



UNIVERSIDAD DE VALLADOLID

Dpto. Estadística e I.O.

La enseñanza de la estadística con herramientas didácticas como “R”

ANEXO 2:A121 2º ESO

**Trabajo Final del Máster Universitario de Profesor en Educación
Secundaria Obligatoria y Bachillerato. Especialidad de Matemáticas.**

Alumno: Julián Rodríguez Vaca.

Tutor: Dr. David Conde del Río.

Valladolid, Junio 2018.

Índice general.

Índice general.....	3
Capítulo 1. Introducción.....	5
Capítulo 2. Contenidos y estándares oficiales.....	7
Capítulo 3. Introducción a “R” Statistics.....	9
Capítulo 4. Contenidos y estándares oficiales con “R” Statistics.	13
Capítulo 5. Estándares de aprendizaje evaluables.....	37
Capítulo 6. Bibliografía.	39

Capítulo 1. Introducción.

El objetivo del presente trabajo es plantear y ofrecer una propuesta para la mejora de la didáctica de la estadística mediante el empleo de un potente software, destinado hasta hoy a estudios superiores.

Por un motivo ético, se ha elegido Software Libre, con el que se pretende fomentar este tipo de herramientas en el aula.

Este software, además de ahorrar una gran cantidad de tiempo, permitirá hacer más dinámica esta parte, gracias al manejo de grandes volúmenes de información, la realización de gráficos estadísticos de manera automática, y permitiendo el análisis de los datos de una forma más adecuada.

Con esto el alumno comenzará a tomar contacto con un software de programación y un lenguaje de alto nivel, lo que le mostrará puertas por abrir, y le aportará una buena ventaja sobre todo si se plantea estudios superiores.

El trabajo se presenta en forma de memoria, donde se recopila cada punto del temario en su versión más extensa, y donde aparecen más de 50 ejemplos de cómo resolver los ejercicios de forma tradicional, y con “R”. Incorpora nueve anexos con el temario preparado para cada uno de los cursos, que el profesor puede proporcionar a sus alumnos. Tanto en la memoria como en los anexos, aparece todo el código utilizado en la elaboración del trabajo. Las versiones de introducción o de repaso de cada punto del temario se han dejado en cada uno esos anexos, para evitar la duplicación de los contenidos en la memoria.

Puesto que el bloque de estadística se presenta en todos los cursos en el último bloque de la asignatura de matemáticas, sufre los retrasos de todos los bloques precedentes, dejando en la mayoría de las ocasiones un tiempo muy reducido para el desarrollo del mismo. Con el uso de este método, no se trata de evitar que el alumno trabaje el tan necesario cálculo mental y manual. Sin embargo, si el grupo llega hasta este punto con retraso, uno de los motivos puede

ser precisamente el llevar trabajando cerca de ocho meses en esta línea. Por ello, se trata de optimizar el poco tiempo del que disponga el profesor, evitando pérdidas en la representación de gráficos a mano, nubes de puntos, o tablas de contingencia.

El BOCYL establece, en sus Ordenes EDU 362 y 363 del 4 de mayo de 2015, que el quinto bloque, «Estadística y probabilidad», es de suma importancia.

Esto no sólo es cierto, sino que además, en la era de la información y de la competitividad, el futuro de las empresas y de los países no dependerá tanto del volumen de información de que dispongan, sino de la mejor explotación que hagan de la misma.

Independientemente de su elección tras acabar la ESO o el Bachillerato, el alumno adquirirá los conceptos y el vocabulario necesarios para poder aplicarlos de manera prácticamente autónoma en su futura profesión.

Así, al finalizar sus estudios será capaz de realizar análisis críticos de una mayor cantidad de información mediante tablas y gráficas, con la ayuda de “R”.

Será capaz de recopilar datos por sí mismo, organizarlos, resumirlos, estudiarlos y explotarlos, lo que le será de gran utilidad en su ámbito profesional.

El contenido del trabajo está adaptado a la comunidad de Castilla y León, según las órdenes EDU 362 y 363 de 2015 por las que se establecen los currículos y se regulan la implantación, evaluación y desarrollo de la educación secundaria obligatoria y del bachillerato en la Comunidad de Castilla y León:

ORDEN EDU/362/2015, de 4 de mayo, por la que se establece el currículo y se regula la implantación, evaluación y desarrollo de la educación secundaria obligatoria en la Comunidad de Castilla y León.

ORDEN EDU/363/2015, de 4 de mayo, por la que se establece el currículo y se regula la implantación, evaluación y desarrollo del bachillerato en la Comunidad de Castilla y León.

Así, establecen los temas para el bloque de estadística que veremos a continuación.

Capítulo 2. Contenidos y estándares oficiales.

2º ESO.

- 2E1.- Población e individuo. Muestra. Variables estadísticas.
- 2E2.- Variables cualitativas y cuantitativas discretas.
- 2E3.- Frecuencias absolutas y relativas.
- 2E4.- Organización en tablas de datos recogidos en una experiencia.
- 2E5.- Diagramas de barras, y de sectores. Polígonos de frecuencias.
- 2E6.- Medidas de tendencia central.
- 2E7.- Medidas de dispersión.

Capítulo 3. Introducción a “R” Statistics.

3.1 Sobre “R”:

R es un lenguaje y entorno para el procesamiento y representación de datos estadísticos. Es un proyecto de GNU similar al lenguaje y al entorno S, que fue desarrollado en Bell Laboratories, por John Chambers y su equipo. Hay algunas diferencias importantes, pero gran parte del código escrito para S corre inalterado bajo R.

R proporciona una amplia variedad de técnicas estadísticas y gráficos, y es altamente extensible mediante la creación de librerías por los usuarios, al ser una herramienta de código abierto.

Uno de los puntos fuertes de R es la facilidad con la que se pueden crear gráficos de calidad, incluyendo símbolos matemáticos y fórmulas si es necesario.

R está disponible como Software Libre bajo los términos de la Licencia Pública General GNU de la Free Software Foundation en forma de código fuente. Se compila y se ejecuta en una amplia variedad de plataformas UNIX y sistemas similares (incluidos FreeBSD y Linux), Windows y MacOS.

Fuente: <https://www.r-project.org/about.html>

3.2 Descarga e instalación de RStudio:

Rstudio es un software gratuito que podemos descargar de forma totalmente legal y sin coste ni publicidad, de la siguiente página:

<https://www.rstudio.com/products/rstudio/download/>

Pulsando en la tecla download, que aparece en la columna Free (gratis), nos dirige a la zona para elegir nuestro sistema operativo:

Installers	Size	Date	MD5
RStudio 1.1.453 - Windows Vista/7/8/10	85.8 MB	2018-05-16	bf287e385aef53829204023087e98735
RStudio 1.1.453 - Mac OS X 10.6+ (64-bit)	74.5 MB	2018-05-16	00a0088424ed06ac434f7a966f602b9c
RStudio 1.1.453 - Ubuntu 12.04-15.10/Debian 8 (32-bit)	89.3 MB	2018-05-16	6cfd86770c7b6dbc13e66f4f59c299ce
RStudio 1.1.453 - Ubuntu 12.04-15.10/Debian 8 (64-bit)	97.4 MB	2018-05-16	63e36e8138e369d19f9aaf4b0e995bbc
RStudio 1.1.453 - Ubuntu 16.04+/Debian 9+ (64-bit)	64.4 MB	2018-05-16	85b3e76c9fad4613bc9cf0de1f34b183
RStudio 1.1.453 - Fedora 19+/RedHat 7+/openSUSE 13.1+ (32-bit)	88.1 MB	2018-05-16	37cade7e162eab62483e6556e39dedee
RStudio 1.1.453 - Fedora 19+/RedHat 7+/openSUSE 13.1+ (64-bit)	90.6 MB	2018-05-16	44cddd285bc31c41e4eaeed174b8eebb

Dependiendo del sistema operativo de nuestro PC, descargamos el que corresponda.

