

MODELO DE PRÁCTICAS DE LABORATORIO CON EL LENGUAJE R, PARA LA  
SOLUCIÓN DE PROBLEMAS DE ESTADÍSTICA 1, EN EL PROGRAMA DE  
INGENIERÍA DE SISTEMAS EN LA UNIVERSIDAD TECNOLÓGICA DE PEREIRA.

JUAN ALEJANDRO RAMIREZ CHUJFI

LUIS FERNANDO DÍAZ HERRERA

TRABAJO DE GRADO

DIRECTOR DE TRABAJO

DR. SAULO DE JESUS TORRES RENGIFO

UNIVERSIDAD TECNOLÓGICA DE PEREIRA

FACULTAD DE INGENIERIAS

INGENIERIA DE SISTEMAS Y COMPUTACION

PEREIRA - RISARALDA

2016

## Tabla de contenido

	Pág.
1. Introducción e Instalación de R .....	5
1.1. Introducción .....	5
1.2. Instalación de R.....	6
1.2.1. Pasos de instalación .....	8
1.3. El editor y la consola de R .....	14
2. Introducción al manejo del lenguaje R. ....	17
2.1. Introducción .....	17
2.2. Tipos de objetos del Lenguaje R .....	17
2.2.1. Data Frame: .....	18
2.2.2. Matrices: .....	19
2.2.3. Array:.....	20
2.2.4. Vectores: .....	20
2.2.5. Factores:.....	21
2.2.6. Listas:.....	23
2.3. Funciones Esenciales en R .....	24
2.4. Operaciones Lógicas .....	24
2.5. Asistente de Ayuda en R .....	26
3. Estadística Descriptiva.....	27
3.1. Tablas de frecuencia.....	27
3.1.1. Frecuencia Absoluta .....	28
3.1.2. Frecuencia relativa.....	28
3.1.3. Suma de frecuencias .....	29

3.1.4. Frecuencia absoluta acumulada .....	30
3.1.5. Frecuencia relativa acumulada .....	30
3.1.6. Función Scan ().....	31
3.1.7. Función Summary ().....	32
3.1.8. Función fivenum () .....	33
3.2. Medidas de tendencia central .....	34
3.2.1. Media aritmética: .....	34
3.2.2. Media geométrica .....	36
3.2.3. Moda.....	37
3.2.4. Mediana .....	38
3.2.5. Cuantiles .....	39
3.3. Medidas de Dispersión .....	41
3.3.1. Rango.....	41
3.3.2. Rango Medio .....	42
3.3.3. Rango Intercuartílico .....	43
3.3.4. Varianza.....	44
3.3.5. Desviación Típica .....	45
4. Ejercicios propuestos .....	46
5. Bibliografía .....	52

## Tabla de imágenes

	Pág.
Imagen 1: Vista principal de la página oficial de R.....	6
Imagen 2: Vista principal de la página oficial de R.....	7
Imagen 3: Vista de opción de descarga de R. ....	7
Imagen 4: Vista de descarga de R.....	8
Imagen 6: Vista de selección del idioma de R.....	9
Imagen 7: Vista de asistente de instalación de R.....	9
Imagen 8: Vista de información de licencia GNU – GPL para instalar R. ....	10
Imagen 9: Vista de selección de carpeta de destino para R. ....	10
Imagen 10: Vista de selección de componentes de R. ....	11
Imagen 11: Vista de opciones de configuración de R.....	11
Imagen 12: Vista de selección de nombre y ubicación para los accesos directos de R.....	12
Imagen 13: Selección de tareas adicionales en la instalación de R. ....	13
Imagen 14: Vista del estado de progreso de instalación de R.....	13
Imagen 15: Vista instalación satisfactoria de R.....	14
Imagen 16: Vista de Consola de R.....	15
Imagen 17: Vista de Consola y Script de R. ....	16

## Capítulo 1

### 1. Introducción e Instalación de R

#### 1.1. Introducción

R es un lenguaje de programación y a su vez un entorno, disponible como software libre y distribuido bajo la licencia GNU GPL. Este lenguaje es de orientación estadística, y se diferencia de otros programas que se suelen usar con una interfaz de ventana, y con varias opciones visuales, que a su vez son más limitadas. R es manejado por medio de consola, para ingresar el código nativo de su lenguaje, para obtener las funciones, gráficas y los resultados buscados.

El lenguaje R es desarrollado por Robert Gentleman y Ross Ihaka del Departamento de Estadística de la *University of Auckland*, ubicada en Auckland, Nueva Zelanda. La base de este lenguaje, fue el lenguaje S, donde los diseñadores iniciales Gentleman e Ihaka hicieron la combinación de las virtudes del Lenguaje S y Scheme (ahora llamada Racket).

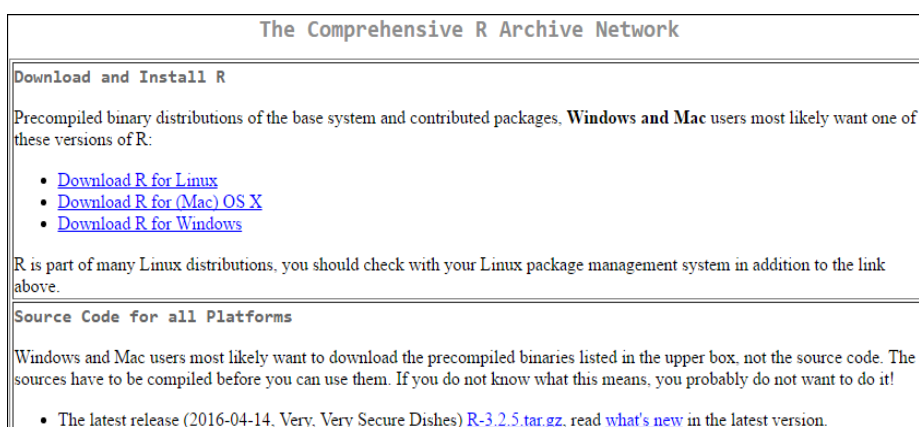
R en un inicio no contenía la robustez e influencia que posee hoy en día en el campo del estudio y el análisis estadístico. Éste gran resultado se debe en gran parte a las diferentes comunidades de usuarios de todo el mundo que trabajan en él, donde sus aportes a R son voluntarios y de carácter colaborativo.

## 1.2. Instalación de R

La instalación de R puede hacerse por consola, mediante dispositivos de almacenamiento que contengan los instaladores o descargándolo desde sitios de internet. Es recomendable hacer la descarga desde la página oficial, en la que se encuentra la versión más actualizada o versiones posteriores:

<https://cran.r-project.org/>

Al ingresar a la página oficial se encuentran disponibles los instaladores de R para los distintos sistemas operativos Windows, Linux y Mac OS X (ver Imagen 1). También se encuentra la fecha y enlace de la última versión de R, y a su lado un enlace en el que se puede leer con detalle los cambios o contenidos nuevos con respecto a otras versiones de R.

