

Estadística con R en el Grado en Administración y Dirección de Empresas (Universidad de Murcia)

Molera Peris, L., Arnaldos García, F., Díaz Delfa, M.T., Faura Martínez, U., Parra Frutos, I., Pérez Castejón, J.J.

Universidad de Murcia

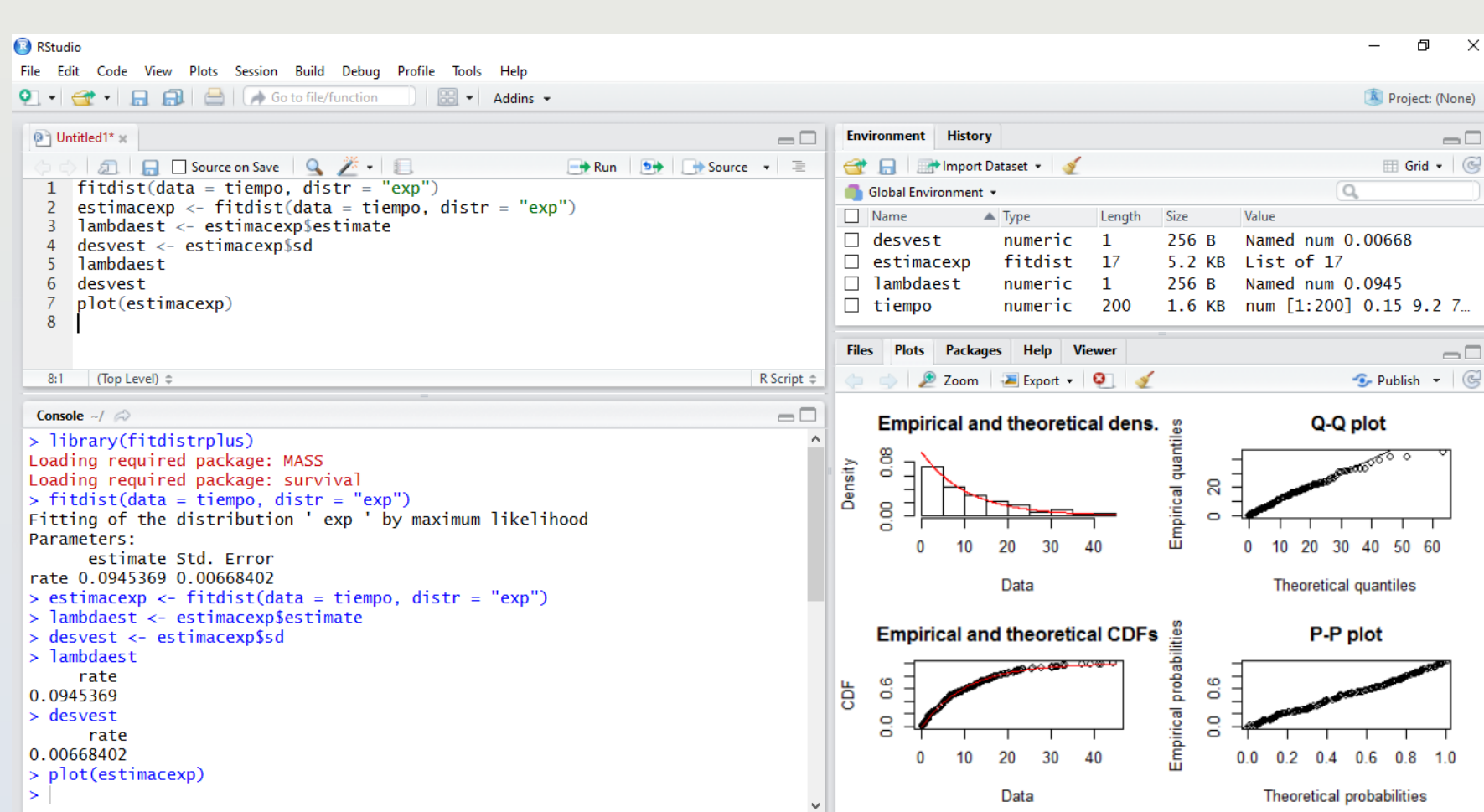
lmolera@um.es, arnaldos@um.es, mdd@um.es, faura@um.es, ipf@um.es, ijpc@um.es

Aspectos que han influido en la decisión de plantear las prácticas de Inferencia Estadística en el Grado en ADE con R: Programa de libre acceso, muy extendido en la comunidad científica, muy versátil, con interfaces sencillas como RStudio, rapidez en la obtención de gráficos y resultados, facilidad para el tratamiento eficiente de grandes cantidades de datos...

