# Giới Thiệu Cơ sở về Ngôn Ngữ R

1. Chúng ta có thể gán dữ liệu cho biến bằng hàm assign([tên biến],[giá trị])

Vd: assign(‘x’,5)…

2. Ta có thể dùng hàm rm() để xóa đối tượng trong R, đối tượng đó có thể là 1 dãy giá trị, hoặc cũng có thể là 1 giá trị đơn lẻ.

Vd: >a <-1:10

>rm(a)

Hoặc: > j<-4

>rm(j)

3. Kiểu dữ liệu:

-muốn kiểm tra kiểu dữ liệu số ta có hàm is.numeric() vd: is.numeric(a).

- Đối với kiểu số nguyên, ta có hàm is.integer(), tuy nhiên giá trị kiểm tra cần phải nối giá trị với L, thì khi đó hàm is.integer() sẽ hoạt động, nếu không kết quả sẽ mặc định là FALSE.Vd:

> i<-5

> is.integer(i)

[1] FALSE

> i<-5L

> is.integer(i)

[1] TRUE

-Đối với kiểu dữ liệu ký tự (character) có thể dùng hàm nchar() để đếm số lượng ký tự có trong đối tượng.

Vd: > x<-'Dai hoc Nha Trang'

> nchar(x)

[1] 17

- Ngoài ra, R còn cung cấp rất nhiều kiểu ngày tháng khác nhau, tuy nhiên, hữu ích nhất là date và POSIXct. Date() chỉ có thể lưu được ngày, trong khi POSIXct có thể lưu được cả ngày lẫn giờ. Vd:

> date2 <- as.Date("2012-06-28 17:42")

> date2

[1] "2012-06-28"

> date2 <- as.POSIXct("2012-06-28 17:42")

> date2

[1] "2012-06-28 17:42:00 +07"

Bên cạnh đó khi sử dụng hàm date() sẽ cho ra ngày giờ hiện tại.

> date()

[1] "Sun Sep 26 15:44:37 2021"

- Kiểu dữ liệu Logic: là kiểu dữ liệu mà kết quả của nó khi trả về chỉ có 2 khả năng True hoặc False. Nếu xét về mặt số học thì True là 1, False là 2. Vì vậy, True \* 5 = 5 và False \* 5 = 0.

> TRUE \* 5

[1] 5

> FALSE \* 5

[1] 0

> class(T)

[1] "logical"

Kiểu Logic cũng có thể trả về kết quả từ 1 biểu thức so sánh giữa 2 số hoặc giữa 2 chuỗi.

> 2 == 3

[1] FALSE

> 2 != 3

[1] TRUE

> 'dai hoc Nha Trang' == 'DAI HOC NHA TRANG'

[1] FALSE

> 'data' > 'states'

[1] FALSE

> 'data' < 'states'

[1] TRUE

## Phép toán với Vector

* Khi nhân 1 vector với 1 số thì tất cả các phần tử của vector sẽ lần lượt nhân với số đó. Vd:

> x \* 3

[1] 3 6 9 12 15 18 21 24 27 30

Không cần bất kỳ vòng lặp nào, các phép toán cộng, trừ, chia cũng hoạt động tương tự như vậy, xảy ra với mọi phép toán với vector.

> x + 2

[1] 3 4 5 6 7 8 9 10 11 12

> x - 3

[1] -2 -1 0 1 2 3 4 5 6 7

> x / 4

[1] 0.25 0.50 0.75 1.00 1.25 1.50 1.75 2.00 2.25 2.50

> x^2

[1] 1 4 9 16 25 36 49 64 81 100

> sqrt(x)

[1] 1.000000 1.414214 1.732051 2.000000 2.236068 2.449490 2.645751

[8] 2.828427 3.000000 3.162278

* So sánh cũng hoạt động trên vector. Ở đây kết quả là 1 vector có độ dài chứa True hoặc False cho mỗi phần tử.