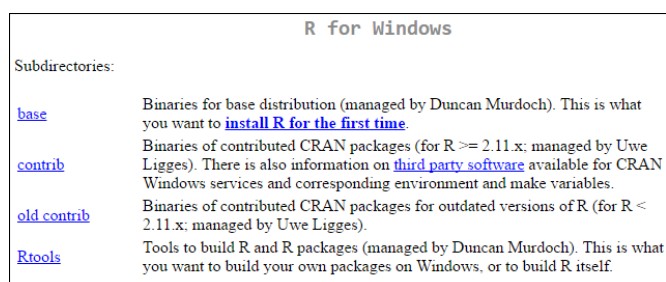


*Imagen 1: Vista principal de la página oficial de R.*

*Fuente: Los autores (2016). Recuperado de <https://cran.r-project.org/>*

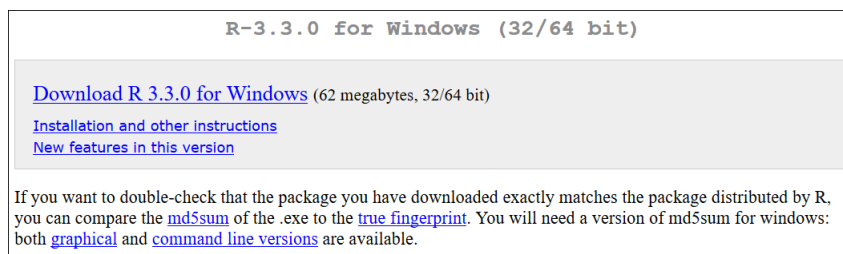
Una vez identificado el instalador para el sistema operativo en el que se desea instalar R, Windows en este caso, se hace clic en la opción respectiva, y a continuación se selecciona la opción base de la otra vista de la página web (ver Imagen 2), y a continuación se descarga la última versión de R ya sea para 32 y 64 bits (ver Imagen 3).

Es recomendable elegir la última versión disponible de R, 3.3.0. en este momento. Si es posible, sin embargo, cada vez se debe procurar actualizar la versión que cada quien tenga en uso. Solo en caso de presentar inconvenientes con la actualización, restaure la versión con la que venía trabajando y reporte su inconveniente en los distintos foros o demás medios de comunicación de comunidades R, para así encontrar o construir una solución a esto.



*Imagen 2: Vista principal de la página oficial de R.*

*Fuente: Los autores (2016). Recuperado de <https://cran.r-project.org/>*

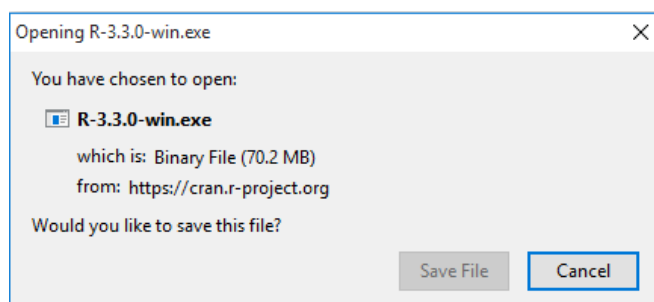


*Imagen 3: Vista de opción de descarga de R.*

*Fuente: Los autores (2016). Recuperado de <https://cran.r-project.org/>*

### 1.2.1. Pasos de instalación

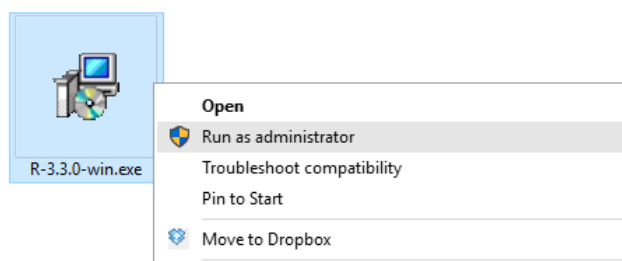
Una vez seguidas las instrucciones anteriormente expuestas, la instalación de R se convierte en algo fútil, porque solo basta con guardar el archivo .exe que se descarga desde su sitio oficial (ver imagen 4)



*Imagen 4: Vista de descarga de R.*

*Fuente: Los autores (2016).*

Después de guardar el instalador de R, se procederá a ejecutarlo como administrador del equipo (ver imagen 5):

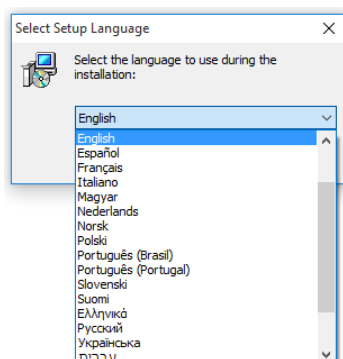


*Imagen 5: Vista de modo de ejecución para R.*

*Fuente: Los autores (2016).*



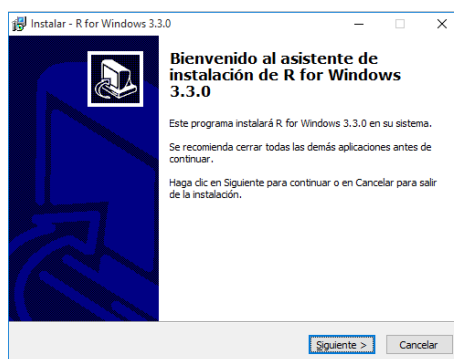
Inmediatamente a la ejecución, se abrirá una ventana para seleccionar nuestro idioma de preferencia (ver imagen 6).



*Imagen 5: Vista de selección del idioma de R.*

*Fuente: Los autores (2016).*

Una vez seleccionado el idioma, se generará una nueva ventaba de bienvenida al asistente de instalación que nos guiará en el transcurso de toda la instalación de R (ver imagen 7), y para dar continuación al mismo, se presiona la opción de siguiente.



*Imagen 6: Vista de asistente de instalación de R.*

*Fuente: Los autores (2016).*

A continuación, se verá los términos y condiciones previas a instalar R, una vez leído y estando de acuerdo se presiona la opción de siguiente, en caso de no estar de acuerdo debe ser presionada la opción cancelar para detener la instalación (ver imagen 8).

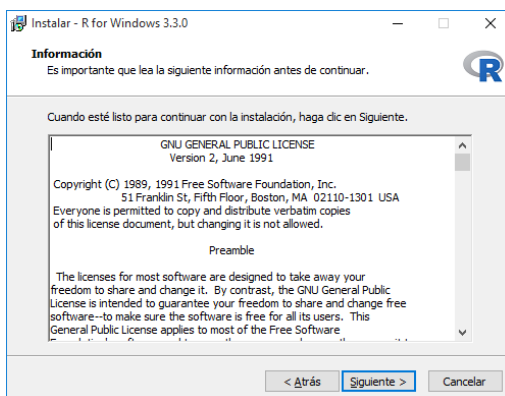


Imagen 7: Vista de información de licencia GNU – GPL para instalar R.

Fuente: Los autores (2016).

En este siguiente paso se selecciona la ruta de destino donde se instalará R en el disco duro, para este caso, se dejará la ruta por defecto (ver imagen 9).

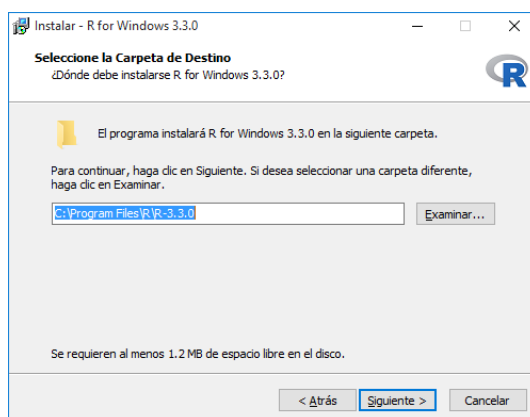
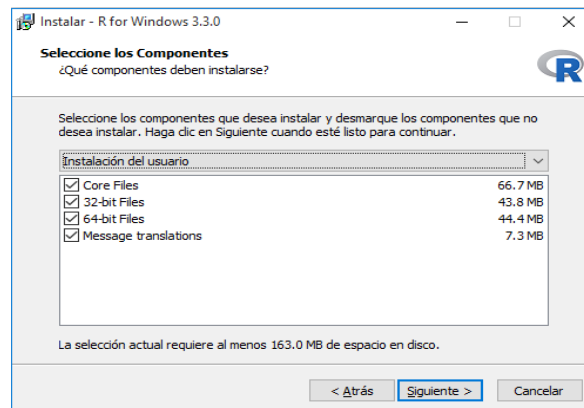


Imagen 8: Vista de selección de carpeta de destino para R.