Operativa: Durante las clases se ha utilizado RStudio y se han elaborado ficheros html con RMarkdown para facilitar al alumno el uso de R y favorecer su autonomía en las sesiones. También se han realizado transparencias con Slidify, así como aplicaciones en Shiny para visualizar datos y reforzar conceptos.

Todos estos materiales están publicados en el sitio de la asignatura del Aula Virtual de la Universidad de Murcia. Además, como resultado de este trabajo está pendiente de publicación un OpenCourseWare en el portal de la Universidad de Murcia titulado **Curso Práctico de Inferencia Estadística con R**.

Uso de R y RStudio para la realización de prácticas



Materiales didácticos elaborados con R Markdown

Documentos html con prácticas resueltas

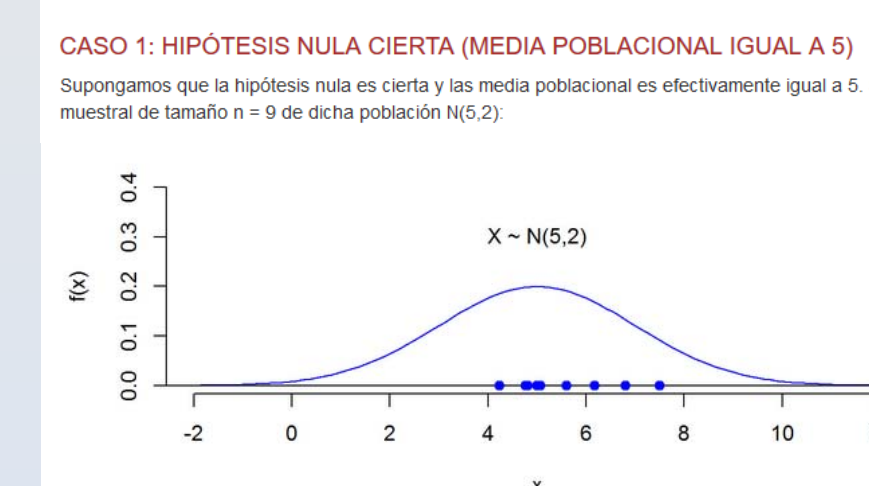
Se han elaborado siete unidades prácticas correspondientes a distintas temáticas de Inferencia Estadística, así como una unidad de introducción a R y RStudio en la que se abordan también algunos cálculos y gráficos básicos de Estadística Descriptiva. En todas ellas se ha utilizado un formato definido en un fichero CSS. Cada unidad comienza con una tabla de contenidos con enlaces, elaborado por el propio R Markdown, y termina con un glosario de funciones.

Documentos html con conceptos teóricos

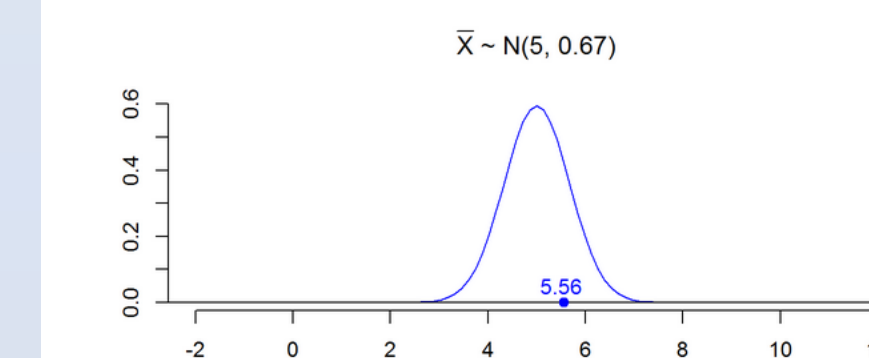
R no solo permite analizar datos sino también reforzar y visualizar conceptos que suelen resultar complicados para los estudiantes: distribución de estadísticos, contrastes de hipótesis, etc. Con este propósito se han elaborado algunos documentos html con R Markdown.