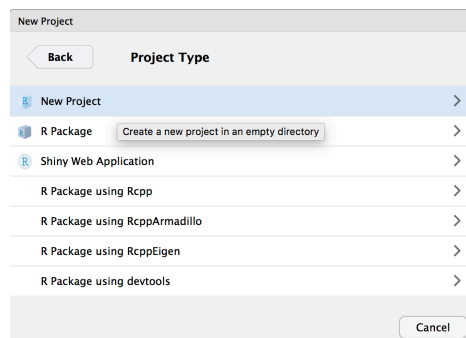
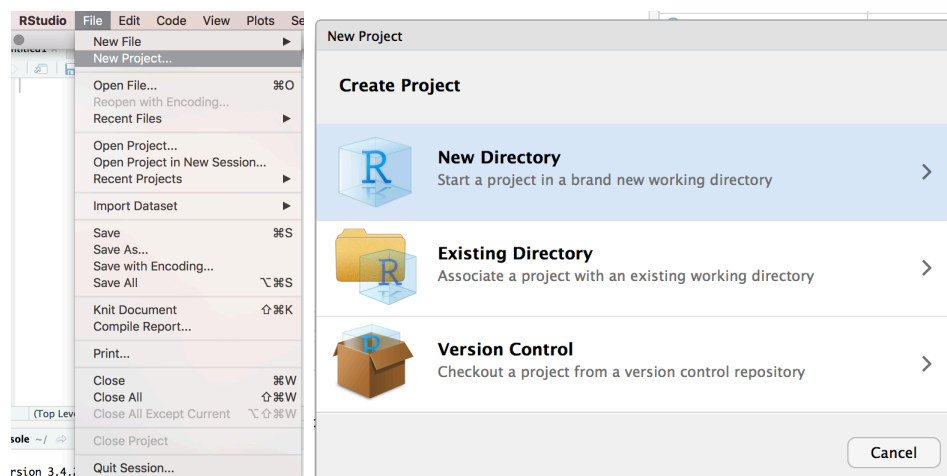
Una vez descargado, se ejecuta el programa instalador, y se van siguiendo los pasos del asistente de instalación, como en cualquier otro programa.

3.3.- Creación de un nuevo proyecto:

1.- Vamos a crear nuevo proyecto con el nombre población y muestra, ubicado en el escritorio del PC.

Para ello, abrir Rstudio y pulsar:

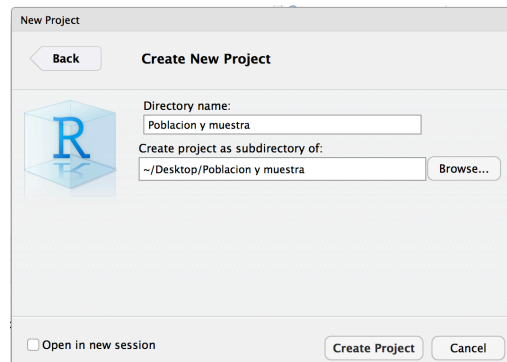
File/New project.../New directory/New Project



En la siguiente ventana, escribimos:

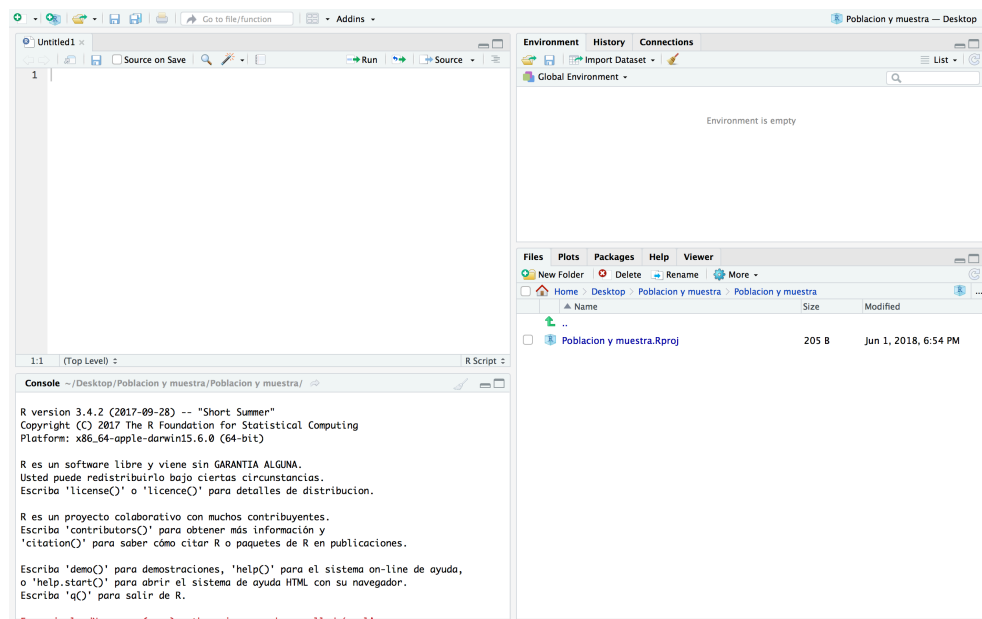
Directory name: Población y muestra.

Create project as subdirectory of: Click en browse... y creamos una carpeta en nuestro escritorio (desktop) que se llame Población y muestra.



Nota: las carpetas podemos crearlas tanto en el escritorio como en un USB o donde queramos, y luego localizarla usando la tecla browse.

Con esto, se nos abre el entrono de “R”, listo para empezar. Tendrá la siguiente pinta:



Los comandos se escriben en la zona inferior izquierda, y los gráficos se mostrarán en la ventana inferior derecha. Las ventanas superiores son para la selección y visualización de tablas y otras variables. Con esto el sistema está listo para comenzar a trabajar

Capítulo 4. Contenidos y estándares oficiales con “R” Statistics.

DeSeCo (2003) define competencia como «la capacidad de responder a demandas complejas y llevar a cabo tareas diversas de forma adecuada».

La competencia «supone una combinación de habilidades prácticas, conocimientos, motivación, valores éticos, actitudes, emociones, y otros componentes sociales y de comportamiento que se movilizan conjuntamente para lograr una acción eficaz».