Fuente: Los autores (2016).

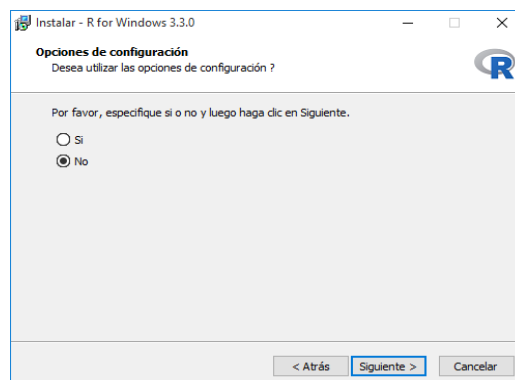
Ahora aparecerá la vista de los componentes que se pueden instalar para R, entre ellas se encuentran archivos de núcleo (core files), las versiones de 32 y/o 64 bits y traducciones de mensajes. Para este caso se dejará marcadas todas las casillas (ver imagen 10).



*Imagen 9: Vista de selección de componentes de R.*

*Fuente: Los autores (2016).*

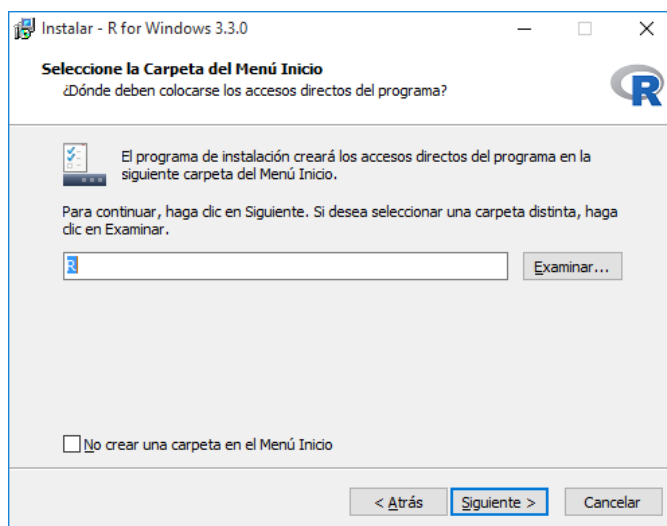
En la vista de opciones de configuración, se dejará todas las opciones de configuración que vienen por defecto, marcando la casilla “No”, después de esto se selecciona la opción siguiente (ver imagen 11).



*Imagen 10: Vista de opciones de configuración de R.*

*Fuente: Los autores (2016).*

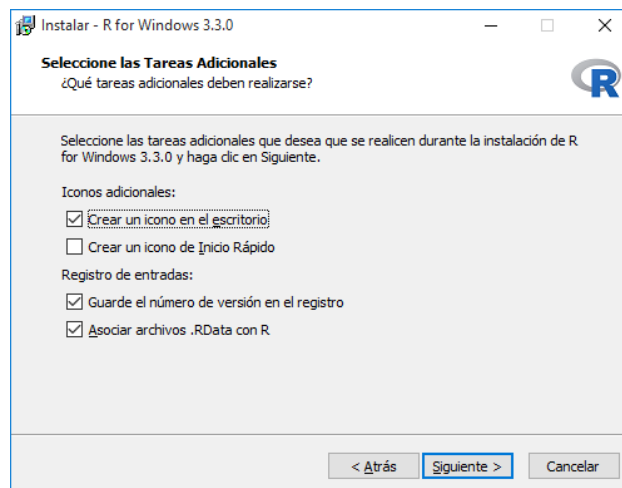
Ahora se procede a especificar el nombre y ubicación para los accesos directos de R, donde es aconsejable dejar esto como se encuentra por defecto, y se procede a dar clic en la opción de siguiente (ver imagen 12).



*Imagen 11: Vista de selección de nombre y ubicación para los accesos directos de R.*

*Fuente: Los autores (2016).*

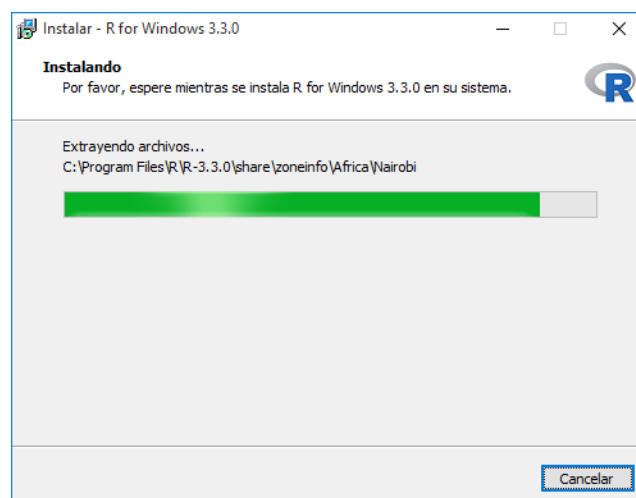
Y como último paso se ven las tareas adicionales que proporciona el instalador de R, como la creación iconos de acceso a R en el escritorio, guardar el número de versión, asociar todos los archivos .RData a lenguaje R. Se dejan las opciones por defecto y se hace clic en la opción siguiente (ver imagen 13).



*Imagen 12: Selección de tareas adicionales en la instalación de R.*

*Fuente: Los autores (2016).*

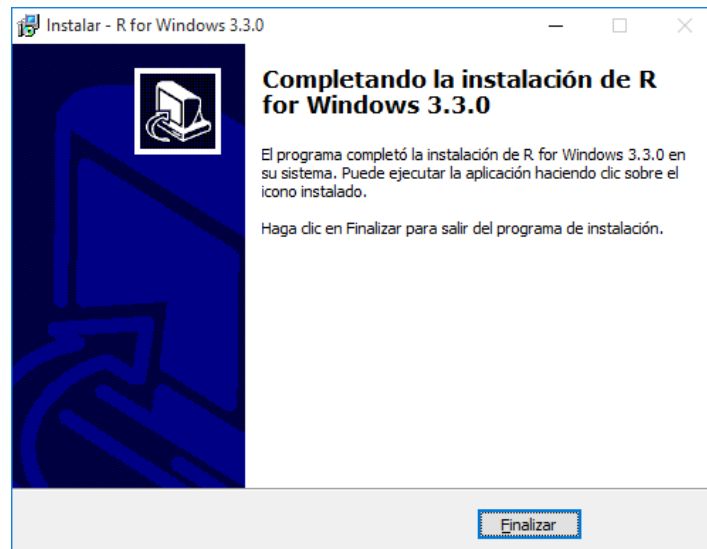
Una vez iniciado el proceso de instalación, se da espera a su culminación (ver imagen 14).



*Imagen 13: Vista del estado de progreso de instalación de R.*

*Fuente: Los autores (2016).*

En una nueva ventana se notifica la instalación de R, al finalizar la misma. Para salir de ella e iniciar a usar R, se procederá a usar la opción Finalizar (ver imagen 15).