Contraste para una media de una población normal con varianza conocida

Sea una población normal X con varianza conocida (por ejemplo, igual a 4) y a partir de la información de una muestra de tamaño, por ejemplo, 9, queremos contrastar si el valor de la media de esa población es, por ejemplo, 5:
 $H_0: \mu = 5$
 $H_1: \mu \neq 5$



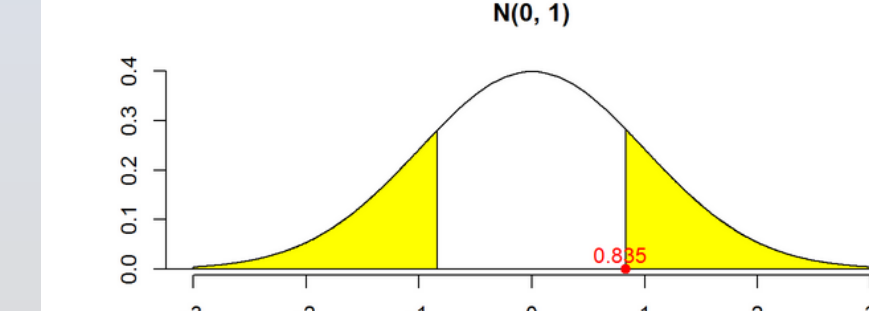
Las observaciones de esta muestra son: 4.77, 4.24, 4.82, 5.07, 5.7, 5.1, 0.18, 5.6, 6.81. El valor de la correspondiente media muestral, 5.56, será una realización muestral de la distribución $N(5, 2/9)$. Recuerde que, como la población es normal, la media muestral sigue una distribución también normal, con la misma media que la población pero con menor dispersión (la varianza será la de la población dividida por el tamaño de la muestra y, por tanto, más pequeña cuanto mayor sea el tamaño muestral).



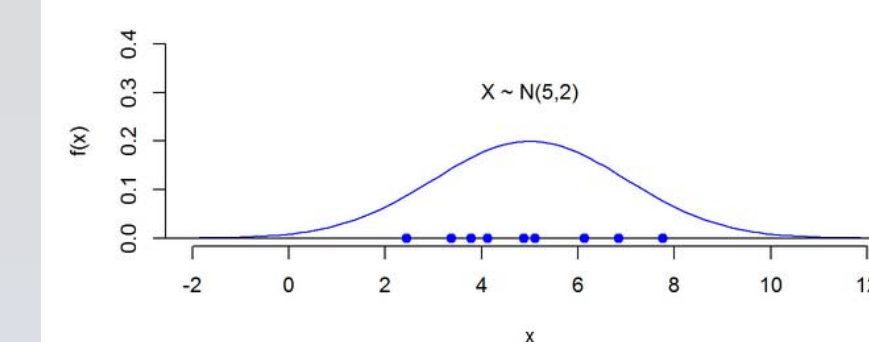
Si a partir de las observaciones de esta muestra, calculamos el valor del estadístico de contraste:

$$T = \frac{\bar{X} - \mu_0}{\sqrt{s^2/n}}$$

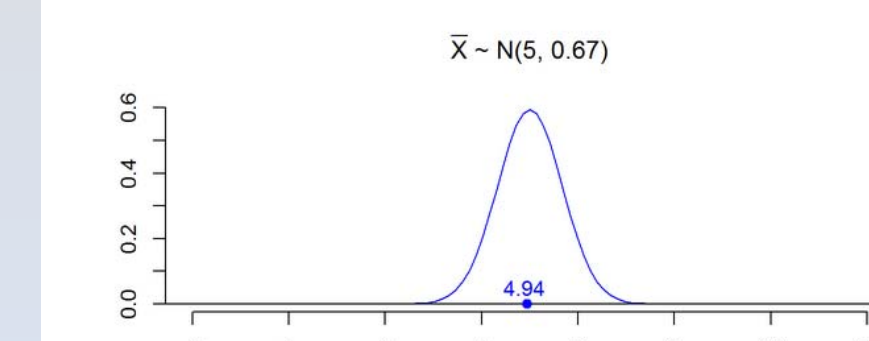
obtenemos el valor 0.936, que será una realización de la distribución $N(0, 1)$ puesto que la hipótesis nula es cierta. El p-valor es igual, entonces, a 0.40362 y coincide con el área coloreada en amarillo en el gráfico.



