



INGENIERÍA INDUSTRIAL
UNIVERSIDAD DE CHILE



fcfm

Auxiliar 1: Introducción a R

IN3242 – Estadística

Profesor: Raimundo Undurraga

Auxiliares: Brandon Galarza – Camila Jáuregui – Leonardo Meneses

Francisca Monetta – Matías Reyes – Bastián Urzúa – Antonia Villegas

¿Qué es R? ¿Cuáles son sus usos?

- R es un lenguaje de programación orientado a Objetos Simplicidad y flexibilidad
 - RStudio es un IDE (entorno de desarrollo integrado) que utiliza R para desarrollar programas estadísticos



- R es el motor (lenguaje) y RStudio es el volante (IDE)
- Utilizado para la computación estadística, creación de gráficas y figuras
- Análisis y visualización de datos
- Se pueden combinar distintas funciones estadísticas para análisis más complejos

Objetos

- Los objetos tienen 2 atributos intrínsecos:
 - 2) Longitud: Número de elementos en el objeto
 - 1) Tipo: numérico / entero / carácter / complejo / lógico

objeto	tipos	varios tipos posibles en el mismo objeto?
vector	numérico, carácter, complejo o lógico	No
factor	numérico o carácter	No
arreglo	numérico, carácter, complejo o lógico	No
matriz	numérico, carácter, complejo o lógico	No
data.frame	numérico, carácter, complejo o lógico	Si
ts	numérico, carácter, complejo o lógico	Si
lista	numérico, carácter, complejo, lógico función, expresión, ...	Si

Tipos de datos básicos en R

- Valores decimales son llamados numerics
- Números enteros son llamados integers. Los integers también son numerics
- Valores tipo texto (o string) son llamados character
- Valores factor guardan información categórica. Ej: altura (alto / medio / bajo)
- Valores booleanos (TRUE / FALSE) son denominados logical

Función	Tipo que verifican
<code>is.integer()</code>	Entero
<code>is.numeric()</code>	Numerico
<code>is.character()</code>	Cadena de texto
<code>is.factor()</code>	Factor
<code>is.logical()</code>	Lógico
<code>is.na()</code>	NA
<code>is.null()</code>	NULL

Tipos de datos básicos en R

- Con la función `class()` se puede chequear el tipo de dato o vector
 - Es importante ser consistentes con los distintos tipos de datos, por ejemplo, no podemos sumar `6 + "cinco"`

Conversión a	Función	Reglas
numérico	<code>as.numeric</code>	<code>FALSE</code> → 0 <code>TRUE</code> → 1 "1", "2", ... → 1, 2, ... "A", ... → NA
lógico	<code>as.logical</code>	0 → <code>FALSE</code> otros números → <code>TRUE</code> "FALSE", "F" → <code>FALSE</code> "TRUE", "T" → <code>TRUE</code> otros caracteres → NA
caracter	<code>as.character</code>	1, 2, ... → "1", "2", ... <code>FALSE</code> → "FALSE" <code>TRUE</code> → "TRUE"

Algunas operaciones básicas

- En su forma más básica, R puede ser usado como una calculadora simple
 - Se pueden manipular o modificar los objetos con operadores y funciones

		Operadores			
Aritméticos		Comparativos		Lógicos	
+	adición	<	menor que	! x	NO lógico
-	substracción	>	mayor que	x & y	Y lógico
*	multiplicación	<=	menor o igual que	x && y	id.
/	división	>=	mayor o igual que	x y	O lógico
^	potencia	==	igual	x y	id.
%%	módulo	!=	diferente de	xor(x, y)	O exclusivo
%/ %	división de enteros				

- Módulo %% devuelve el resto de la división del número a la izquierda, por el número a su derecha. Ej: 5 %% 3 es igual a 2, 28 %% 6 es igual a 4