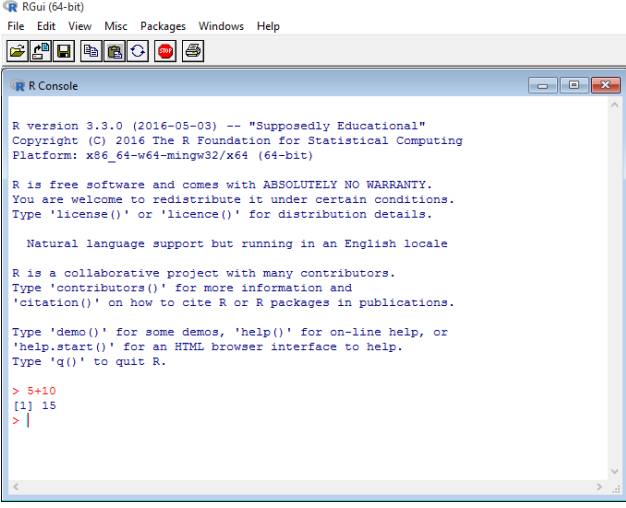


*Imagen 14: Vista instalación satisfactoria de R.*

*Fuente: Los autores (2016).*

### 1.3. El editor y la consola de R

La consola es lo primero que aparece como una ventana al momento al abrir R, donde se puede manejar R mediante la escritura de su código. Para comprobar que R esté funcionando correctamente en nuestro PC, se ingresa operaciones aritméticas en la consola. Por ejemplo, escribir  $5+10$  en ella, y a continuación se pulsa la tecla Intro para ejecutar la operación como tal, lo que devolverá como resultado en la misma consola es el resultado de la operación, que sería el valor de 15 (ver imagen 16).



```

RGui (64-bit)
File Edit View Misc Packages Windows Help

R Console

R version 3.3.0 (2016-05-03) -- "Supposedly Educational"
Copyright (C) 2016 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> 5+10
[1] 15
> |

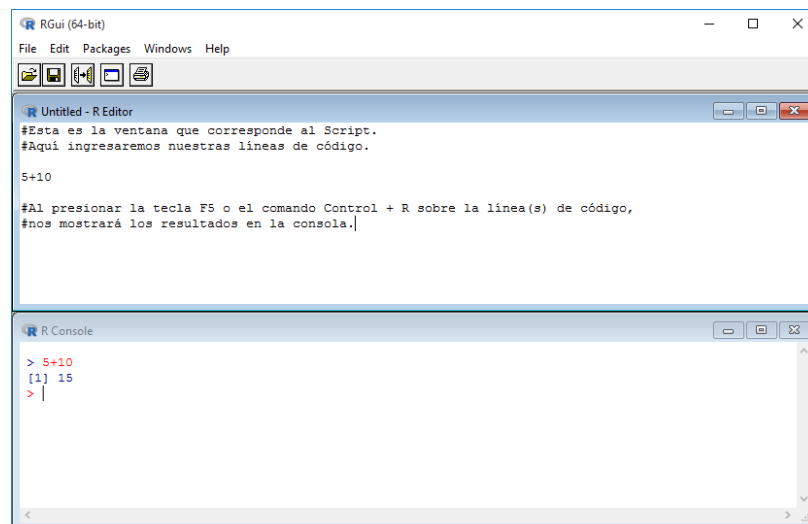
```

*Imagen 15: Vista de Consola de R.*

*Fuente: Los autores (2016).*

Vale la pena tener presente que el trabajo directo desde la consola no es apropiado para trabajos más robustos y complejos que las operaciones aritméticas, por la gran cantidad de entradas de datos y líneas de código que se requieren, así como sus correcciones, modificaciones, mejoras, repeticiones y guardar continuamente el trabajo entre las funciones. Por las razones anteriormente expuestas entre otras, la forma más apropiada para trabajar el lenguaje R es desde el editor.

Para iniciar un nuevo trabajo New Script por el editor de R, se selecciona la opción que se encuentra en la ventana principal llamada File, y allí se desplegará varias opciones, donde se dará clic en New Script, posterior a esto se abrirá una nueva ventana para el nuevo Script. Para trabajar de una forma más cómoda en las dos ventanas (Consola y Script) al mismo tiempo, elegirá la herramienta Windows y luego clic en la opción Tile Horizontally. Una vez hecho esto, el aspecto deberá quedar como se muestra en la imagen 17.



*Imagen 16: Vista de Consola y Script de R.*

*Fuente: Los autores (2016).*

En la imagen 17, se pueden apreciar varias cosas, entre ellas está la forma que se ingresa comentarios en R, que corresponde al carácter numeral (#), donde todo lo que se escriba después de este símbolo, no será leído por R. Y para el resto de código realizado, se puede ejecutar de dos formas, una de ellas es presionando la tecla F5, ubicándose en cualquier posición de esa línea (no necesariamente en el final), y la otra forma es usando el comando Control + R. En caso de querer ejecutar más de una línea, bastará con que sea seleccionadas y ejecutadas también con Control + R o F5.

Para guardar el trabajo realizado en el Script, se selecciona la opción *File* y se hace clic en la opción *Save* o *Save as*, y posterior a esto se selecciona el sitio donde se guardará el trabajo como tipo de archivo *.R*. Es importante guardar todo el trabajo realizado en el Script, ya que esto permitirá abrirlo en otro momento seleccionando la opción *Open Script*, para continuar trabajando en él.



## Capítulo 2

### 2. Introducción al manejo del lenguaje R.

#### 2.1. Introducción

En este capítulo se hablará sobre varias de las entidades que son manejadas en lenguaje R, más conocidas como objetos, en otros lenguajes son llamadas estructuras de datos, como lo son las matrices, listas, vectores, etc. Serán abordados los objetos más importantes en R para este capítulo, teniendo como objetivo enseñar su correcto uso y concepto.

#### 2.2. Tipos de objetos del Lenguaje R

Los objetos o elementos que son manejados en el lenguaje R, deben ser declarados y distinguidos de forma clara para su correcto uso. Como por ejemplo en el caso de las matrices, siendo estos definidos como un contenedor de una serie de elementos (números, caracteres, cadenas, etc.) del mismo tipo, configurada su ubicación en filas y columnas. Pero si necesita reunir vectores que contiene en ellos números, y estos vectores no manejan una misma dimensión, entonces es allí donde se requiere usar como objeto las listas, que, aunque no son una matriz, se puede guardar en ellas vectores de distintas dimensiones, pero a la hora de querer realizar operaciones matriciales, no lo podría hacer.

En lo anteriormente expuesto, se hace con intención de aclarar la importancia de tener claro el concepto de los distintos objetos de R, para la aplicación de los mismos, en que se podrá realizar a cabalidad los propósitos que sean tenidos en este lenguaje estadístico. Los objetos más pertinentes y que serán manejados posteriormente, son los siguientes:

- Data Frame
- Matrices
- Array
- Vectores
- Factores
- Listas

#### 2.2.1. Data Frame:

Los Data Frame son usados para almacenar tablas de datos. Son una tabla de vectores de igual longitud. Por ejemplo, la siguiente variable “df” es un Data Frame que contiene 3 vectores a, b, c:

```
> a = c (1, 2, 3)
```

```
> b = c (“A”, ” B”, ”C”)
```

```
> c = c (FALSE, TRUE, TRUE)
```

```
> #df es un DataFrame
```

```
> df = data.frame (a,b,c)
```

### 2.2.2. Matrices:

Existen varias alternativas para crear una matriz, cuando se va a construir una, e ingresando directamente los datos, la matriz se va llenando en orientación de las columnas, esto lo realiza por defecto. Por ejemplo, en el siguiente código, el contenido de B se va llenando a lo largo de las columnas de formas consecutiva:

```
> BM = matrix(  
c (2, 4, 6, 8, 10, 12),  
nrow=3,  
ncol=2)  
  
> # BM tiene 3 filas y 2 columnas  
  
> BM  
      [,1] [,2]  
[1,]  2   8  
[2,]  4  10  
[3,]  6  12
```