> x <= 5

[1] TRUE TRUE TRUE TRUE TRUE FALSE FALSE FALSE FALSE FALSE

> x > y

[1] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE

> x < y

[1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

Hàm nchar cũng hoạt động trên từng phần tử của vector.

Vd: > q <- c("Hockey", "Football", "Baseball", "Curling", "Rugby", + "Lacrosse", "Basketball", "Tennis", "Cricket", "Soccer")

> nchar(q)

[1] 6 8 8 7 5 8 10 6 7 6

Tương tự như hàm nchar, hàm as.numeric() cũng đếm số lượng chữ cái của từng phần tử trong 1 vector.

> as.numeric(q)

[1] 6 5 1 4 8 7 2 10 3 9 6 7 6 11 6 7

# KIỂM ĐỊNH THỐNG KÊ

Chú ý:

-Trong TH KDTK đối vs 2 mẫu:

H0: chỉ có đẳng thức

|  |  |
| --- | --- |
| H0 | H1 |
| =1 | ≠1 |
| =1 | >1 |
| =1 | <1 |

* Kích thước 2 mẫu có thể khác nhau.

So sánh phương sai

F = có phân phối Fisher.

|  |  |
| --- | --- |
| H0 | H1 |
| =1 | ≠1 |
| =1 | >1 |
| =1 | <1 |

* Khi kiểm định chỉ cần xét F =
* P(F>)= P

Nếu th1: p=2\*P(F<) = 2 \* pf(f,n-1,m-1)

Nếu th2: p=2\*[1- P(F<)] = 2 \*[1-pf(f,n-1,m-1)]

=

* Trong R: qf();df = (n-1,m-1)
* Chú ý: ước lượng tỉ lệ phương sai với mức ý nghĩa :

( FL,FU). trong đó:

FU = , FL =

**KIỂM ĐỊNH GIA THUYẾT THỐNG KÊ VỚI 2 MẪU KHÔNG ĐỘC LẬP**

(Kiểm định với 2 mẫu có sự tương thích)

Matched or paired sample

Lưu ý:

1. Lấy mẫu ngẫu nhiên.
2. Kích thước mẫu nhỏ.
3. Mẫu kiểm định rút trích từ các cặp tương thích.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 1. X1 | 1. X11 | 1. .. | 1. X1i | 1. .. | 1. X1n |
| 1. X2 | 1. X21 | 1. .. | 1. X2i | 1. .. | 1. X2n |
| 1. D | 1. X11-X21 | 1. .. | 1. X1i-X2i | 1. .. | 1. X1n-X2n |

1. Kiểm định trên sự khác biệt giữa các cặp cá thể (D)
2. Giả thiết:

* Hoặc X1, X2 tuân theo phân phối chuẩn.
* Hoặc D có phân phối chuẩn.

So sánh giá trị trung bình của 2 tổng thể:

Gọi lần lượt kỳ vọng vs X1, X2. Xét kỳ vọng của D

Tính

Trong đó : : giá trị của D, ; Độ lệch chuẩn của D; n ;kích thước của D.

Thông thường

Chú ý:

t.test(x1,x2,paired = T, alt = ‘tow.side|less| greater’)

**KIỂM ĐỊNH KHI BÌNH PHƯƠNG**

1. Kiêm định:

* Dữ liệu có cung phân phối xs.
* Phân phối xs của dữ liệu.
* Tính độc lập của dữ liệu.

1. Chú ý:

* Giả thuyết có thể là câu/ mệnh đề/ biểu thức toán học.

1. Kiểm định liên quan đến đại lượng:

Trong đó:

O: giá trị quan sát (dữ liệu mẫu)

E: giá trị kỳ vọng

K : số khoảng/giá trị dữ liệu khác nhau.

H0 (dữ liệu thỏa lý thuyết) bị bác bỏ:

Q2 ≥

Trong đó: r: số tham số thống kê:

R: phân phối chuẩn 2

R: phân phối poisson 1

…

qchisq(1-alpha,K-r-1)

p-value:

p = 1 – P(X2 ≤ Q2) == 1 – pchisq(Q2,K-r-1)