Se contemplan, pues, como conocimiento en la práctica, es decir, un conocimiento adquirido a través de la participación activa en prácticas sociales y, como tales, se pueden desarrollar tanto en el contexto educativo formal, a través del currículo, como en los contextos educativos no formales e informales.

Fuente: <https://www.mecd.gob.es/>

En este trabajo se ha buscado contribuir a las competencias en:

Comunicación lingüística, mediante el fomento de un uso del vocabulario apropiado, de la lectura y sobre todo de la interpretación de los enunciados, que contribuyen finalmente a expresarse y comunicarse con propiedad.

Competencia matemática, mediante el análisis matemático del comportamiento de las variables de estudio de la población, extrayendo conclusiones en función de la regresión lineal y correlación de los datos de las variables, y bajo la interpretación conjunta de parámetros estadísticos.

Competencia digital, mediante el fomento de un uso ético, cívico y crítico de las nuevas tecnologías, y mediante el empleo de una herramienta Software de alto nivel.

Competencias sociales y cívicas, mediante el análisis de datos de nuestro entorno, como PIB, IPC, extrayendo conclusiones de posibles desigualdades salariales en poblaciones, o identificando las malas prácticas de las presentaciones de datos de forma interesada.

Competencia cultural, representando e interpretando la información con relación a ejemplos de plantas y otros datos del entorno.

La competencia aprender a aprender, mediante el ejemplo de la búsqueda de información para mejorar la utilización del software R de manera casi autodidacta.

Sentido de la iniciativa y espíritu emprendedor, mostrando al alumno el inicio de un camino, que con su propia iniciativa podrá recorrer hasta donde le lleve su curiosidad científica. Por su potencia y escasa inversión, el alumno será capaz de imaginar escenarios de emprendimiento, donde con un ordenador y este software como herramientas podrá realizar estudios de alto valor a nivel profesional.

2E1.- Población e individuo. Muestra. Variables estadísticas.

La Estadística es una ciencia que se ocupa del estudio de los métodos y procedimientos para recoger, clasificar, resumir y analizar datos observados sobre una población de individuos.

El objetivo de cualquier estudio estadístico es obtener información acerca de las características de los individuos de cierto colectivo, llamado población estadística.

Población: Es el conjunto total de individuos sobre los que se quieren estudiar unos datos determinados.

Individuo: Cada uno de los componentes de la población. Pueden ser personas, animales, plantas, u objetos.

Cuando la población o colectivo sea muy grande, se hará difícil el estudio de la misma. Estos inconvenientes pueden ser superados mediante la elección de muestras.

Muestra: Es una parte de la población representativa de la misma. Tiene por tanto características similares. Ha de elegirse al azar. Se utiliza cuando la población es muy grande, o difícil de estudiar.

Variable estadística: Es el dato o característica que se quiere estudiar. Por ejemplo: la estatura, la nota de Matemáticas, el sexo de una persona, o su peso.

Ejemplo. Población.

Vamos a descargar un archivo de población ya creado, con los resultados de la última evaluación internacional PISA, que a día de hoy, es la de 2015.

Lo descargamos de este enlace:

<https://www.mecd.gob.es/inee/evaluaciones-internacionales/pisa/pisa-2015.html>.

Descargamos el libro: Cap. 2. Los resultados de los alumnos en PISA: ciencias, lectura y matemáticas. Enlace externo, se abre en ventana nueva XLS.

Lo abrimos con Libreoffice Calc, y hacemos clic en la pestaña "Tabla 2.7", donde aparece la Tendencia de los resultados en ciencias por comunidades autónomas y media OCDE.

Copiamos la tabla, y la pegamos en una nueva hoja de cálculo.

Modificamos las cabeceras de la tabla, y las dejamos así:

CCAA	PM15	E15	PM12	E12	PM09	E09	PM06	E06
OCDE	493,20	0,429	501,20	0,492	500,84	0,494	500,00	0,531
España	492,79	2,066	496,45	1,829	488,25	2,053	488,42	2,565
Andalucía	473,05	4,148	486,03	4,260	469,22	5,344	473,77	3,968
Aragón	507,60	4,649	504,11	5,174	505,43	4,336	513,36	3,916
Asturias	501,26	3,878	516,86	4,686	501,63	4,901	508,49	4,944
Balears (Illes)	484,61	4,522	482,97	4,461	460,85	5,711		
Canarias	475,40	3,557			451,97	4,087		
Cantabria	495,62	5,603	500,51	3,672	500,19	4,660	509,41	3,562
Castilla y León	519,00	3,524	519,00	4,200	515,67	4,919	519,85	3,909
Cataluña	504,08	4,727	491,93	4,231	497,27	5,883	491,43	5,074
Extremadura	474,21	3,794	482,61	4,494				
Galicia	511,90	3,138	511,54	4,797	506,12	4,896	504,51	3,432
Madrid	515,77	3,454	517,48	3,997	507,60	4,239		
Murcia (R. de)	483,73	3,792	479,19	4,733	483,74	5,255		
Navarra	511,90	4,089	514,14	3,532	508,55	3,223	511,32	2,943
País Vasco	483,06	3,019	505,65	2,388	494,69	2,489	494,67	3,455
Rioja (La)	498,03	5,501	509,71	2,122	509,33	2,602	519,57	2,510

La guardamos en la carpeta del proyecto Población y muestra con el nombre Pisa2015.xlsx, y la importamos a R:

```
>Pisa2015 <- read_excel("~/Desktop/Población y muestra/Pisa2015.xlsx")
```

Ejemplo 2. Muestra:

Tomar una muestra representativa de la población que acabamos de cargar.

Vamos a tomar como muestra a 5 comunidades Autónomas al azar:

```
> MuestraPisa<-sample(Pisa2015$CCAA,size=5)
> MuestraPisa
[1] "Canarias"      "Aragón"        "Murcia (R. de)"
[4] "Balears (Illes)" "Extremadura"
```

El programa ha seleccionado aleatoriamente a Canarias, Aragón, Murcia, Islas Baleares y Extremadura.

2E2.- Variables cualitativas y cuantitativas.

Variable cualitativa: Describen cualidades que no puede expresarse por números. Por ejemplo, provincias españolas, colores favoritos, qué libro lectura prefieren los adolescentes, o el coche más vendido.

Variable cuantitativa: Las variables cuantitativas toman valores numéricos. Por ejemplo, la estatura, o la nota de una asignatura. Pueden ser:

Variable cuantitativa discreta: Los valores de la variable son números enteros 1, 2, 3, 4, 5. Por ejemplo el número de compras de un producto en un mes, no puede ser 4,8.

Variable cuantitativa continua: Pueden tomar todos los valores dentro de un intervalo. La temperatura, la humedad, o la estatura, son ejemplos de variables cuantitativas continuas. Los valores razonables de la variable temperatura en una persona pueden valer desde 34,5°C hasta 42°C, pudiendo tomar cualquier valor intermedio.

Ejemplo: Variables cuantitativas.

Vamos a trabajar con la columna PM15, que muestra los resultados de la prueba del año 2015.