### 2.2.3. Array:

También llamadas variables indexadas es un objeto en que se podrá guardar elementos, es decir, es una generalización de una matriz, pero este es multidimensional. La estructura para crear un array es: array (datos, dimensiones). Por ejemplo, la siguiente variable “ar”, es un array, que contiene los números del 1 al 12, y a su vez contendrá en ella 2 matrices, donde cada una de ellas tienen como dimensión 2 filas y 3 columnas:

```
> ar = array (1:12, c (2,3,2)) # ar es un array
,, 1
 [1] [2] [3]
[1,]  1  3  5
[2,]  2  4  6
,, 2
 [1] [2] [3]
[1,]  7  9 11
[2,]  8 10 12
```

### 2.2.4. Vectores:

Es una forma u objeto en el que se puede almacenar una serie de elementos del mismo tipo (números, cadenas, valores lógicos, etc.), y cada uno de los elementos allí almacenados, tiene una posición asociada para poder acceder a ellos. La posición o índice de los elementos

siempre será un número entero positivo. Por ejemplo, la variable “vec”, contendrá 4 elementos numéricos.

```
> # Vector numérico
```

```
> vec = c (1,3,5,7)
```

```
> # Vector lógico
```

```
> vec_log = c (F, T, F, T, F)
```

```
> #Vector de cadenas
```

```
> vec_cad = c (“pereira”, “cartago”, “manizales”)
```

#### 2.2.5. Factores:

Los factores se asocian a la definición de un vector, con la diferencia de que éste se usa para clasificar de forma discreta los elementos de vectores de igual longitud. Por ejemplo, en el siguiente vector se tendrán almacenados las ciudades de origen de 7 personas:

```
> personas.origen = c (“cartago”, ”pereira”, ”armenia”, ”salento”, ”cartago”, ”cartago”,
”pereira”)
```

```
> personas.origen
```

```
[1] “cartago” “pereira” “armenia” “salento” “cartago” “cartago”
```

```
[7] “pereira”
```

En la anterior instrucción, se ha creado un vector que contiene 7 ciudades, que representa la ciudad de origen de personas, A continuación, se procederá a crear una variable de tipo factor, precediendo de un vector existente:

```
> fac.estudiantes = as.factor (personas.origen)
```

```
> fac.estudiantes
```

```
[1] "cartago" "pereira" "armenia" "salento" "cartago" "cartago"
```

```
[7] "pereira"
```

Con la función levels, se pide los tipos de factores que hay en la variable fac.estudiantes:

```
> levels (fac.estudiantes)
```

```
[1] "cartago" "pereira" "armenia" "salento"
```

Por último, se emplea la función summary, usada para pedir el sumario de la variable fac.estudiantes, y el resultado será una tabla que nos muestra los niveles de factor, junto con el número de incidencias, que corresponde a las repeticiones de las ciudades.

```
> summary (fac.estudiantes)
```

```
Cartago pereira armenia salento
```

```
3      2      1      1
```

### 2.2.6. Listas:

Las listas son un tipo de estructura de dato, que nos permite concatenar objetos, con la ventaja de que no necesariamente deben tener la misma estructura, característica que no pasa por ejemplo en los arrays o en las matrices. En el siguiente ejemplo, se creará una lista a la que será llamada comida:

```
> comida = list (carne = "pollo", postre = "chocolate", cantidad.jugos = 3, sabor.jugos = c  
("fresa", "naranja", "mora"), precio.jugos = c (3000, 2500, 2300), ciudad = "pereira")
```

```
> comida
```

```
carne
```

```
[1] "pollo"
```

```
postre
```

```
[1] "chocolate"
```

```
cantidad.jugos
```

```
[1] 3
```

```
sabor.jugos
```

```
[1] "fresa" "naranja" "mora"
```

```
precio.jugos
```

```
[1] 3000 2500 2300
```

```
ciudad
```

```
[1] "pereira"
```

### 2.3. Funciones Esenciales en R

```
> sort ()      # Ordena un vector de menor a mayor

> max () y min ()    # Extrae el valor máximo o mínimo de un argumento

> exp ()       # Representa la operación exponencial aplicada a un valor

> log ()       # Realiza la función de un logaritmo natural

> log10 ()     # Representa el logaritmo en base 10

> sqrt ()      # Opera la raíz cuadrada de un valor

> sum ()       # Realiza la suma de un argumento

> rowSums () y colSums ()    # Función encargada de realizar la suma de las filas o

                                # columnas de un Data Frame

> prod ()      # Realiza el producto de un argumento

> cumsum ()    # Realiza una suma acumulada de un vector
```

### 2.4. Operaciones Lógicas

Lenguaje R permite realizar operaciones lógicas, involucrando los objetos de R. A continuación, se darán algunos casos con los operadores lógicos más frecuentes.

- Los siguientes caracteres `<`, `>`, `<=`, `>=` representan los operadores Menor, Mayor, Menor que y Mayor que.



Ejemplo:

```
> x = c(1, 5, 9)
```

```
> x >= 6
```

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

- Los dobles signos de igual (=), representa el operador de igualdad.

Ejemplo:

```
> color = c(rojo, verde, azul)
```

```
> color == "verde"
```

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

- Los caracteres & y |, representan los operadores lógicos AND y OR.

Ejemplo:

```
#Crea una matriz de 3x3 con los números del 1 al 9
```

```
> ma.num = matrix(c(1:9), nrow=3, ncol=3)
```

```
#Evalúa los valores mayores de 4 y menores de 7
```

```
> (ma.num > 4) & (ma.num < 7)
```

[,1]	[,2]	[,3]
[,1] FALSE	FALSE	FALSE
[,2] FALSE	TRUE	TRUE
[,3] FALSE	FALSE	FALSE

## 2.5. Asistente de Ayuda en R

R brinda un asistente de ayuda, en caso de desconocer el uso o sintaxis de una función específica, para hacer uso de esta ayuda, solo basta con ingresar el signo de interrogación (?) seguido de la función, como por ejemplo, ?sqrt ingresando de esta forma en R, proporcionará una vista en el explorador predeterminado, permitiendo observar información detallada de la función, sintaxis, uso e ilustrativos ejemplos.

## Capítulo 3

### 3. Estadística Descriptiva

La estadística descriptiva es una rama de la estadística, encargada del análisis y representación de series de datos, como, por ejemplo, notas de estudiantes, edades de una determinada población, temperaturas registradas de una región, etc. La estadística descriptiva puede ser aplicada a todos los campos de la investigación, producción, medicina, etc. Pudiendo extraer conclusiones a partir de ciertos procesos de datos, logrando obtener, analizar, elaborar y simplificar información, para luego interpretar de forma más rápida y cómoda, el comportamiento de todas las variables (cuantitativa o cualitativas) involucradas en el estudio.

#### 3.1. Tablas de frecuencia

A continuación, será tomada como ejemplo, la muestra de las edades de un grupo de 22 estudiantes de un colegio, en el que se tiene como resultado: {15, 16, 15, 17, 16, 18, 17, 17, 15, 16, 16, 17, 18, 1,6 16, 15, 18, 18, 16, 15, 15, 16}.

### 3.1.1. Frecuencia Absoluta

Con los datos obtenidos se hará el cálculo de la frecuencia absoluta de cada valor de la muestra, con la función “table ()”:

```
> #frecuencia absoluta:

> edades = c (15, 16, 15, 17, 16, 18, 17, 17, 15, 16, 16, 17, 18, 16, 16, 15, 18, 18, 16, 15,
15, 16)

> edades

[1] 15 16 15 17 16 18 17 17 15 16 16 17 18 16 16 15 18 18 16 15
15 16

> table (datos)

15 16 17 18

6 8 4 4
```

### 3.1.2. Frecuencia relativa

Ya calculada la frecuencia absoluta de la muestra de las edades de los estudiantes, se hará a continuación, el cálculo de la frecuencia relativa, dividiendo la función “table ()” entre la longitud del conjunto de los datos, que se obtiene con la función “length ()”.

