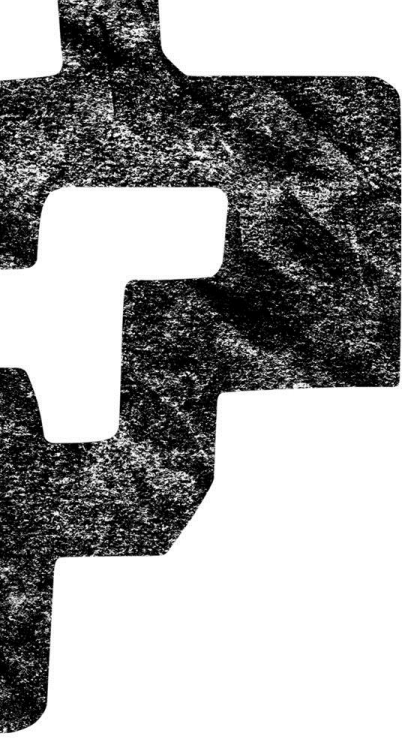


FREE R SOFTWARE

5. LESSON

Xabier Erdocia
Itsaso Leceta

eman ta zabal zazu
UPV EHU

OBJECTIVES

- ✓ Know how to install free R Studio software on the computer
- ✓ Be able to handle in R Studio environment and know how to perform simple operations
- ✓ Be able to define discrete and continuous distributions in R Studio
- ✓ Be able to calculate different probabilities using probability, density or distribution functions

INDEX

5.1. Installation of R

General concepts of R

- It is an environment designed for data processing, calculation, statistical analysis and graphics obtaining .
- It is a complete programming language that makes R a very versatile program
- The R programming language is included in GNU, a collaborative project of free software (https://www.gnu.org/ for more information).
- It is free software so users can run, copy, distribute, change and make improvements to the software with complete freedom.
- All the information about R can be found on the following website: https://www.r-project.org/

R Studio

- It is the work environment that will be used.
- In order to use R Studio, R must be previously installed.
- It is also free software and can be run on different operating systems (Windows, Linux, Mac)
- Similarly, using *RStudio Server* it can be run from the web page.

Installation of R Commander

- Installation is done through CRAN (R The Comprehensive Archive Network: https://cran.r-project.org/). In the same web page the necessary information to install the program is obtained.

WINDOWS

- From the official R web page download the latest version of R (4.0.2).
- Click on "Download R for Windows" and then click "install R for the first time".
- Download the R-4.0.2-win.exe file and click to start the installation.
- Follow the steps indicated by the installation program.

Mac OSX

- From the official R web page download the selected package of R (R.4.0.2.pkg, R.3.6.3.nn.pkg, R.3.3.3.pkg or R.3.2.1-snowleopard.pkg).
- Click on "Download R for (Mac) OS X"
- Install the chosen package following the steps indicated.

Installation of R Studio

- From the official web page of R Studio download the latest version depending on the operating system you use.
- Follow the following steps::
  - ✓ Click on the first icon on the left: *Download RStudio Desktop (Open Source License) FREE.*
  - ✓ It will ask if we have R installed. If we do not have it, we must install it in order to use R Studio. (Follow the steps indicated on the previous slide).
  - ✓ Once R is installed, download the R Studio installation package depending on the operating system.
    - **Windows:** RStudio 1.3.1073.exe - Windows 10/8/7
    - **Mac OS:** RStudio 1.3.1073.dmg - Mac OS 10.13+
  - ✓ Follow the steps indicated by the installation program.

5.2. First steps with R

- After clicking on the R Studio icon, the following screen will appear for the first time.