Por ejemplo, vamos a buscar la CCAA con mayor puntuación. Para ello podemos utilizar la función `order`, que nos devuelve las posiciones en orden creciente de las variables de estudio:

```
> order(Pisa2015$PM15)
[1] 3 11 7 16 14 6 2 1 8 17 5 10 4 12 15 13 9
```

Esto significa que la CCAA con menor nota ha sido la 3, y la CCAA con mayor nota, ha sido la 9. Para saber cuál es cada una, podemos poner el nombre de la tabla, “Pisa2015” y nos la muestra con las posiciones en la primera columna. Vemos, por tanto, que la CCAA con menor nota ha sido la 3, Andalucía, y la CCAA con mayor nota, la 9, Castilla y León.

```
> Pisa2015
```

```
# A tibble: 17 x 9
```

	CCAA	PM15	E15	PM12	E12	PM09	E09	PM06	E06
	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	OCDE	493.	0.429	501.	0.492	501.	0.494	500.	0.531
2	España	493.	2.07	496.	1.83	488.	2.05	488.	2.57
3	Andalucía	473.	4.15	486.	4.26	469.	5.34	474.	3.97
4	Aragón	508.	4.65	504.	5.17	505.	4.34	513.	3.92
5	Asturias	501.	3.88	517.	4.69	502.	4.90	508.	4.94
6	Balears (Illes)	485.	4.52	483.	4.46	461.	5.71	NA	NA
7	Canarias	475.	3.56	NA	NA	452.	4.09	NA	NA
8	Cantabria	496.	5.60	501.	3.67	500.	4.66	509.	3.56
9	Castilla y León	519.	3.52	519.	4.20	516.	4.92	520.	3.91
10	Cataluña	504.	4.73	492.	4.23	497.	5.88	491.	5.07
11	Extremadura	474.	3.79	483.	4.49	NA	NA	NA	NA
12	Galicia	512.	3.14	512.	4.80	506.	4.90	505.	3.43
13	Madrid	516.	3.45	517.	4.00	508.	4.24	NA	NA
14	Murcia (R. de)	484.	3.79	479.	4.73	484.	5.26	NA	NA
15	Navarra	512.	4.09	514.	3.53	509.	3.22	511.	2.94
16	País Vasco	483.	3.02	506.	2.39	495.	2.49	495.	3.46
17	Rioja (La)	498.	5.50	510.	2.12	509.	2.60	520.	2.51

Para buscar el mayor valor de todas las notas, podemos escribir lo siguiente:

```
> max(Pisa2015$PM15)
[1] 518.9996
```

Esto nos devuelve la mayor puntuación, que ahora ya sabemos que es la obtenida por Castilla y león, con 518,9996 puntos.

2E3.- Frecuencias absolutas y relativas.

Frecuencia absoluta (n_i): Es el número de veces que aparece cada valor (x_i) de la variable.

La suma de las frecuencias absolutas es el número total de datos (N).

En "R", usaremos la función `table` para crear tablas de frecuencias absolutas:

```
> FA_Pelo<-table(Población$Pelo)
> FA_Pelo
```

castaño	moreno	pelirrojo	rubio
9	7	2	5

Frecuencia relativa (f_i): es el resultado de dividir la frecuencia absoluta entre el número total de datos (N):

$$f_i = \frac{n_i}{N}$$

En "R", usaremos la tabla de frecuencias absolutas, y la aplicaremos la función `prop.table`:

```
> prop.table(FA_Pelo)
```

castaño	moreno	pelirrojo	rubio
0.39130435	0.30434783	0.08695652	0.21739130

Estos ejercicios se realizan mucho más rápido utilizando el ordenador, como vamos a ver a continuación.

Frecuencias acumuladas

La frecuencia absoluta acumulada (N_i) de un valor X_i del conjunto (X_1, X_2, \dots, X_N) es la suma de las frecuencias absolutas de los valores menores o iguales a X_i , es decir: $N_i = n_1 + n_2 + \dots + n_i$.

En "R", usaremos la tabla de frecuencias absolutas, y la aplicaremos la función `cumsum`:

```
> cumsum(FA_pelo)
  castaño    moreno pelirrojo    rubio
      9      16      18      23
```

2E4.- Organización en tablas de datos recogidos en una experiencia.

Es muy común encontrarnos los datos colocados en tablas y gráficas para ayudar a su interpretación. Estos datos pueden haber sido obtenidos mediante encuestas, de muy diversos tipos.

Encuesta: Una encuesta es un procedimiento que permite obtener los datos para hacer un estudio de ellos. Puede ser oral o escrita. En función de las preguntas realizadas, se construye una tabla donde se representan los datos.

Ejemplo. Tablas en "R":

Los alumnos de 1º de la ESO, han hecho una encuesta, con la que han creado la tabla Población.xlsx. En ella, podríamos seguir insertando columnas, con las preguntas que se nos ocurriesen.

1.- Crea nuevo proyecto con el nombre población y muestra, ubicado en el escritorio del PC.

2.- Importa la tabla Población.xlsx. La llamaremos Población, escribiendo lo siguiente en la ventana inferior derecha:

```
>Población<-read_excel("~/Desktop/Población y muestra/Población.xlsx").
```

Vamos a estudiar las frecuencias relativas y absolutas de los alumnos rubios, por un lado, y las frecuencias de tener 2 hermanos por otro, sin la necesidad de hacer cuentas.

Nota: Los nombres de las Variables, escritos en cada cabecera, no deben contener espacios ni caracteres especiales. Para poner espacios, usaremos el guión bajo _.

Construimos una tabla con las frecuencias absolutas de la variable Pelo. Para ello usamos la función table, escribiendo lo siguiente:

```
> FA_Pelo<-table(Población$Pelo)
```

A la tabla de frecuencias absolutas la estaremos llamando FA_Pelo.

Mediante el símbolo <-, que representa una especie de flecha, estaremos

diciéndole al programa que escriba dentro de la variable lo que nosotros queramos. En este caso, como queremos escribir la tabla de frecuencias absolutas de la variable Pelo de nuestra población, hemos escrito a continuación de la flecha, la función que queremos utilizar, "table". A continuación y entre paréntesis, hemos escrito la tabla de donde tiene que buscar los datos, (Población\$Pelo). El símbolo dollar indica la columna a la que queremos referirnos.

Para visualizar lo que acabamos de escribir en la tabla FA_Pelo escribimos:

```
> FA_Pelo
```

El programa nos muestra lo siguiente:

castaño	moreno	pelirrojo	rubio
9	7	2	5

Por lo tanto, podemos ver que hay 9 alumnos castaños, 7 morenos, 2 pelirrojos, y 5 rubios.

La **moda**, en este caso, sería tener el pelo castaño, por ser el valor más repetido.

Las frecuencias absolutas del número de hermanos serían:

```
> FA_Hermanos<-table(Población$Hermanos)
> FA_Hermanos
0 1 2 3
4 9 8 2
```

La **moda** de nuestra clase sería tener un único hermano.

