \text{t distribution}
\]
\[
\text{pf}(q, \text{df1}, \text{df2}) \quad \text{F distribution}
\]
\[
\text{pbeta}(q, \text{shape1}, \text{shape2}) \quad \text{Beta distribution}
\]
\[
\text{punif}(q, \text{min}, \text{max}) \quad \text{uniform distribution}
\]
\[
\text{pcauchy}(q, \text{location}, \text{scale}) \quad \text{Cauchy distribution}
\]
\[
\text{pbinom}(q, \text{size}, \text{prob}) \quad \text{binomial distribution}
\]
\[
\text{pnbinom}(q, \text{size}, \text{prob}) \quad \text{negative binomial distribution}
\]
\[
\text{pgeom}(q, \text{prob}) \quad \text{geometric distribution: same as pnbinom}(q,1,\text{prob})
\]
\[
\text{ppois}(q, \text{lambda}) \quad \text{Poisson distribution}
\]
8 Example: Empirical distribution

Take a sample; tabulate the results.
   Create a sample:
   \[ x \leftarrow \text{rbinom}(100,8,1/2) \]
   Create a vector:
   \[ n \leftarrow 0:8 \]
   Tabulate the sample:
   \[ \text{for}(i \text{ in } 1:9) \ n[i] \leftarrow \text{sum}\left[ x == i-1 \right] \]
   Plot the table:
   \[ \text{plot}(0:8, n) \]

9 Example: The Bernoulli process

Compare the number of successes up to \( n \) with the time of the \( i \)th success.
   Take an independent Bernoulli sample:
   \[ x \leftarrow \text{rbinom}(100,1,1/7) \]
   Create a vector:
   \[ s \leftarrow 1:100 \]
   Find the number of successes in the first \( n \) trials:
   \[ h \leftarrow 1:100 \]
   \[ \text{for}(n \text{ in } 1:100) \ s[n] \leftarrow \text{sum}\left( x[ h <= n] \right) \]
   Create another vector:
   \[ t \leftarrow 1:100 \]
   Find the time of the \( i \)th success:
   \[ \text{for}(i \text{ in } 1:100) \ t[i] \leftarrow \text{min}\left( h[ s >= i] \right) \]
   Extract the useful part of this vector:
   \[ tt \leftarrow t[1:13] \]

10 File input

To read in a vector:
   \[ x \leftarrow \text{scan(”filename.txt”)} \]
   To read in a list of two vectors:
   \[ xy \leftarrow \text{scan(”filename.txt”, list(0,0))} \]
   To extract the individual vectors:
   \[ x \leftarrow xy[1] \]
   \[ y \leftarrow xy[2] \]