Si se dispusiera de otra realización muestral de tamaño $n = 9$ de una distribución $N(5, 2)$, los resultados serían más o menos similares a los anteriores. Es decir, el valor del estadístico de contraste sería siempre una realización de la distribución $N(0, 1)$, de modo que, si se establece 0.05 como nivel de significación, sólo en el 5% de los casos nos equivocamos rechazando la hipótesis nula (esto es, obteniéndamos un p-valor menor que 0.05). Por ejemplo, con esta otra realización muestral de tamaño 9 de la misma población,



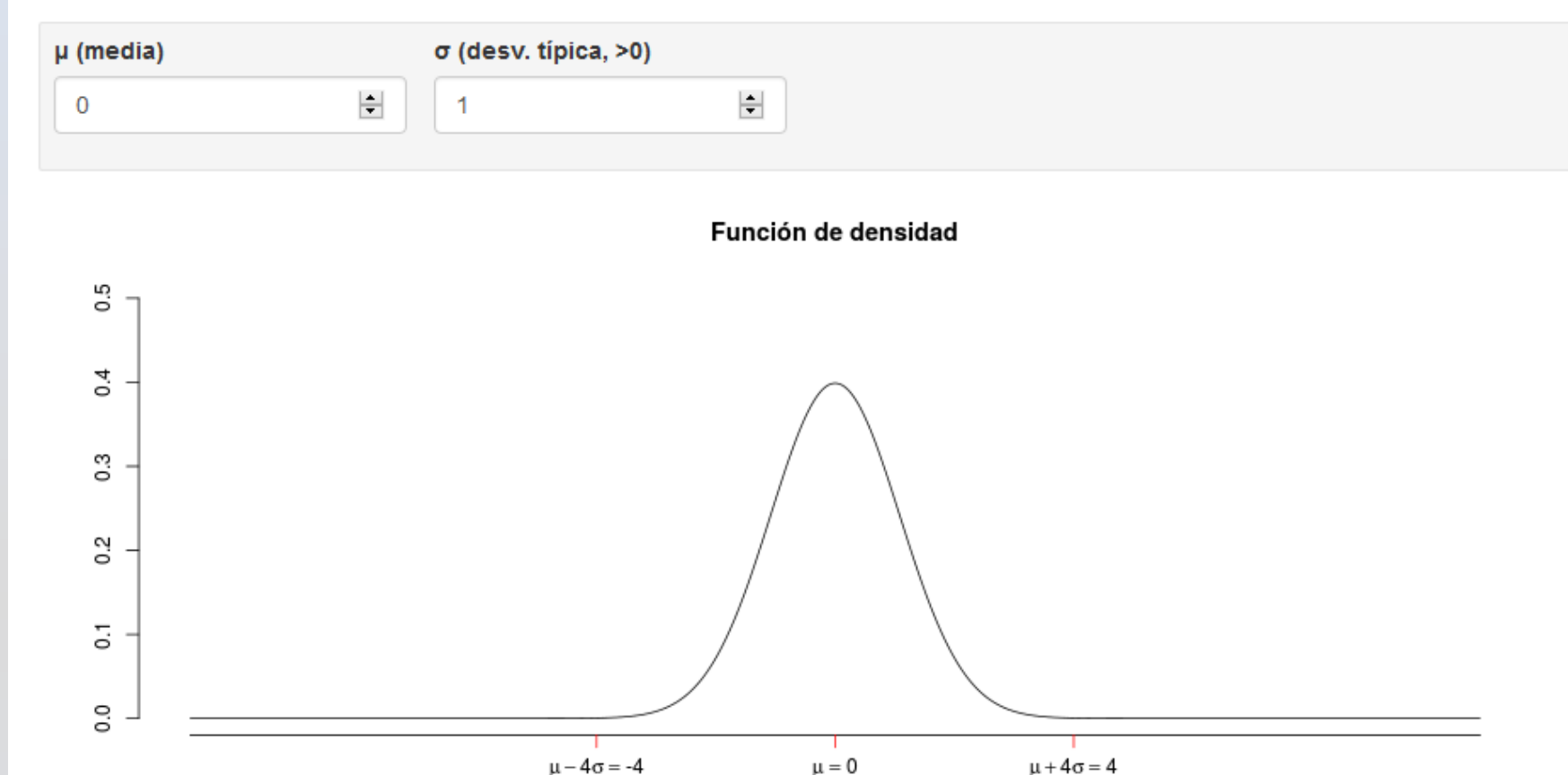
el valor de la media muestral sería ahora 4.94, también realización de la distribución $N(5, 2/9)$.