Para las frecuencias relativas, usaremos la tabla de frecuencias absolutas, y la aplicaremos la función prop.table:

```
> FR_Pelo<-prop.table(FA_Pelo)
> FR_Pelo
```

castaño	moreno	pelirrojo	rubio
0.39130435	0.30434783	0.08695652	0.21739130

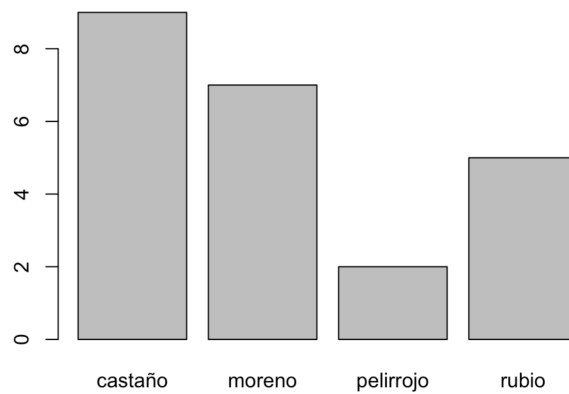
Con esta instrucción, el programa divide cada miembro de la tabla de frecuencias absolutas, por el número de miembros, y coloca el resultado en la nueva tabla.

Hallamos las frecuencias relativas de la variable Hermanos:

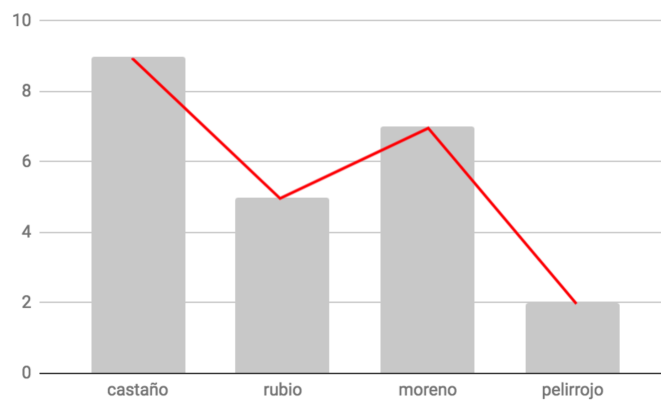
```
> FR_Hermanos<-(FA_Hermanos)/margin.table(FA_Hermanos)
> FR_Hermanos
      0      1      2      3
0.17391304 0.39130435 0.34782609 0.08695652
```

2E5.- Diagramas de barras, y de sectores. Polígonos de frecuencias.

El diagrama de barras muestra las frecuencias absolutas de los datos. Cuanto más alta es la barra más se da el valor al que corresponde.



Uniando el punto medio de la parte superior de cada barra del diagrama de barras se obtiene el polígono de frecuencias (en rojo). En el ejemplo de la tabla:



Ejercicio: Construye un diagrama de barras de las variables Pelo y Hermanos, e intérpretalas. Para ello, responde a las siguientes preguntas:

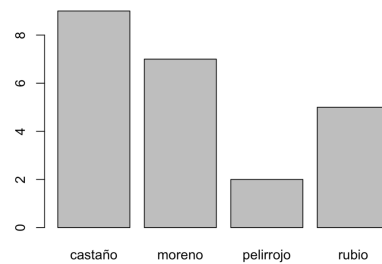
¿Quién predomina más en la clase, los rubios o los morenos?

¿Qué es más común, tener un hermano, o dos?

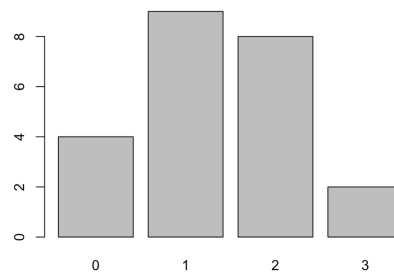
¿Qué color de pelo es el menos frecuente?

Para construir diagramas de barras, usaremos la función `barplot`.

```
> barplot(FA_Pelo)
```



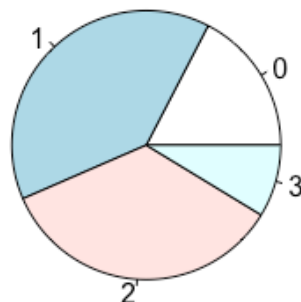
```
> barplot(FA_Hermanos)
```



En un diagrama de sectores, la amplitud de cada sector circular representa el valor de la variable:

En "R" se representa con la función `pie`

```
> pie(FA_Hermanos)
```



2E6.- Medidas de tendencia central.

En ocasiones interesa resumir la información de una muestra, en un solo valor, para hacernos una idea de cómo se comporta la variable y poder realizar comparaciones.

Las medidas de tendencia central más habituales, son la media, la mediana y la moda.

Media Aritmética: Es la suma de todos los datos dividida entre el número total de datos.

Es la medida de tendencia central más utilizada en Estadísticas y en todos los campos de la vida.

Tanto las empresas, como los países, como los medios de comunicación, constantemente hablan de medias de datos, como el gasto medio, el salario medio, la estatura media, o la media de la edad de los alumnos de primer curso de universidad.

Hay que tener cuidado, porque si no se usa debidamente, puede conducir a interpretaciones erróneas. Por ejemplo, si Pedro come dos porciones de pizza, y mi Juan ninguna, la media diría que han comido una cada uno. Y eso no es cierto.

Media Ponderada: La media ponderada (MP) de un conjunto de valores de una variable, consiste en dar a cada observación del conjunto de datos unos pesos, que indican la importancia que se quiere dar a cada uno de los valores que toma la variable.

Media Geométrica: La media geométrica de un conjunto de datos es el resultado de multiplicarlos entre si y aplicar la n -ésima raíz.

Si en la media aritmética sumábamos los valores para luego dividirlos, ahora debemos multiplicarlos para luego aplicar la n -ésima raíz pertinente.

La media geométrica necesita que no haya números negativos o que estos sean un número par, para evitar aplicar una raíz a un número negativo.

Mediana: La mediana divide en dos partes iguales a la distribución

estadística

Moda: Es el valor más repetido, o el valor de la variable estadística que tiene la frecuencia absoluta más alta. Pueden ser varias variables a la vez.

Ejercicios:

Media: La media la calcularemos con la función mean.

Mediana: La mediana la calcularemos con la función median.

Moda: Es el valor cuya frecuencia absoluta es más alta. Calculamos la tabla de frecuencias absolutas de las notas, y buscamos el mayor valor.

Ejercicio: Calcular la media, la mediana y la moda de las notas de la evaluación de PISA de 2015

Media:

```
> mean(Pisa2015$PM15)
[1] 495.5995
```

Mediana:

```
> median(Pisa2015$PM15)
[1] 495.6154
```

Moda:

Calculamos las frecuencias absolutas de las notas de la evaluación de PISA de 2015:

```
> table(Pisa2015$PM15)
```

```
473.05 474.2136 475.397 483.0577 483.727 484.6059 492.7861 493.2017
      1      1      1      1      1      1      1      1
495.6154 498.0305 501.2559 504.0753 507.6018 511.9015 511.9034 515.7698
      1      1      1      1      1      1      1      1
518.9996
```

En este caso, no aparecen valores repetidos, al ser cifras de cuatro decimales. Podemos redondear, para eliminar los decimales, y buscar de esta manera un valor que nos sirva como moda:

```
> table(round(Pisa2015$PM15))
```

```
473 474 475 483 484 485 493 496 498 501 504 508 512 516 519
  1   1   1   1   1   1   2   1   1   1   1   1   2   1   1
```

Al hacerlo así, aparecen dos modas: el 493, y el 512

2E7.- Medidas de dispersión.