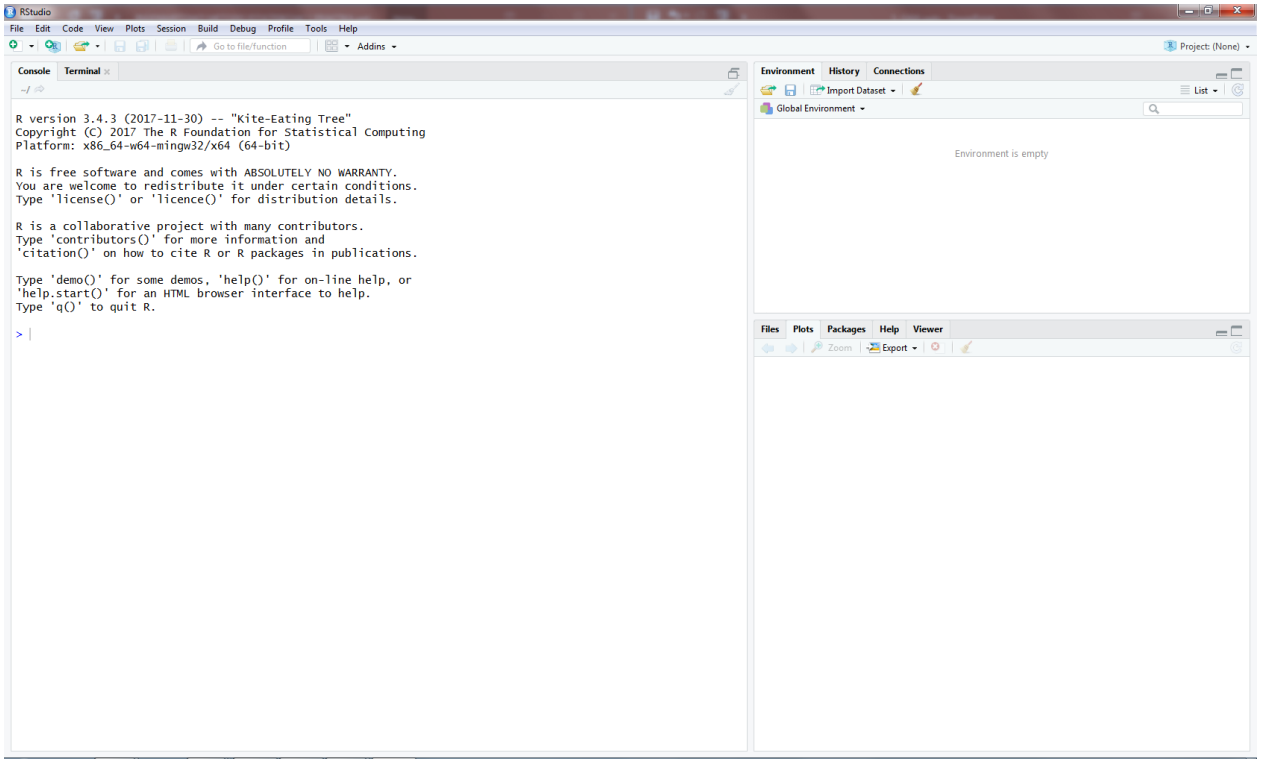


Figure 1. R Studio software initial screen.

1. Session management

- First of all, it is recommended to open a **Script**: *File->New File-> R Script (Ctrl+Shift+N)*
- A new window will open and on it we will write all the commands to use.
- To save what is written in the Script: *File->Save as…* Assign the desired name and the location where it is wanted to be saved.
- The files will be saved with the extension .R .
- To open a previously saved Script: *File->Open File->…*
- What is written in the Script will be executed in a window called **Console**.
- It can be executed line by line(*Ctrl+Enter*) or execute all the written in the Script in one go (*Ctrl+Alt+R*). It can also chosen which part of the code to execute using *Code->Run Region->…*
- It can even be written directly to the Console and run it at the same time (not recommended).

1. Session management

- While you are working, to save the files, to open them, etc., we are in a certain working directory.
- To know the working directory in which we are: getwd()
- To change the working directory: *Session->Set Working Directory->Choose Directory... (Ctrl+Shift+H)*
- To display session orders: history (inf) or view them on the panel History (up to the right).
- To save the used objects: on the panel Environment (up to the right) click on Save.
- To get help: *Help->R Help* or on the panel Help (bottom right) write the question you have.

2. Basic operations

- Basic arithmetic operations can be performed.
- Notation of the arithmetic operators: +,-,*,/ and ^ (addition, subtraction, multiplication, division and empowerment respectively).
- The operations of the line will be read from left to right taking into account the **priority of the operators**. Parentheses will be used to give priority to the operations deemed appropriate.
- Different functions or operations can be entered on the same line, separated by a semicolon (;) .

Example

Executed on the Console

```
2*3-5
[1] 1
 5/4^3;-1^2-9;(-1)^2-9;(-1)^(2-9)
[1] 0.078125
[1] -10
[1] -8
[1] -1
(9-11)^2/(5-4)*3^2;(9-11)^2/((5-4)*3^2)
[1] 36
[1] 0.4444444
```

2. Basic operations

- In R there are numerous mathematical and statistical commands.
- As a command is being typed, R will help to type it correctly.
- After each command, parentheses will be placed writing the necessary arguments inside them.
- To introduce comments that are not wanted to be executed, the # symbol will be placed in front of them.

Example

Executed on the Console

```
>sqrt(7)#square root
[1] 2.645751
>exp(6)#exponential
[1] 403.4288
>sin(pi/2)/cos(pi)#sine and cosine commands
[1] -1
>log(7);log(2,7)#natural logarithm command; logarithm to base 2
[1] 1.94591 [1] 0.3562072
>factorial(9)#factorial command
[1] 362880
>choose(6,3)#combination
[1] 20
```

5.3. Discrete distributions with R

- **Discrete distribution commands**

Table 1: Commands for discrete distributions in R.