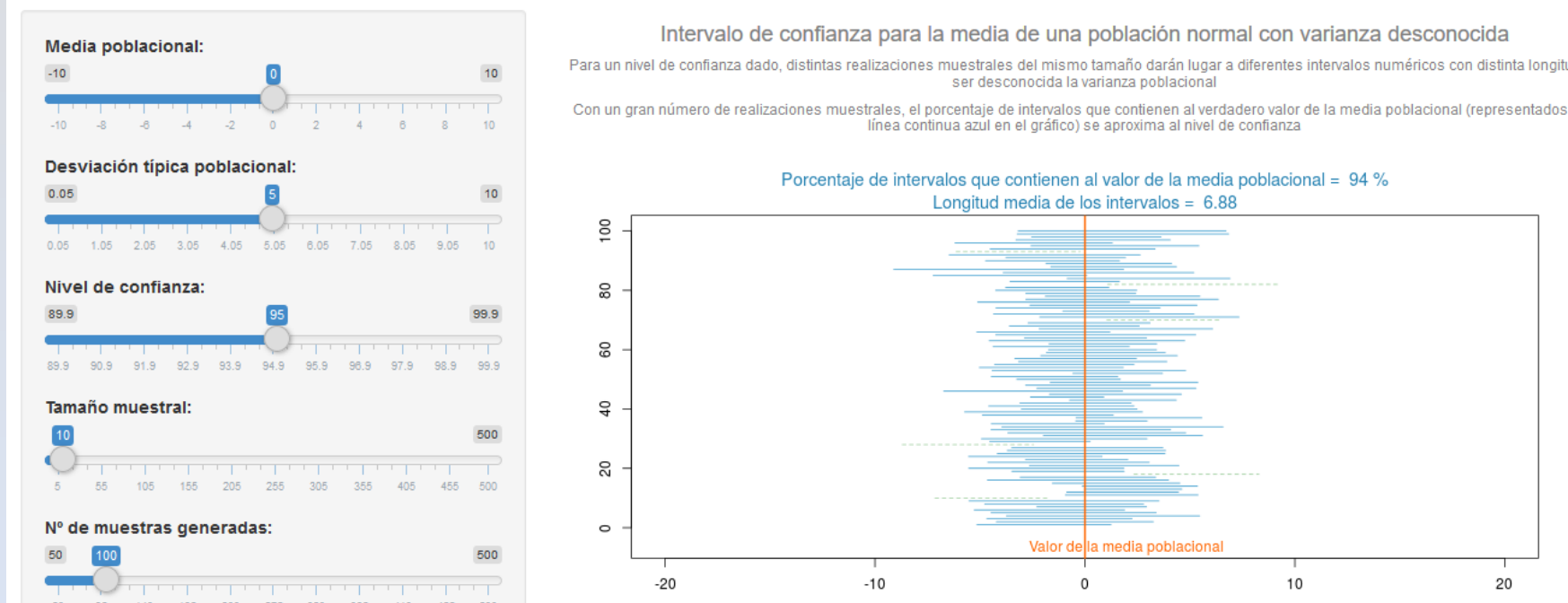
Aplicaciones interactivas diseñadas con Shiny y R Markdown

DISTRIBUCIÓN NORMAL $N(\mu, \sigma)$

El siguiente gráfico interactivo permite seleccionar valores para μ y σ , y observar cómo influyen en la forma de la función de densidad de una $N(\mu, \sigma)$. El lector debe deducir de la gráfica de dicha función, las propiedades probabilísticas que se acaban de comentar.