Las medidas de dispersión son una serie de valores que nos informan cómo se encuentran los datos de agrupados o desagrupados.

Desviación media: mide la distancia media que hay entre todos los valores de la muestra y el valor medio.

Cuando mayor sea la desviación media quiere decir que mayor es la dispersión de la muestra, los datos están más separados, en cambio si el valor de la desviación es reducido significa que los valores de la media están muy concentrados.

Rango: mide la diferencia entre el valor mayor y el valor menor de la muestra.

Mientras mayor sea el rango más dispersos estén los valores.

Varianza: Al igual que la media, la varianza es un indicador que se utiliza para medirla dispersión (a mayor varianza, mayor dispersión), dando una información más precisa que la media.

Se calcula como suma de las diferencias al cuadrado de cada valor respecto a la media de la muestra. Esta suma se divide entre el número de datos. La varianza se suele representar con la letra V .

Desviación típica: La desviación típica es otra medida de dispersión y se calcula como raíz cuadrada de la varianza. Es la medida de dispersión que más se utiliza.

Ejercicios:**Desviación media:** Se calcula con la función `sd(x)`**Rango:** Se calcula con la función `range(x)`**Varianza:** Se calcula con la función `var(x)`**Desviación típica:** Se calcula con la función `sd(x)`

Ejercicio: Calcular la Desviación media, el rango, la varianza y la desviación típica de las notas de la evaluación de PISA de 2015

```
> sd(Pisa2015$PM15)
[1] 14.88819
> range(Pisa2015$PM15)
[1] 473.0500 518.9996
> var(Pisa2015$PM15)
[1] 221.6582
```

En este ejemplo los datos aparecen bastante centrados en la media. Por eso, la desviación típica es pequeña en comparación con la media.

El rango nos da los extremos de las notas mínima y máxima.

Ejercicio.

Calcular las medias de las longitudes de los pétalos de la tabla iris que contiene la muestra de 150 plantas de 3 especies diferentes:

La tabla iris es una tabla de 150 plantas. En esta tabla aparecen la longitud de los sépalos (`Sepal.Length`), la anchura de los sépalos (`Sepal.Width`), la longitud de los pétalos (`Petal.Length`), la anchura de los pétalos (`Petal.Width`), y la especie a la que pertenece la planta estudiada (`Species`). Tenemos tres tipos de especies, que pueden verse en las siguientes fotos:



Cargar la tabla de datos interna de “R” que se llama iris

```
> iris
```

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
1	5.1	3.5	1.4	0.2	setosa
2	4.9	3.0	1.4	0.2	setosa
3	4.7	3.2	1.3	0.2	setosa
4	4.6	3.1	1.5	0.2	setosa
5	5.0	3.6	1.4	0.2	setosa
6	5.4	3.9	1.7	0.4	setosa
7	4.6	3.4	1.4	0.3	setosa
8	5.0	3.4	1.5	0.2	setosa
9	4.4	2.9	1.4	0.2	setosa
10	4.9	3.1	1.5	0.1	setosa
11	5.4	3.7	1.5	0.2	setosa
12	4.8	3.4	1.6	0.2	setosa
13	4.8	3.0	1.4	0.1	setosa
14	4.3	3.0	1.1	0.1	setosa
15	5.8	4.0	1.2	0.2	setosa
16	5.7	4.4	1.5	0.4	setosa
17	5.4	3.9	1.3	0.4	setosa
18	5.1	3.5	1.4	0.3	setosa
19	5.7	3.8	1.7	0.3	setosa
20	5.1	3.8	1.5	0.3	setosa
21	5.4	3.4	1.7	0.2	setosa
22	5.1	3.7	1.5	0.4	setosa
23	4.6	3.6	1.0	0.2	setosa
24	5.1	3.3	1.7	0.5	setosa
25	4.8	3.4	1.9	0.2	setosa
26	5.0	3.0	1.6	0.2	setosa
27	5.0	3.4	1.6	0.4	setosa
28	5.2	3.5	1.5	0.2	setosa
29	5.2	3.4	1.4	0.2	setosa
30	4.7	3.2	1.6	0.2	setosa
31	4.8	3.1	1.6	0.2	setosa
32	5.4	3.4	1.5	0.4	setosa
33	5.2	4.1	1.5	0.1	setosa
34	5.5	4.2	1.4	0.2	setosa
35	4.9	3.1	1.5	0.2	setosa
36	5.0	3.2	1.2	0.2	setosa

37	5.5	3.5	1.3	0.2	setosa
38	4.9	3.6	1.4	0.1	setosa
39	4.4	3.0	1.3	0.2	setosa
40	5.1	3.4	1.5	0.2	setosa
41	5.0	3.5	1.3	0.3	setosa
42	4.5	2.3	1.3	0.3	setosa
43	4.4	3.2	1.3	0.2	setosa
44	5.0	3.5	1.6	0.6	setosa
45	5.1	3.8	1.9	0.4	setosa
46	4.8	3.0	1.4	0.3	setosa
47	5.1	3.8	1.6	0.2	setosa
48	4.6	3.2	1.4	0.2	setosa
49	5.3	3.7	1.5	0.2	setosa
50	5.0	3.3	1.4	0.2	setosa
51	7.0	3.2	4.7	1.4	versicolor
52	6.4	3.2	4.5	1.5	versicolor
53	6.9	3.1	4.9	1.5	versicolor
54	5.5	2.3	4.0	1.3	versicolor
55	6.5	2.8	4.6	1.5	versicolor
56	5.7	2.8	4.5	1.3	versicolor
57	6.3	3.3	4.7	1.6	versicolor
58	4.9	2.4	3.3	1.0	versicolor
59	6.6	2.9	4.6	1.3	versicolor
60	5.2	2.7	3.9	1.4	versicolor
61	5.0	2.0	3.5	1.0	versicolor
62	5.9	3.0	4.2	1.5	versicolor
63	6.0	2.2	4.0	1.0	versicolor
64	6.1	2.9	4.7	1.4	versicolor
65	5.6	2.9	3.6	1.3	versicolor
66	6.7	3.1	4.4	1.4	versicolor
67	5.6	3.0	4.5	1.5	versicolor
68	5.8	2.7	4.1	1.0	versicolor
69	6.2	2.2	4.5	1.5	versicolor
70	5.6	2.5	3.9	1.1	versicolor
71	5.9	3.2	4.8	1.8	versicolor
72	6.1	2.8	4.0	1.3	versicolor
73	6.3	2.5	4.9	1.5	versicolor
74	6.1	2.8	4.7	1.2	versicolor
75	6.4	2.9	4.3	1.3	versicolor
76	6.6	3.0	4.4	1.4	versicolor
77	6.8	2.8	4.8	1.4	versicolor