> #Frecuencia relativa de cada valor de la variable edades:

```
> table(edades) / length(edades)
```

```
edades
```

```
15      16      17      18
0.2727273 0.3636364 0.1818182 0.1818182
```

> #Otra forma de hallar la frecuencia relativa de cada valor de la variable edades:

```
> table (edades) / sum (table (edades))
```

```
15      16      17      18
0.2727273 0.3636364 0.1818182 0.1818182
```

### 3.1.3. Suma de frecuencias

Ahora se va a calcular la frecuencia absoluta y el total de la suma de las frecuencias de cada uno de los datos, de la muestra, para esto se usará la función “addmargins ()”.

```
> addmargins (table (edades))
```

```
Edades
```

```
15 16 17 18 Sum
6  8  4  4  22
```

### 3.1.4. Frecuencia absoluta acumulada

Se procede a continuación a hacer el cálculo de la frecuencia absoluta acumulada y el total de la variable.

```
> # Frecuencia absoluta acumulada, total de la variable edades
```

```
> cumsum (table (edades))
```

```
15 16 17 18
```

```
6 14 18 22
```

### 3.1.5. Frecuencia relativa acumulada

Por último, se hará el cálculo de la frecuencia relativa acumulada.

```
> # Frecuencia relativa acumulada
```

```
> cumsum (table (edades) / length (edades))
```

```

      15          16          17          18
0.2727273 0.6363636 0.8181818 1.0000000
```

### 3.1.6. Función Scan ()

Después de haber hecho los cálculos de la frecuencia absoluta, frecuencia absoluta acumulada, frecuencia relativa y la frecuencia relativa acumulada, se calculará las medidas de dispersión, para esto, serán manejados nuevos datos, que representarán los pesos de un grupo de estudiantes. Con la ayuda de la función “scan ()”, se ingresarán cada uno de los valores respectivos, y en el momento de terminar el ingreso, se presiona dos veces la tecla “Enter”, para el uso de ésta función.

```
> pesos = scan()
```

```
1: 70
```

```
2: 65
```

```
3: 63
```

```
4: 65
```

```
5: 70
```

```
6: 63
```

```
7: 80
```

```
8: 80
```

```
9: 65
```

```
10: 70
```

```
11:
```

Read 10 items

```
> pesos
```

```
[1] 70 65 63 65 70 63 80 80 65 70
```

### 3.1.7. Función Summary ()

Una vez que se tiene las variables cuantitativas de la muestra, se hará un resumen estadístico de las variables, para esto, se empleará la función “summary”, que realiza esta tarea. La función mostrará el valor mínimo, el primer cuartil, la mediana, la media, el tercer cuartil y el valor máximo, como se muestra a continuación:

```
> # valor mínimo, primer cuartil, mediana, media, tercer cuartil y valor máximo
```

```
> pesos
```

```
[1] 70 65 63 65 70 63 80 80 65 70
```

```
> summary (pesos)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
63.0	65.0	67.5	69.1	70.0	80.0



### 3.1.8. Función `fivenum ()`

Otra función muy útil para resúmenes estadísticos de las variables cuantitativas, es “`fivenum()`”, ésta calcula 5 elementos elementales, que permiten hacer observaciones del comportamiento de la distribución de los datos.

```
> # Valor mínimo, primer cuartil, mediana, tercer cuartil y valor máximo
```

```
> fivenum(pesos)
```

```
[1] 63.0 65.0 67.5 70.0 80.0
```

### 3.1.9. Función `Stat.desc ()`

La función “`stat.desc ()`”, también realiza cálculos para hacer un resumen estadístico, dando como resultado 14 datos (número de variables, número de valores nulos, número de valores faltantes, valor mínimo, suma total de los valores, media, mediana, etc.), que comprende las medidas de tendencia central y de dispersión para la variable con contenidos numéricos. Para usar esta función es requerido instalar el paquete “`pastecs`”, esto se realiza con la instrucción “`install.packages('pastecs')`”, una vez instalado, se podrá dar uso haciendo su llamado con instrucción “`library( "pastecs" )`”, como se muestra a continuación.

```
> install.packages('pastecs')
```

```
> install.packages('pastecs')
```

```
> stat.desc(edades)
```

nbr.val	nbr.null	nbr.na	min	max	range
22.00000000	0.00000000	0.00000000	15.00000000	18.00000000	3.00000000
sum	median	mean	SE.mean	CI.mean.0.95	var
358.00000000	16.00000000	16.27272727	0.22964134	0.47756530	1.16017316
std.dev	coef.var				
1.07711335	0.06619132				

## 3.2. Medidas de tendencia central

### 3.2.1. Media aritmética:

Dado un conjunto de  $n$  números  $\{a_1, a_2, a_3, \dots, a_n\}$ , se define la media aritmética como:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n a_i = \frac{a_1 + a_2 + \dots + a_n}{n}$$

Por ejemplo, la media aritmética de los valores  $\{3,4,5\}$  es igual a:

$$\bar{x} = \frac{3 + 4 + 5}{3} = 4$$

En lenguaje R, la media se puede calcular por medio del comando “mean ()”, como se ve a continuación:

```
> val = c (3, 4, 5)
```

```
> mean (val)
```

```
[1] 4
```

Se puede observar en el ejemplo anteriormente expuesto, que para el cálculo de la media se puede realizar por medio de la instrucción “mean ()”, a partir de un conjunto de datos creado, declarando una variable como vector “val = c (3, 4, 5)”. A continuación, se dará uso a la instrucción “data.entry ()”, ésta permite ingresar los valores por medio de una ventana que se abrirá, tipo “hola de cálculo”, para ingresar los valores deseados, por medio de una interfaz.

```
> # Se introducirán los datos por medio de la función data.entry
```


```
> # Primero se crea un vector vacío
```

```
> x = c (0)
```

```
> # Ahora se ejecuta la instrucción usando el vector declarado
```

```
> data.entry(x)
```

	x	var2	var3	var4	var5
1	0				
2					
3					



	x	var2	var3	var4	var5
1	3				
2	4				
3	5				

Una vez ingresados los datos en la tabla que proporciona la función “data.entry ()”, se imprime el vector, con el fin de verificar que los cambios hayan sido guardados.

```
> #Imprimir vector y después calcular la media
```

```
> x
```

```
[1] 3 4 5
```

```
> mean(x)
```

```
[1] 4
```

### 3.2.2. Media geométrica

La media geométrica es una forma más en la que se describe el valor central de un conjunto de datos, y está definida como:

$$\log \bar{x} = \frac{1}{n} \sum_{i=1}^n \log x_i$$

```
> #La media geométrica es hallada, calculando la media del logaritmo.
```

```
> val = c (12,16,19,11,15,17)
```

```
> val
```

```
[1] 12 16 19 11 15 17
```

```
> mean (log (val))
```

```
[1] 2.690182
```

El cálculo anterior corresponde a la media geométrica expresada en escala logarítmica, en caso de ser deseado convertir a la escala original, es empleada la función de exponencial anti-logarítmica, de la siguiente manera:

```
> exp(mean(log(val)))
```

```
[1] 14.73436
```

### 3.2.3. Moda

La moda en estadística, es el mayor valor de frecuencia en una distribución de datos cuantitativos. Está representado de la siguiente forma, en donde los datos cuando se encuentran agrupados, se puede describir que la moda es un punto que divide al intervalo modal en dos partes, que serían (p) y (c-p).

$$\frac{p}{c - p} = \frac{n_i - n_{i-1}}{n_i - n_{i+1}}$$

Como en lenguaje R no existe una función para el cálculo de la moda, debemos instalar la librería “modeest”, y después usar la función “mfv ()”. Para este ejemplo, seguiremos usando la variable “edades”.