INTERVALOS DE CONFIANZA CON SIMULACIÓN (II)



El p-valor: concepto y cálculo



Unidad 2: Generación de números aleatorios y simulación del comportamiento muestral de estadísticos

- Instrucciones
- Sección 2.1. Simulación de experimentos aleatorios
 - Ejercicio 1. Lanzamiento de un dado
 - Ejercicio 2. Lanzamiento un dado trucado
 - Ejercicio 3. Extracción de bolas de una urna
 - Ejercicio 4. Extracción de una muestra procedente de una variable discreta
- Sección 2.2. Generación de números aleatorios
 - Ejercicio 5. Distribución continua
 - Ejercicio 6. Distribución discreta
- Sección 2.3. Comportamiento muestral de los estadísticos
 - Ejercicio 7. Media muestral en una población normal
 - Ejercicio 8. Cuasivarianza muestral en una población normal
- MATERIAL DE APOYO
- RECOPILEDARIO DE FUNCIONES UTILIZADAS
 - Funciones para la generación de números aleatorios

OCW Curso Práctico de Inferencia Estadística con R
M.T. Díaz, F. Arnaldos, U. Faura, L. Molera, I. Parra

Instrucciones

- Establezca un directorio de trabajo para la práctica. Para saber cuál es el directorio de trabajo actual use la función `getwd()`. Para cambiarlo desde `RStudio` puede pinchar en el siguiente menú: `Session/Session Working Directory/Choose Directory`.
- Guarde en ese directorio un R Script con los códigos de `R` utilizados en esta práctica. Para crear un R Script debe pinchar en menú `File/New File/R Script`.

Volver al principio

Sección 2.1. Simulación de experimentos aleatorios

Función a utilizar: `sample`

Su estructura es `sample(x, size, replace = FALSE, prob = NULL)` y los argumentos se explicarán a lo largo de los ejercicios de esta sección y en el recopilatorio de funciones al final de la unidad.

Ejercicio 1. Lanzamiento de un dado

Simule 10 lanzamientos de un dado.

Los posibles resultados del lanzamiento de un dado son los valores 1, 2, 3, 4, 5 y 6, por lo que escribimos 1:6 en el argumento `x` de la función `sample`. Además, como queremos realizar 10 lanzamientos con reemplazamiento, hay que asignar el valor 10 al argumento `size` y el valor `TRUE` al argumento `replace` (ya que `replace = FALSE` por defecto).

```
sample(1:6, 10, replace = TRUE)
## [1] 3 3 5 5 2 6 4 3 6 3
```

Nota: Observe que la muestra que ha obtenido no tiene por qué coincidir con la descrita en este ejercicio. Para que fuera la misma deberíamos haber utilizado el mismo valor inicial para arrancar el generador de números aleatorios. Este número se denomina semilla, y debe colocarse como argumento de la función `set.seed`. La utilización de la semilla será tratada en el ejercicio 6.

RECOPILEDARIO DE FUNCIONES UTILIZADAS

Funciones para la generación de números aleatorios

- Función `rbinom`: Se utiliza para generar muestras procedentes de una distribución binomial.

Argumentos
<code>n</code> : cantidad de números aleatorios a generar
<code>size</code> : número de ensayos (parámetro n de la distribución binomial)
<code>prob</code> : probabilidad de éxito en el experimento (parámetro p de la distribución binomial)

Diapositivas para las clases prácticas

Para cada una de las unidades prácticas se han elaborado diapositivas que permiten a los estudiantes seguir con mayor facilidad las clases en el aula de informática.