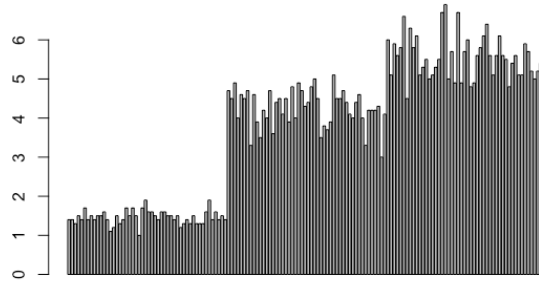
78	6.7	3.0	5.0	1.7	versicolor
79	6.0	2.9	4.5	1.5	versicolor
80	5.7	2.6	3.5	1.0	versicolor
81	5.5	2.4	3.8	1.1	versicolor
82	5.5	2.4	3.7	1.0	versicolor
83	5.8	2.7	3.9	1.2	versicolor
84	6.0	2.7	5.1	1.6	versicolor
85	5.4	3.0	4.5	1.5	versicolor
86	6.0	3.4	4.5	1.6	versicolor
87	6.7	3.1	4.7	1.5	versicolor
88	6.3	2.3	4.4	1.3	versicolor
89	5.6	3.0	4.1	1.3	versicolor
90	5.5	2.5	4.0	1.3	versicolor
91	5.5	2.6	4.4	1.2	versicolor
92	6.1	3.0	4.6	1.4	versicolor
93	5.8	2.6	4.0	1.2	versicolor
94	5.0	2.3	3.3	1.0	versicolor
95	5.6	2.7	4.2	1.3	versicolor
96	5.7	3.0	4.2	1.2	versicolor
97	5.7	2.9	4.2	1.3	versicolor
98	6.2	2.9	4.3	1.3	versicolor
99	5.1	2.5	3.0	1.1	versicolor
100	5.7	2.8	4.1	1.3	versicolor
101	6.3	3.3	6.0	2.5	virginica
102	5.8	2.7	5.1	1.9	virginica
103	7.1	3.0	5.9	2.1	virginica
104	6.3	2.9	5.6	1.8	virginica
105	6.5	3.0	5.8	2.2	virginica
106	7.6	3.0	6.6	2.1	virginica
107	4.9	2.5	4.5	1.7	virginica
108	7.3	2.9	6.3	1.8	virginica
109	6.7	2.5	5.8	1.8	virginica
110	7.2	3.6	6.1	2.5	virginica
111	6.5	3.2	5.1	2.0	virginica
112	6.4	2.7	5.3	1.9	virginica
113	6.8	3.0	5.5	2.1	virginica
114	5.7	2.5	5.0	2.0	virginica
115	5.8	2.8	5.1	2.4	virginica
116	6.4	3.2	5.3	2.3	virginica
117	6.5	3.0	5.5	1.8	virginica
118	7.7	3.8	6.7	2.2	virginica

119	7.7	2.6	6.9	2.3	virginica
120	6.0	2.2	5.0	1.5	virginica
121	6.9	3.2	5.7	2.3	virginica
122	5.6	2.8	4.9	2.0	virginica
123	7.7	2.8	6.7	2.0	virginica
124	6.3	2.7	4.9	1.8	virginica
125	6.7	3.3	5.7	2.1	virginica
126	7.2	3.2	6.0	1.8	virginica
127	6.2	2.8	4.8	1.8	virginica
128	6.1	3.0	4.9	1.8	virginica
129	6.4	2.8	5.6	2.1	virginica
130	7.2	3.0	5.8	1.6	virginica
131	7.4	2.8	6.1	1.9	virginica
132	7.9	3.8	6.4	2.0	virginica
133	6.4	2.8	5.6	2.2	virginica
134	6.3	2.8	5.1	1.5	virginica
135	6.1	2.6	5.6	1.4	virginica
136	7.7	3.0	6.1	2.3	virginica
137	6.3	3.4	5.6	2.4	virginica
138	6.4	3.1	5.5	1.8	virginica
139	6.0	3.0	4.8	1.8	virginica
140	6.9	3.1	5.4	2.1	virginica
141	6.7	3.1	5.6	2.4	virginica
142	6.9	3.1	5.1	2.3	virginica
143	5.8	2.7	5.1	1.9	virginica
144	6.8	3.2	5.9	2.3	virginica
145	6.7	3.3	5.7	2.5	virginica
146	6.7	3.0	5.2	2.3	virginica
147	6.3	2.5	5.0	1.9	virginica
148	6.5	3.0	5.2	2.0	virginica
149	6.2	3.4	5.4	2.3	virginica
150	5.9	3.0	5.1	1.8	virginica

La tabla iris es una tabla de 150 plantas, en las que aparecen la longitud de sus sépalos (Sepal.Length), la anchura de sus sépalos (Sepal.Width), la longitud de los pétalos (Petal.Length), la anchura de los pétalos (Petal.Width), y la especie a la que pertenece la planta estudiada (Species)

Dibujamos el diagrama de barras de la longitud de los pétalos:

```
> barplot (iris$Petal.Length)
```



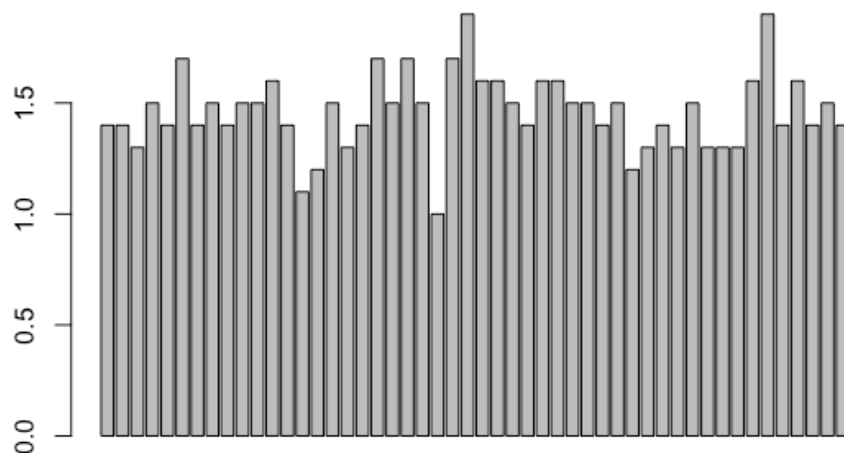
En el gráfico se aprecian tres escalones. Si nos fijamos en la tabla, las 50 primeras muestras, son plantas de la especie Setosa, las 50 siguientes, de la especie versicolor y las 50 últimas, de la especie virgínica.

Podemos decir, por tanto, que la primera especie tiene de media los pétalos más cortos que las otras dos especies.

Para hacer un estudio más riguroso de estas plantas, podría interesar estudiarlas por separado.