```
> install.packages('modeest')
```

```
> library( "modeest" )
```

```
> mfv(edades)
```

```
[1] 16
```

Otra manera de calcular la moda, sin el uso de una librería, es de la siguiente manera:

```
> table (edades)
```

```
edades
```

```
15 16 17 18
```

```
6 8 5 4
```

```
> max (table (edades))
```

```
[1] 8          #Teniendo este valor, se concluye que el resultado es 16
```

#### 3.2.4. Mediana

Puede definirse como un indicador estadístico, en donde divide la información en dos partes iguales. Esta puede ser usada cuando la media aritmética tiene datos muy elevados que pueden llegar a sensibilizar el promedio, recordando que la media aritmética es sensible a valores extremos, llegado el caso de que esto suceda el mejor indicador es la mediana. Para éste calcula se usará la instrucción “median ()”, como se muestra a continuación:

```
> edades
```

```
[1] 15 16 15 17 16 18 17 17 15 16 16 17 18 16 16 15 18 18 16 15 15 16 17
```

```
> median(edades)
```

```
[1] 16
```

Solo como ejercicio comparativo, se usará la instrucción “summary ()”, esta fue empleada anteriormente. Ahora se volverá a usar, con el fin de observar que las dos maneras de hacer el cálculo de la mediana, coincidan.

```
> summary(edades)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
15.0	15.5	16.0	16.3	17.0	18.0

### 3.2.5. Cuantiles

Los cuantiles son aquellos valores que representan los porcentajes de división, en partes iguales de una distribución, es decir, son intervalos que representan una misma cantidad de valores. Estos porcentajes pueden variar, donde los más usados son los percentiles, cuartiles y deciles.

Los percentiles son aquellos 99 valores en la que se divide un conjunto de datos en 100 partes iguales. Esto quiere decir que los percentiles son los valores que corresponden al 1%, 2%, hasta el 99% de los datos.

Los cuartiles se refieren a los tres valores de un conjunto de datos ya divididos en partes iguales, estos estarán ordenados en cuatro partes iguales, normalmente estas partes son representadas por  $Q_1$ ,  $Q_2$  y  $Q_3$ . Cada una de ellas son los valores correspondientes al 25%, 50% y al 75% de los datos. También el valor  $Q_2$  es coincidencial al valor de la mediana.

Los deciles corresponden a los nueve valores, en que se divide un conjunto de datos, esta división se hace en diez partes iguales. Los deciles representan los valores del 10%, 20% ... hasta el 90% de los datos. El quinto decil, que por lo general es representado por “D<sub>5</sub>” es coincidencial con la mediana.

> # Se calculará el primer y tercer cuartil

> # el primer cuartil deja el 25% a la izquierda

> quantile (edades, probs=.25)

25%

15.5

> # El tercer cuartil deja el 75% a la izquierda

> quantile (edades, probs=.75)

75%

17

> # El tercer decil deja el 30% a la izquierda

> quantile (edades, probs=.3)

30%

16

> # El octavo decil deja el 80% a la izquierda

> quantile (edades, probs=.8)

80%



17

> # La octogésima primera parte deja el 81% a la izquierda

> quantile (edades, probs=.81)

81%

17

### 3.3. Medidas de Dispersión

#### 3.3.1. Rango

En la estadística descriptiva el rango estadístico o también llamado recorrido estadístico, es el intervalo de menor tamaño de un conjunto de datos, se calcula tomando el valor mínimo y se resta con el valor máximo de los datos. Obteniendo así, una idea de la dispersión de los datos.

$$\text{Rango} = \max(x) - \min(x)$$

> edades = c (15, 16, 15, 17, 16, 18, 17, 17, 15, 16, 16, 17, 18, 16, 16, 15, 18, 18, 16, 15,  
15, 16)

> range (edades)

[1] 15 18

> # Veamos que coinciden con los resultados de summary

```
> summary(edades)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 15.0   15.5   16.0   16.3   17.0   18.0
```

```
> # Lo creamos nosotros
```

```
> min (edades)
```

```
[1] 15
```

```
> max (edades)
```

```
[1] 18
```

```
> rango = c (min (edades), max (edades))
```

```
> rango
```

```
[1] 15 18
```

### 3.3.2. Rango Medio

El rango medio, es la media del valor mínimo y el valor máximo de un conjunto de valores, o en otras palabras es la mitad del camino entre los valores extremos (mínimo y máximo) del conjunto de datos.

$$\text{Rango Medio} = \frac{\max(x) + \min(x)}{2}$$

```
> edades = c (15, 16, 15, 17, 16, 18, 17, 17, 15, 16, 16, 17, 18, 16, 16, 15, 18, 18, 16, 15,  
15, 16)
```

```
> range (edades)
```

```
[1] 15 18
```

```
> range_medio = range (edades)
```

```
> mean (range_medio)
```

```
[1] 16.5
```

### 3.3.3. Rango Intercuartílico

El rango intercuartílico o rango intercuartil, es una medida de dispersión estadística, y esta resulta de la diferencia entre el tercer cuartil ( $Q_3$ ) y el primer cuartil ( $Q_1$ ). Este rango es usado para comparar distribuciones de la misma variable o también para observar la variabilidad de una variable.

$$RQ = Q_3 - Q_1$$

```
> # Cálculo del rango intercuartil
```

```
> quantile(edades,.75)-quantile(edades,.25)
```

```
75%
```

```
1.5
```

### 3.3.4. Varianza

La varianza es una medida de dispersión, que representa la media aritmética del cuadrado de las desviaciones respecto a la media de una distribución estadística. La varianza se representa por " $\sigma^2$ ".

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{N}$$

```
> edades = c (15, 16, 15, 17, 16, 18, 17, 17, 15, 16, 16, 17, 18, 16, 16, 15, 18, 18, 16, 15,
15, 16)
```

```
> var (edades)
```

```
[1] 1.130435
```

```
> sum ((edades - mean (edades))^2) / (length (edades))
```

```
[1] 1.081285
```

```
> sum ((edades - mean (edades))^2) / (length (edades)-1)
```

```
[1] 1.130435
```

```
> #Si se quiere calcular la poblacional multiplicamos por n-1/n
```

```
> n = length (edades)
```

```
> var (edades) * (n-1) / n
```

```
[1] 1.081285
```

### 3.3.5. Desviación Típica

La desviación típica o también llamada desviación estándar, es una medida de dispersión para variables cuantitativas o variables racionales, esta representa la raíz cuadrada de la varianza. En otras palabras, es la raíz cuadrada de la media de los cuadrados de las puntuaciones de desviación y es representada por " $\sigma$ ".

Cómo lenguaje R no calcula directamente la varianza poblacional, sino que calcula la muestral, se debe poner como divisor (n - 1) en vez de "n". Y ahora para calcularla se usará la función "sd()".