Distribution	Command
Binomial	binom
Geometric	geom
Negative binomial	nbinom
Hypergeometric	hyper
Poisson	pois

- **Prefixes**

Table 2: Prefixes for obtaining different functions.

Functions	Prefix
Probability function	d
Distribution function	p
Create random values	r
Quantile function	q

Examples

1. **Let be a random variable following a binomial distribution. Parameters n=7, p=0.15. Calculate:**

   a) The probability that it takes the value of 2.

```
> dbinom(2,size=7,prob=0.15)#Probability function
   [1] 0.2096508
```

   In a simplified way

```
> dbinom(4,7,0.15)#Probability function in a simplified way
   [1] 0.2096508
```

   b) Probability that it takes at most the value of 3.

```
> pbinom(3,7,0.15)#Distribution function in a simplified way
   [1] 0.9878968
```

Examples

2. Let be a random variable following a geometric distribution. Parameter p=0.15. Calculate:

a) The probability that it takes the value of 3.

```
> dgeom(3,prob=0.15)#Probability function
   [1] 0.09211875
```

In a simplified way

```
> dgeom(3,0.15)#Probability function in a simplified way
   [1] 0.09211875
```

b) Probability that it takes at most the value of 4.

```
> pgeom(4,0.15)#Distribution function in a simplified way
   [1] 0.5562947
```

Examples

3. Let be a random variable following a negative binomial distribution. Parameters n=12, p=0.34. Calculate:

a) The probability that it takes the value of 6.

```
> dnbinom(6, size=12,prob=0.34)#Probability function
   [1] 0.00244113
```

In a simplified way

```
> dnbinom(6,12,0.34)#Probability function in a simplified way
   [1] 0.00244113
```

b) Probability that it takes at most the value of 12.

```
> pnbinom(12,12,0.34)#Distribution function in a simplified way
   [1] 0.07754832
```

Examples

4. Let be a random variable following an hypergeometric distribution. Parameters N=30 (size of the population), m=8 (number of successes in the population), n= 22 (number of failures in the population) and k=6 (number of tests). Calculate:

a) The probability that it takes the value of 5.

```
> dhyper(5,m=8,n=22,k=6)#Probability function
   [1] 0.00207486
```

In a simplified way

```
> dhyper(5,8,22,6)#Probability function in a simplified way
   [1] 0.00207486
```

b) Probability that it takes at least the value of 5.

```
> phyper(4,8,22,6,lower.tail=F)#Distribution function in a simplified
  way and using the command lower.tail=F, P(X>x) is calculated
   [1] 0.002122016
```

Examples

5. Let be a random variable following a Poisson distribution. Parameter λ=6. Calculate:

a) The probability that it takes the value of 7.

```
> dpois(7,lambda=6)#Probability function
   [1] 0.137677
```

In a simplified way

```
> dpois(7,6)#Probability function in a simplified way
   [1] 0.137677
```

b) The minimum value of the random variable for a cumulative probability of 0.9. P(X≤*x*)=0.9

```
> qpois(0.9,6)#Quantile function in a simplified way
   [1] 9
```

5.4. Continuous distributions with R

- **Continuous distribution commands**

Table 3: Commands for continuous distributions in R.

Distribution	Command
Uniform	unif
Exponential	exp
Normal	norm

- **Prefixes**

Table 4: Prefixes for obtaining different functions.

Functions	Prefix
Density function	d
Distribution function	p
Create random values	r
Quantile function	q

Examples

6. Let be a random variable following a uniform distribution. Parameters a=6 and b=20. Calculate:

a) The probability that it takes at least the value of 7 and at most the value of 10.

```
> punif(10,min=6,max=20)-punif(7,min=6,max=20)#Distribution function
  [1] 0.2142857
```

In a simplified way

```
> punif(10,6,20)-punif(7,6,20)#Distribution function in a simplified
  way
  [1] 0.2142857
```

b) The value of the random variable that leaves the 20% of the distribution on the right.

```
> qunif(0.2,6,20,lower.tail=F)#Quantile function in a simplified
  way and using the command lower.tail=F, P(X>x)=0.20 is calculated
  [1] 17.2
```

Examples

7. **Let be a random variable following an exponential distribution. Parameter λ=1/3. Calculate:**

a) The probability that it takes at most the value of 5.

```
> pexp(5,rate=1/3)#Distribution function
   [1] 0.8111244
```

In a simplified way

```
> pexp(5,1/3)#Distribution function in a simplified way
   [1] 0.8111244
```

b) The value of the random variable that leaves the 50% of the distribution on the left.

```
> qexp(0.5,1/3)#Quantile function in a simplified way
   [1] 2.079442
```

