

Introduction to R

Jasmine Sellers

2024-01-16

Establishing our Basics

What is programming?

Imagine you are a tutor helping a friend learn a new language. Your friend, though very smart, doesn't understand English, so you need to use a special language they do understand. This language is like a programming language, and your friend is the computer.

What is R?

R is a programming language designed for statistical computing and data analysis.

Terms that might show up

Compiler: converts the stuff you write to stuff machines can understand

Basic R Syntax

1. **Comments:** These are used to explain what is going on and ignored by the R interpreter. Comments are preceded the '#' symbol

```
# This is a comment
```

2. **Assignment:** You can use the '<-' or '=' symbol for setting variables equal to something.

```
x <- 5  
x = 5
```

3. **Print Output:** The 'print()' function is used to display the value of a variable.

```
print(x)
```

```
## [1] 5
```

4. **Data Types:**

1. *Numeric:* represent real numbers (integers and decimals)
2. *Character:* represent characters or strings (sequence of characters)

```
#character variable, you can use either ' ' or " " to surround the characters  
char_var = 'Hello World'  
char_var2 = "Hello World"
```

```
print(char_var)
```

```
## [1] "Hello World"
```

```
print(char_var2)
```

```
## [1] "Hello World"
```

3. *Logical*: represents boolean values (i.e. 'TRUE' or 'FALSE')

```
boolean_var = 5 > 6
```

```
print(boolean_var)
```

```
## [1] FALSE
```

4. *Vector*: one-dimensional array that can hold element of the SAME data type

```
# Numeric vector
```

```
numeric_vector <- c(1, 2, 3, 4, 5)
```

```
print(numeric_vector)
```

```
## [1] 1 2 3 4 5
```

```
# Character vector
```

```
char_vector <- c("apple", "grapes", "banana")
```

```
print(char_vector)
```

```
## [1] "apple" "grapes" "banana"
```

5. *Other Data Types (You can skip this :)*

- *Matrix*: a matrix is a 2D array with rows and columns

```
# Matrix
```

```
matrix_var <- matrix(c(1, 2, 3, 4, 5, 6), nrow = 2, ncol = 3)
```

```
print(matrix_var)
```

```
##      [,1] [,2] [,3]
```

```
## [1,]    1    3    5
```

```
## [2,]    2    4    6
```

- *Array*: multi-dimensional generalization of a matrix

```
# Creating a 3D array
```

```
array_var <- array(c(1, 2, 3, 4, 5, 6), dim = c(2, 2, 2))
```

```
# Displaying the array
```

```
print(array_var)
```

```
## , , 1
```

```
##
```

```
##      [,1] [,2]
```

```
## [1,]    1    3
```

```
## [2,]    2    4
```

```
##
```

```
## , , 2
```

```
##
```

```
##      [,1] [,2]
```

```
## [1,]    5    1
```

```
## [2,]    6    2
```

- *List*: collection of different data types (e.g. characters, vectors, numeric, etc.)

```
# List
list_var <- list(name = "John", age = 25, is_student = TRUE)

print(list_var)

## $name
## [1] "John"
##
## $age
## [1] 25
##
## $is_student
## [1] TRUE
```

- Data Frame: A data frame is a 2D dimensional table with rows and columns where each column can be a different data type

```
# Data frame
df_var <- data.frame(name = c("Alice", "Bob", "Charlie"), age = c(25, 30, 22))

print(df_var)

##      name age
## 1  Alice  25
## 2   Bob  30
## 3 Charlie  22
```

6. **Indexing:** R uses 1-base indexing meaning the index of the first element in a vector is 1.

```
numeric_vector[1] # Access the first element

## [1] 1

char_vector[2]    # Access the second element

## [1] "grapes"
```

Functions

functions are defined using the ‘function()’ keyword. The ‘return()’ statement is used to return a value from a function

```
# Example of a function that adds two numbers together
add_numbers <- function(a, b) {
  result <- a + b
  return(result)
}

# Call the function
sum_result <- add_numbers(3, 5)
print(sum_result)

## [1] 8
```

Using R for Statistics

Descriptive Statistics

Statistics Concept	Function in R	parameters
Mean	mean(x)	x: vector containing data of interest
Median	median(x)	x: vector containing data of interest
sample standard deviation	sd(x)	x: vector containing data of interest
sample variation	var(x)	x: vector containing data of interest
range	range(x)	x: vector containing data of interest
Five-number summary	summary(x)	x: vector containing data of interest

Data Visualization

Statistics Concept	Function in R	parameters
histogram	hist(x)	x: vector containing data of interest
box-and-whisker plot	boxplot(x)	x: vector containing data of interest x: vector containing data of interest

Probability Distribution

Statistics Concept	Function in R	Output
Normal distribution	pnorm(x, mean = mean, sd = standard deviation)	probability of a value in the sample being less than or equal to x
Binomial Distribution	dbinom() pbinom()	probability of a value being equal to x
Geometric Distribution	dgeom() pgeom()	

Hypothesis Testing

Statistics Concept	Function in R
z-test	z.test(data, mu = population mean, sigma = population sd)
t-test	t.test(data)

Other Resources

- https://www.w3schools.com/r/r_stat_intro.asp