Ej. 3. Resolución

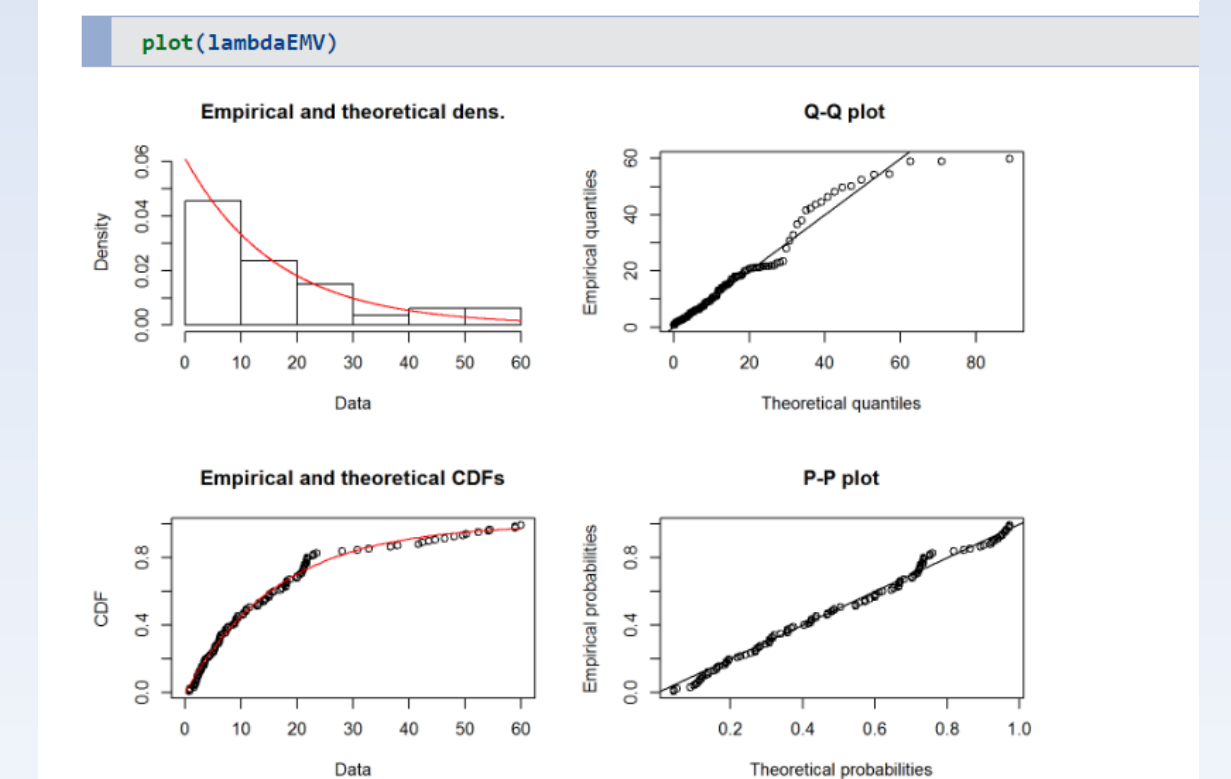
Al 99% de confianza, el intervalo se calcula:

```
t.test(x=coste.novedoso,y=coste.convencional,var.equal=TRUE,conf.level=0.99)
```

```
## Two Sample t-test
## data: coste.novedoso and coste.convencional
## t = 2.5426, df = 68, p-value = 0.0136
## alternative hypothesis: true difference in means is not equal to 0
## 99 percent confidence interval:
## -0.02694339 1.19122447
## sample estimates:
## mean of x mean of y
## 7.661600 7.079459
```

Interpretación: Con un 99% de confianza, la misma variación media creemos que está entre -0,03 y 1,19 euros.

Aptdo. h. Resolución.



Conclusiones: Se han combinado las prestaciones principales de R y su entorno para ofrecer una material interactivo y dinámico que facilite el aprendizaje autónomo del alumno. La experiencia ha sido bastante favorable, aun teniendo en cuenta el coste de entrada del uso del software y el perfil de los estudiantes de titulaciones de ciencias sociales.