Podemos analizar la longitud de las plantas por separado, limitando la tabla a la especie setosa, con la instrucción: `iris[iris$Species == "setosa",]`.

```
> barplot (iris[iris$Species == "setosa",]$Petal.Length)
```



Igualmente, podemos extraer la media de la longitud de los pétalos de todas estas plantas:

```
> mean (iris[iris$Species == "setosa",]$Petal.Length)
[1] 1.462
```

Podemos hacer lo mismo para las otras especies:

```
> mean (iris[iris$Species == "versicolor",]$Petal.Length)
[1] 4.26
```

```
> mean (iris[iris$Species == "virginica",]$Petal.Length)
[1] 5.552
```

O para la anchura de los pétalos:

```
> mean (iris[iris$Species == "setosa",]$Petal.Width)
[1] 0.246
```

```
> mean (iris[iris$Species == "versicolor",]$Petal.Width)
[1] 1.326
```

```
> mean (iris[iris$Species == "virginica",]$Petal.Width)
[1] 2.026
```

En este caso, la anchura de los pétalos se comporta de manera similar a las longitudes.

Si calculamos las medianas:

```
> median (iris[iris$Species == "setosa",]$Petal.Width)
[1] 0.2
```

```
> median (iris[iris$Species == "versicolor",]$Petal.Width)
[1] 1.3
```

```
> median (iris[iris$Species == "virginica",]$Petal.Width)
[1] 2
```

Vemos que hay algo de diferencia con respecto a las medias, pero no mucha.

Podemos obtener más información de las variables con la función `summary`:

```
> summary(iris[iris$Species == "setosa",]$Petal.Width)
  Min. 1st Qu. Median   Mean 3rd Qu.   Max.
 0.100  0.200  0.200   0.246  0.300   0.600
```

Esto nos devolvería los valores mínimo, máximo, la media, la mediana y los cuartiles.

Capítulo 5. Estándares de aprendizaje evaluables.

Estándares 2º ESO. Aplicadas

1.1. Define población, muestra e individuo desde el punto de vista de la estadística, y los aplica a casos concretos.

1.2. Reconoce y propone ejemplos de distintos tipos de variables estadísticas, tanto cualitativas como cuantitativas.

1.3. Organiza datos, obtenidos de una población, de variables cualitativas o cuantitativas en tablas, calcula sus frecuencias absolutas y relativas, y los representa gráficamente.

1.4. Calcula la media aritmética, la mediana y la moda y los emplea para resolver problemas.

2.1. Emplea la calculadora y herramientas tecnológicas para organizar datos, y calcular las medidas de tendencia central.

3.1. Identifica los experimentos aleatorios y los distingue de los deterministas.

3.2. Calcula la frecuencia relativa de un suceso mediante la experimentación.

3.3. Realiza predicciones sobre un fenómeno aleatorio a partir del cálculo exacto de su probabilidad o la aproximación de la misma mediante la experimentación.

4.1. Describe experimentos aleatorios sencillos y enumera todos los resultados posibles, apoyándose en tablas, recuentos o diagramas en árbol sencillos.

Capítulo 6. Bibliografía.

Software Rstudio:

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

Manual de R:

Título: R para profesionales de los datos: una introducción

Autor: Carlos J. Gil Bellosta

Fecha: 2018-04-22

https://www.datanalytics.com/libro_r/index.html

Histogramas en R:

<https://www.cs.waikato.ac.nz/~fbravoma/teaching/explora.pdf>

Librerías en R:

<http://ggplot2.tidyverse.org/>

<http://rstudio-pubs->

static.s3.amazonaws.com/324830_8985f6dac8d34633b6cf23a92ff3e64c.html#sepal.length
[gth](#)

Estructuras de datos en R:

http://www.dm.uba.ar/materias/analisis_de_datos/2009/2/practicas/TP2-2009.pdf

Creación de data.frames en R:

<http://r-econ.blogspot.com/2012/07/unir-varios-dataframes-en-un-solo-paso.html>

Curso de introducción a la Estadística:

<http://www.postdata-statistics.com/IntroEstadistica/Tutoriales/Tutorial-00.pdf>

<http://www.postdata-statistics.com/IntroEstadistica/Tutoriales/Tutorial-02.pdf>

<http://www.postdata-statistics.com/IntroEstadistica/Tutoriales/Tutorial-04.pdf>

<http://www.postdata-statistics.com/IntroEstadistica/Tutoriales/Tutorial-05.pdf>

<http://www.postdata-statistics.com/IntroEstadistica/Tutoriales/Tutorial-06.pdf>

<http://www.postdata-statistics.com/IntroEstadistica/Tutoriales/Tutorial-07.pdf>

<http://www.postdata-statistics.com/IntroEstadistica/Tutoriales/Tutorial-08.pdf>

<http://www.postdata-statistics.com/IntroEstadistica/Tutoriales/Tutorial-09.pdf>

<http://www.postdata-statistics.com/IntroEstadistica/Tutoriales/Tutorial-10.pdf>
<http://www.postdata-statistics.com/IntroEstadistica/Tutoriales/Tutorial-11.pdf>

Contenidos y estándares oficiales:

Real Decreto 1105/2014, de 26 de diciembre, por el que se establece el currículo básico de la Educación Secundaria Obligatoria y del Bachillerato.

<https://www.boe.es/boe/dias/2015/01/03/pdfs/BOE-A-2015-37.pdf>

ORDEN EDU/362/2015, de 4 de mayo, por la que se establece el currículo y se regula la implantación, evaluación y desarrollo de la educación secundaria obligatoria en la Comunidad de Castilla y León.

<http://bocyl.jcyl.es/boletines/2015/05/08/pdf/BOCYL-D-08052015-4.pdf>

Libros de texto:

Fundamentos y métodos de Estadística, 3ª Edición,:

Autores: M. López Cachero

Editorial: Ed Piramide

ISBN 84-368-0171-7

Estadística aplicada a las ciencias de la educación:

Autores: Joan Welkowitz, Robert B. Ewen, Jacob Cohen

Editorial: Ed Santillana

ISBN: 84-294-1903-9

Libros de texto IES Cristo Rey, curso 2017-12018

Matemáticas 1º ESO:

Autores: Francisco Javier García Crespo; Ruth Martín Escanilla

Editorial: EDITEX S.A

1ª ed. (2015)

ISBN: 8490784949 ISBN-13: 9788490784945

Matemáticas 2º ESO

Autores: Fernando ... [et al.] Alcaide Guindo

Pelorroto; Juan Antonio Rocafort (il.)

Editorial: EDICIONES SM

1ª ed. (01/05/2016)

ISBN: 8467586885 ISBN-13: 9788467586886

Apuntes de Complementos de Matemáticas del Máster en profesor de educación secundaria obligatoria y bachillerato, formación profesional y enseñanzas de idiomas

http://campusvirtual2017.uva.es/pluginfile.php/415026/mod_resource/content/1/CM2017-18-Material%20Estad%C3%ADstica-2.pdf

Apuntes de estadística para Ingenieros Técnicos Industriales. Curso 2005, Escuela Universitaria Politécnica de Valladolid.

Otros:

Definiciones de Medidas de centralización y de dispersión:
Wikipedia

Referencia censos bíblicos:
<https://www.bible.com/es/bible/149/NUM.1.RVR1960?parallel=149>

Definiciones. Encuesta, censos:
http://www.ine.es/explica/explica_pasos_primera_encuesta.htm

Imágenes flores iris:
<http://www.lac.inpe.br/~rafael.santos/Docs/R/CAP394/WholeStory-Iris.html>