Examples

8. Let be a random variable following a normal distribution. Parameters $\mu=35$ and $\sigma=7$. Calculate:

a) The probability that the random variables has a maximum difference of 5 from the mean.

```
> pnorm(40,mean=35,sd=7)-pnorm(30,mean=35,sd=7)#Distribution function
  [1] 0.5249495
```

In a simplified way

```
> pnorm(40,35,7)-pnorm(30,35,7)#Distribution function in a simplified
  way
  [1] 0.5249495
```

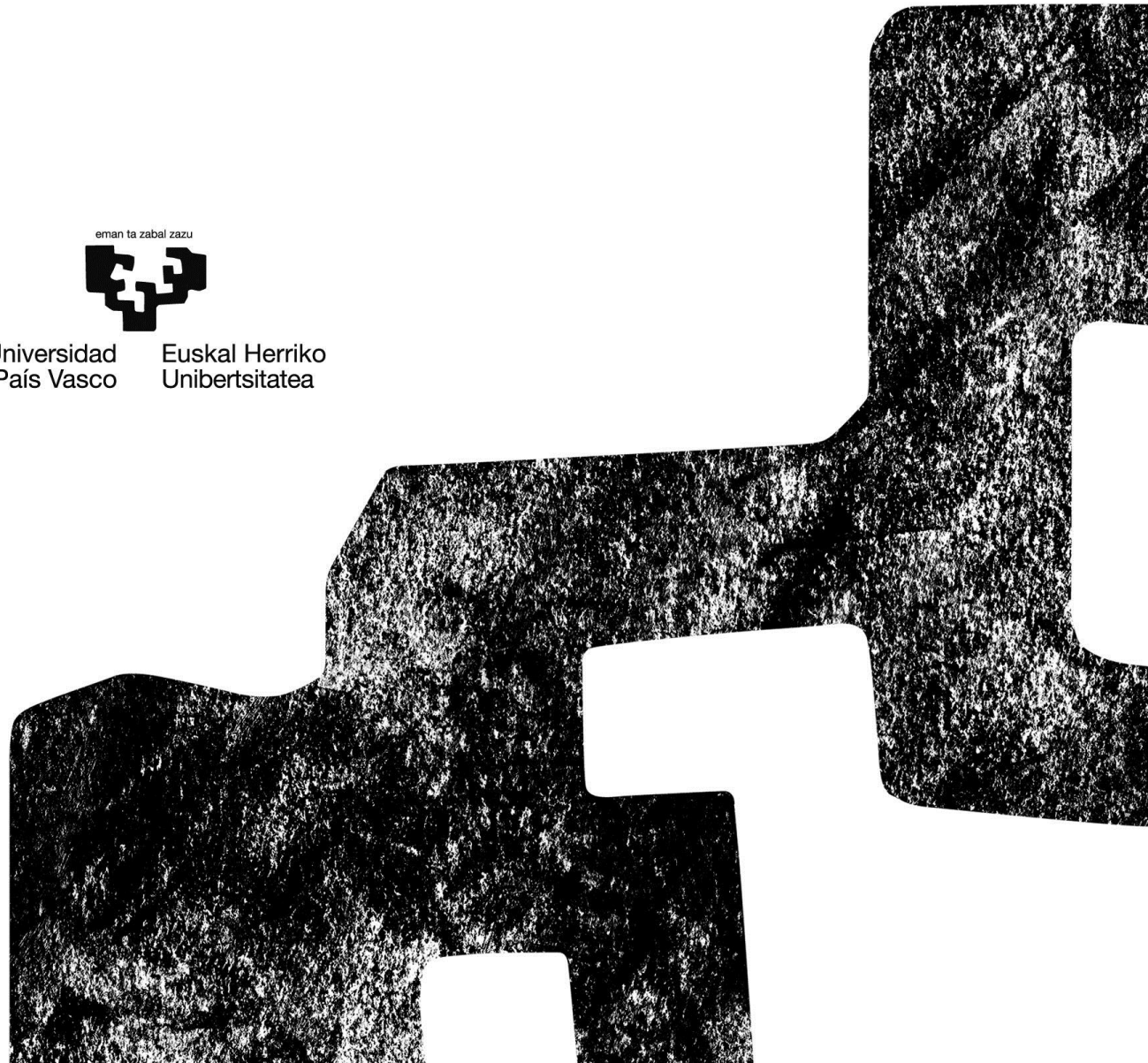
b) The probability that it takes at most the value of 43.

```
> pnorm(43,35,7)#Distribution function
  [1] 0.8734
```

eman ta zabal zazu

Universidad del País Vasco

Euskal Herriko Unibertsitatea



OCW