Introduction to R and RStudio

Instituto Nacional de Medicina Genomica

Daniel Rivas & **Juan Manuel Mejia Arangure**

August 18, 2023

Outline - R Programming Fundamentals

- Data structures
- Common operations
  - Control Flow
  - Loops
  - Functions
- Statistics

Data structures

Primitives

Logical => TRUE(T) | FALSE(F)

Numeric => 357, 3.141596, 0, -59

Character => ‘char’, “char”

Integer => 9L

Complex => 46+5i

Primitives

Vectors -> one-dimensional collection of elements

Matrices -> two-dimensional collection of elements (size must be predefined)

Arrays -> *n*-dimensional collection of elements

Primitives

Factors -> one-dimensional collection of elements with predefined possible values

Lists -> arbitrary collection of any elements

Data Frames -> flexible two-dimensional representation of data

Workshop III

Common operations

Control Flow

Problem

Control Flow

Query data

Decision making

```
# # if statement
# if (condition) {
#   statement
# } else {
#   statement
# }
```

Loops

Problem

Loops

Defined or conditional repetition

Iterative

```
# # for loop
# for (value in vector) {
#    statements
# }
```

Functions

Problem

Functions

Customized work flow with variable inputs

```
# # no arguments
# funcx <- function() {
#   print('I am a function')
# }

# # arguments
# funcy <- function(x) {
#   x * 10
# }
```

Workshop IV

Statistics

Examples

Mean

```
# mean(x, trim = 0, na.rm = FALSE, ...)
```

Median

```
# median(x, na.rm = FALSE)
```

Summary

```
# summary(object, ...)
```

Examples

linear regression => y = ax + b

```
# lm(formula, data)
```

multiple regression => y = a + x1 + x2 + x3 …

```
# lm(y ~ x1 + x2 + x3..., data)
```

logistic regression

```
# glm(formula, data, family)
```

Examples

ancova analysis

```
# aov(formula, data = NULL, projections = FALSE, qr = TRUE, contrasts = NULL, ...)
```

time series analysis

```
# survfit(formula)
```

chi square

```
# chisq.test(data)
```

Workshop V

Daniel Rivas

danielrivasmd@gmail.com

Questions & Comments