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{N}}$$

```
> #Calculo Muestral
```

```
> sd (edades)
```

```
[1] 1.063219
```

```
> sqrt (var (edades))
```

```
[1] 1.063219
```

```
> #Calculo Varianza Poblacional
```

```
> sd (edades) * (n-1) / n
```

```
[1] 1.016992
```

```
> sqrt (var (edades)) * (n-1) / n
```

```
[1] 1.016992
```

## Capítulo 4

## 4. Ejercicios propuestos

Ejercicio 1:

Se lanza 25 veces un dado, obteniendo los siguientes valores:

5	1	6	4	2
6	4	3	2	6
1	6	4	3	2
5	6	3	2	2
1	5	4	3	3

Los siguientes ejercicios deben ser hechos en lenguaje R:

- Ingresar los datos anteriores del dado en un vector
- Calcular la frecuencia absoluta
- Calcular la frecuencia relativa
- Calcular la frecuencia acumulada
- Calcular la frecuencia relativa acumulada

## Ejercicio 2:

En un colegio fueron encuestados varios alumnos, para registrar su mes de nacimiento, como resultado se obtuvo que: 5 nacieron en el mes 1 (enero), 6 nacieron en el mes 2 (febrero), 3 nacieron en el mes 3 (marzo), 7 nacieron en el mes 4 (abril), 4 nacieron en el mes 5 (mayo), 2 nacieron en el mes 6 (junio), 3 nacieron en el mes 7 (julio), 2 nacieron en el mes 8 (agosto), 3 nacieron en el mes 9 (septiembre), 1 nació en el mes 10 (octubre), 2 nacieron en el mes 11 (noviembre), 2 nacieron en el mes 12 (diciembre). Construir la tabla de frecuencias y hacer los cálculos en lenguaje R.

- a) Construir la siguiente tabla de frecuencias.

Datos (Meses)	Frecuencia Absoluta (Alumnos)	Frecuencia Relativa	Frecuencia Absoluta Acumulada	Frecuencia Relativa Acumulada
Enero				
Febrero				
Marzo				
Abril				
Mayo				
Junio				
Julio				
Agosto				
Septiembre				
Octubre				
Noviembre				
Diciembre				
$\Sigma$				

- b) Almacenar los datos de los alumnos en un vector en lenguaje R.
- c) Realizar en R los cálculos respectivos para hallar las frecuencias solicitadas en la tabla.

## Ejercicio 3:

Un grupo de 45 estudiantes del departamento de Risaralda nacieron en las ciudades de: 8 en Pereira, 9 en la Virginia, 6 en Santa Rosa, 10 en Dosquebradas, 9 en Santuario y 6 en La Celia. Construir la tabla de frecuencias y hacer los cálculos en el lenguaje R.

Datos (Ciudades)	Frecuencia Absoluta (Estudiantes)	Frecuencia Relativa	Frecuencia Absoluta Acumulada	Frecuencia Relativa Acumulada
Pereira				
Virginia				
Santa Rosa				
Dosquebradas				
Santuario				
Celia				

- Ordenar los datos
- Almacenar los datos de los estudiantes en un vector en lenguaje R.
- Calcular la frecuencia absoluta.
- Calcular la frecuencia relativa.
- Calcular la frecuencia acumulada.
- Calcular la frecuencia relativa acumulada.



## Ejercicio 4.

A continuación, se darán los datos de edad de varias personas que oscilan entre los 30 y 40 años:

Valores (Edad)	Frecuencia Absoluta	Frecuencia Relativa	Frecuencia Absoluta Acumulada	Frecuencia Relativa Acumulada
30	3			
31	6			
32	6			
33	3			
34	3			
35	4			
36	3			
37	4			
38	1			
39	4			
40	3			
$\Sigma$				

Calcular:

- Ordenar las edades.
- Frecuencia absoluta.
- Frecuencia relativa.
- Frecuencia absoluta acumulada.
- Frecuencia relativa acumulada.
- Tamaño muestral.
- La Media de las edades.
- La mediana de las edades.
- La Varianza de las edades.

- j) La desviación típica.
- k) El rango intercuartil.
- l) Los percentiles del 25%, 50% y 75%.
- m) La media recortada al 10% y al 5%.

### Ejercicio 5

Las estaturas de los alumnos de una clase oscilan entre 1.20 m y 1.30 m, dibuje la tabla de frecuencias y calcule sus datos en el lenguaje R.

Datos (Altura en metros)	Frecuencia Absoluta	Frecuencia Relativa	Frecuencia Absoluta Acumulada	Frecuencia Relativa Acumulada
1.20	1			
1.21	4			
1.22	4			
1.23	2			
1.24	1			
1.25	2			
1.26	3			
1.27	3			
1.28	4			
1.29	3			
1.30	3			
$\Sigma$				

Calcular:

- a) Frecuencia absoluta.
- b) Frecuencia relativa.
- c) Frecuencia absoluta acumulada.
- d) Frecuencia relativa acumulada.

- e) La media de las alturas.
- f) La mediana de las edades.
- g) La varianza de las edades.
- h) La desviación típica.
- i) El rango intercuartil.
- j) La desviación estándar.
- k) Los percentiles del 25%, 50% y 75%.
- l) La media recortada al 10% y al 5%.
- m) Graficar el diagrama de tallo y hojas con la función “stem ()” de las alturas.
- n) Graficar el diagrama de caja y bigotes de las edades con la función “boxplot ()”.
- o) Graficar el histograma de las alturas con la función “hist ()”

## 5. Bibliografía

- Anderson, David R., Sweeney, Dennis J., Williams, Thomas A., “Estadística para Administración y Economía” 10ª Edición, Cengage Learning Editores, 2008.
  
- Antonio José Sáez Castillo, “Métodos Estadísticos con R y R Commander”, versión 2.1, julio de 2010. Dpto de Estadística e Investigación Operativa Universidad de Jaén.
  
- Arriaza A. J., Fernández F. y col. (2008), “Estadística Básica con R y R-Commander”, Servicio de Publicaciones de la Universidad de Cádiz
  
- Carmona F., Ejercicios de Análisis de la Varianza con R
  
- Crawley M. J. (2007), The R Book, Hardcover
  
- Douglas Montgomery, George Runner “Probabilidad y Estadística Aplicadas a la Ingeniería”, Mc GrawHill, 1997. Signatura Biblioteca 519M787 (préstamo de un día para otro).
  
- González A. y González S. (2000), Introducción a R: Notas sobre R: Un entorno de programación para Análisis de Datos y Gráficos, R Development Core Team.

- John E. Freund, Ronald E. Walpole; “Estadística Matemática con aplicaciones” 4<sup>a</sup> Edición. P.H.H. 1990.
  
- José M. Contreras, Elena Molina, Pedro Arteaga Cezón, “Introducción a la programación estadística con r para profesores”, Esta obra forma parte de los proyectos: SEJ2007-60110/EDUC (MEC-FEDER) y EDU2010-14947 (subprograma EDUC); con la colaboración de la beca FPI BES-2008-003573 y la beca FPU AP2007-03222.
  
- Martínez Bencardino, Ciro “Estadística y muestreo”, ECOE Ediciones, 12<sup>a</sup> edición, 2005.
  
- Miller y Freund “Probabilidad y estadística para ingenieros”, - 5<sup>a</sup> Edición, PrenticeHall HISPANOAMERICANA, S.A, 1997. Signatura Biblioteca 519M648 (préstamo ocho días).
  
- Murray R. Spiegel; “Teoría y Problemas Resueltos” – Serie Schaum, Mc. Graw–Hill, 1987.
  
- Paradis E. (2002), R para Principiantes, Institut des Sciences de l'Évolution Universit Montpellier II

- Paúl Meyer; “Probabilidad y Aplicaciones Estadísticas” Edición revisada Addison-Wesley-Iberoamericana 1992.
  
- Saulo de Jesús Torres Rengifo (2013), “VerWeb - Metodología Para La Evaluación De Accesibilidad Web Aplicable A La Realidad De Colombia.”. Tesis Doctoral, Universidad Pontificia De Salamanca, España.
  
- Walpole, Ronald E., Myers, Raymond H., Myers, Sharon L., “Probabilidad y Estadística para Ingenieros”. – 6ª Ed. México: PrenticeHall Hispanoamérica, S.A, 2007. Signatura Biblioteca 519W218.