Funciones I: Más utilizadas

Función	Detalle
<i>sum(x)</i>	Entrega la suma los elementos de x
<i>prod(x)</i>	Entrega el producto de los elementos de x
<i>sqrt(x)</i>	Entrega la raíz cuadrada de x
<i>log(x)</i>	Entrega el logaritmo natural de x
<i>exp(x)</i>	Entrega el exponencial de x
<i>max(x)</i>	Entrega el máximo de los valores en x
<i>min(x)</i>	Entrega el mínimo de los valores en x
<i>range(x)</i>	Entrega el rango de x o $c(\min(x), \max(x))$
<i>length(x)</i>	Entrega la cantidad de elementos de x
<i>round(x,n)</i>	Entrega el valor de x redondeado a n dígitos
<i>mean(x)</i>	Entrega el promedio de los valores contenidos en x
<i>median(x)</i>	Entrega la mediana de los elementos de x
<i>sd(x)</i>	Entrega la desviación estándar de los valores en x
<i>var(x) / cov(x)</i>	Entrega la varianza (o matriz de var-cov) de los elementos de x

Funciones II: Distribuciones

Distribución/función	función
Gausse (normal)	<code>rnorm(n, mean=0, sd=1)</code>
exponencial	<code>rexp(n, rate=1)</code>
gamma	<code>rgamma(n, shape, scale=1)</code>
Poisson	<code>rpois(n, lambda)</code>
Weibull	<code>rweibull(n, shape, scale=1)</code>
Cauchy	<code>rcauchy(n, location=0, scale=1)</code>
beta	<code>rbeta(n, shapel, shape2)</code>
'Student' (t)	<code>rt(n, df)</code>
Fisher-Snedecor (F)	<code>rf(n, df1, df2)</code>
Pearson (χ^2)	<code>rchisq(n, df)</code>
binomial	<code>rbinom(n, size, prob)</code>
geométrica	<code>rgeom(n, prob)</code>
hypergeométrica	<code>rhyper(nn, m, n, k)</code>
logística	<code>rlogis(n, location=0, scale=1)</code>
lognormal	<code>rlnorm(n, meanlog=0, sdlog=1)</code>
binomial negativa	<code>rnbinom(n, size, prob)</code>
uniforme	<code>runif(n, min=0, max=1)</code>
Estadístico de Wilcoxon's	<code>rwilcox(nn, m, n), rsignrank(nn, n)</code>

Funciones III: Inferencia Estadística

Función	Detalle
<i>t.test()</i>	Inferencia sobre la media en una población
<i>qqplot()</i>	Evaluar gráficamente la hipótesis de que la variable es normal
<i>shapiro.test()</i>	Evaluar la normalidad de la variable a través de un test formal
<i>wilcox.test()</i>	Contrasta si dos muestras proceden de la misma distribución
<i>sigma.test()</i>	Contrastar el valor de la varianza de una variable distribuida normal
<i>var.test()</i>	Lleva a cabo un test F para comparar las varianzas de dos poblaciones

Paquetes más utilizados

Instalar y cargar paquetes:

1) `install.packages("nombre_paquete")`

2) `library(nombre_paquete)`

Paquete	Detalle
<i>readxl</i>	Librería para cargar bases de datos de Excel
<i>dplyr</i>	Gramática más flexible para manipulación de datos
<i>car</i>	Para usar el Factor Inflador de Varianza
<i>Rmisc</i>	Para usar función <code>summarySE()</code>
<i>ggplot2</i>	Librería para realizar visualizaciones
<i>xlsx</i>	Librería para cargar bases de datos de Excel

Gráficos

Función	Detalle
<i>plot(x)</i>	Gráfico de los valores de x (en el eje y) ordenados en el eje x
<i>plot (x,y)</i>	Gráfico bivariado de x (en el eje x) e y (en el eje y)
<i>boxplot(x)</i>	Gráfico tipo ' <i>box-and-whiskers</i> ', o bien, diagrama de caja y bigotes
<i>hist (x)</i>	Histograma de las frecuencias de x
<i>barplot (x)</i>	Histograma de los valores de x
<i>qqnorm (x)</i>	Cuartiles de x con respecto a lo esperado bajo una distribución normal
<i>qqline (x)</i>	Recta que pasa por los cuartiles de la distribución y los datos
<i>qqplot (x,y)</i>	Cuartiles de y con respecto a los cuartiles de x
<i>ggplot2 ()</i>	Combinación de elementos básicos en gráficos distintos mediante la sintaxis: <i>ggplot(data.frame, aes(x = variable)) + geom_forma()</i>