



UNIVERSIDAD  
DE GRANADA



**MÁSTER UNIVERSITARIO EN PROFESORADO  
DE EDUCACIÓN SECUNDARIA Y BACHILLERATO**

# **Propuesta de enseñanza para la estadística de 1º de Bachillerato de Ciencias Sociales a través del lenguaje de programación R**

Luis Felipe Del Río López

Especialidad: Matemáticas  
Tutor: Rafael Ramírez Uclés  
Curso académico 2020-2021

Luis Felipe Del Río López

<b>1. Justificación y marco teórico</b>	<b>3</b>
1.1. Introducción . . . . .	3
1.2. Descripción y clasificación del material . . . . .	5
1.3. Fundamentación . . . . .	7
<b>2. Análisis de Contenido</b>	<b>10</b>
2.1. Organización del contenido . . . . .	10
2.1.1. Ámbito Conceptual . . . . .	11
2.1.2. Ámbito Procedimental . . . . .	13
2.1.3. Ámbito Actitudinal . . . . .	13
2.2. Mapa Conceptual . . . . .	14
2.3. Sistemas de representación . . . . .	15
2.4. Situaciones . . . . .	17
<b>3. Análisis Cognitivo</b>	<b>18</b>
3.1. Expectativas de aprendizaje . . . . .	18
3.2. Limitaciones del aprendizaje . . . . .	22
3.3. Oportunidades de aprendizaje . . . . .	23
<b>4. Análisis de instrucción</b>	<b>24</b>
4.1. Atención a la diversidad . . . . .	38
<b>5. Análisis evaluativo</b>	<b>40</b>
<b>6. Conclusiones</b>	<b>43</b>
<b>Bibliografía</b>	<b>45</b>
<b>Anexos</b>	<b>47</b>

# 1 JUSTIFICACIÓN Y MARCO TEÓRICO

## 1.1. Introducción

El presente Trabajo de Fin de Máster, perteneciente a la modalidad de validación de materiales didácticos, tiene como objetivo introducir y promover el uso de las nuevas tecnologías en el aula, en particular el programa RStudio<sup>1</sup>. Para contextualizar este trabajo realizaremos el análisis de contenido y cognitivo de la Unidad Didáctica de Estadística de 1º de Bachillerato de Ciencias Sociales (incluyendo un repaso de la estadística unidimensional vista en la ESO) y, finalmente, propondremos una serie de tareas en las que pondremos de manifiesto la utilidad del uso de dicho software.

En [Gea et al. \(2015a\)](#) se afirma que la presencia de recursos tecnológicos en los libros de texto para el tema de estadística bidimensional es escasa. En consecuencia, esto implica que los conjuntos de datos utilizados en los ejercicios sean pequeños y, por consiguiente, que las estimaciones de la correlación y regresión sean poco fiables. Además de ello, los alumnos dedican mucho tiempo a realizar cálculos en vez de dedicarlo a interpretar y a reflexionar sobre los datos tratados ([Galindo, 2017](#)). Es por ello que utilizaremos el lenguaje de programación R ya que, gracias a él, podremos realizar dichos cálculos con grandes conjuntos de datos y, además, utilizar el tiempo que ahorraremos en hacer los cálculos para aumentar la comprensión, interpretación y el espíritu crítico del alumnado frente a los datos tratados.

Todo lo dicho anteriormente es válido para gran cantidad de programas, es por ello que una de las preguntas que el lector podría hacerse es: ¿por qué usar RStudio cuando, por ejemplo, Excel es más conocido y su uso es más sencillo? Ciertamente es que muchas de las funciones que usaremos en RStudio están presentes en Excel y, además, de forma más intuitiva. Sin embargo, enseñar R en el aula tiene, en mi opinión, cuatro

---

<sup>1</sup>Resaltar que usaremos RStudio y no R. La diferencia entre ambos, a efectos prácticos del alumnado, es la sencillez. RStudio posee editor y gran cantidad de menús desplegados, mientras que R únicamente posee ventana de comandos. RStudio utiliza el lenguaje de programación R.

principales beneficios:

1. **R es poco conocido fuera de la universidad.** R es, junto con SPSS, uno de los software más importantes en el ámbito de la estadística. Sin embargo, no es generalmente hasta la universidad cuando el alumno conoce de su existencia. Un primer contacto antes de ella mejorará el rendimiento de los alumnos en programación a la vez que trabajan la competencia digital.
2. **Es una apuesta de futuro.** Para problemas con muestras de pequeño tamaño, o que requieran poca potencia de cálculo, Excel es, a priori, mucho más cómodo. Sin embargo, en los problemas estadísticos reales se usan enormes cantidades de datos que requieren gran potencia de cálculo. Si el alumno aprende a usar R desde un primer momento podrá, más adelante, enfrentarse a problemas que no podría tratar de forma óptima con Excel. Además, R tiene muchas más aplicaciones, también fuera del ámbito de la estadística, lo cual lo hace un lenguaje de programación muy versátil.
3. **Presentación de los resultados.** Gracias al paquete RMarkdown todos los gráficos, cálculos y comentarios que realice el alumno en RStudio podrá ser exportados en formato pdf, pudiendo crear así informes muy fácilmente<sup>2</sup>. De esta forma, los alumnos podrán resolver los ejercicios en el propio programa y, sólo con darle a un botón, exportarlos a pdf para, por ejemplo, enviárselos a su profesor.
4. **Es Software Libre.** RStudio es Software Libre y, en consecuencia, es accesible a todo el alumnado de forma gratuita. Además, este software no necesita grandes requerimientos computacionales, por lo que podrá ejecutarse en prácticamente cualquier ordenador, eliminando una posible brecha económica-social.

---

<sup>2</sup>Todos los documentos incluidos en el Anexo están generados con RMarkdown.

## 1.2. Descripción y clasificación del material

El material en el cual centramos nuestro trabajo es el lenguaje de programación R, en particular el entorno de desarrollo integrado RStudio. R es uno de los software de referencia en el mundo de la estadística, de hecho es el séptimo lenguaje de programación más popular según el índice PPYL en Abril de 2021<sup>3</sup>. El lenguaje de programación R fue desarrollado inicialmente por los profesores Robert Gentleman y Ross Ihaka del Departamento de Estadística de la Universidad de Auckland en 1993, los cuales combinaron los puntos fuertes de dos lenguajes de programación ya existentes en la época, S y Scheme. En la actualidad su desarrollo está a cargo del *R Development Core Team*.

El programa que utilizaremos será RStudio, el cual utiliza R como lenguaje de programación. Su principal ventaja con respecto a R es que incluye un editor de sintaxis, múltiples opciones de trazado y depuración y una interfaz que facilita la gestión del espacio de trabajo. Está disponible para los principales sistemas operativos (Windows, Mac y Linux) e, incluso, está disponible para navegadores conectados a los servidores de RStudio.

Las principales características del programa, las cuales se pueden consultar en <https://www.rstudio.com/products/rstudio/>, son las siguientes:

- El resaltado de sintaxis, auto completado de código y sangría inteligente.
- Ejecutar código R directamente desde el editor de código fuente.
- Salto rápido a las funciones definidas.
- Administración sencilla de múltiples directorios de trabajo mediante proyectos.
- Navegación en espacios de trabajo y visor de datos.
- Depurador interactivo para diagnosticar y corregir los errores rápidamente.

---

<sup>3</sup><https://pyp1.github.io/PYPL.html>

A continuación mostramos una imagen del programa, en la cual podemos distinguir claramente las cuatro partes de su interfaz:

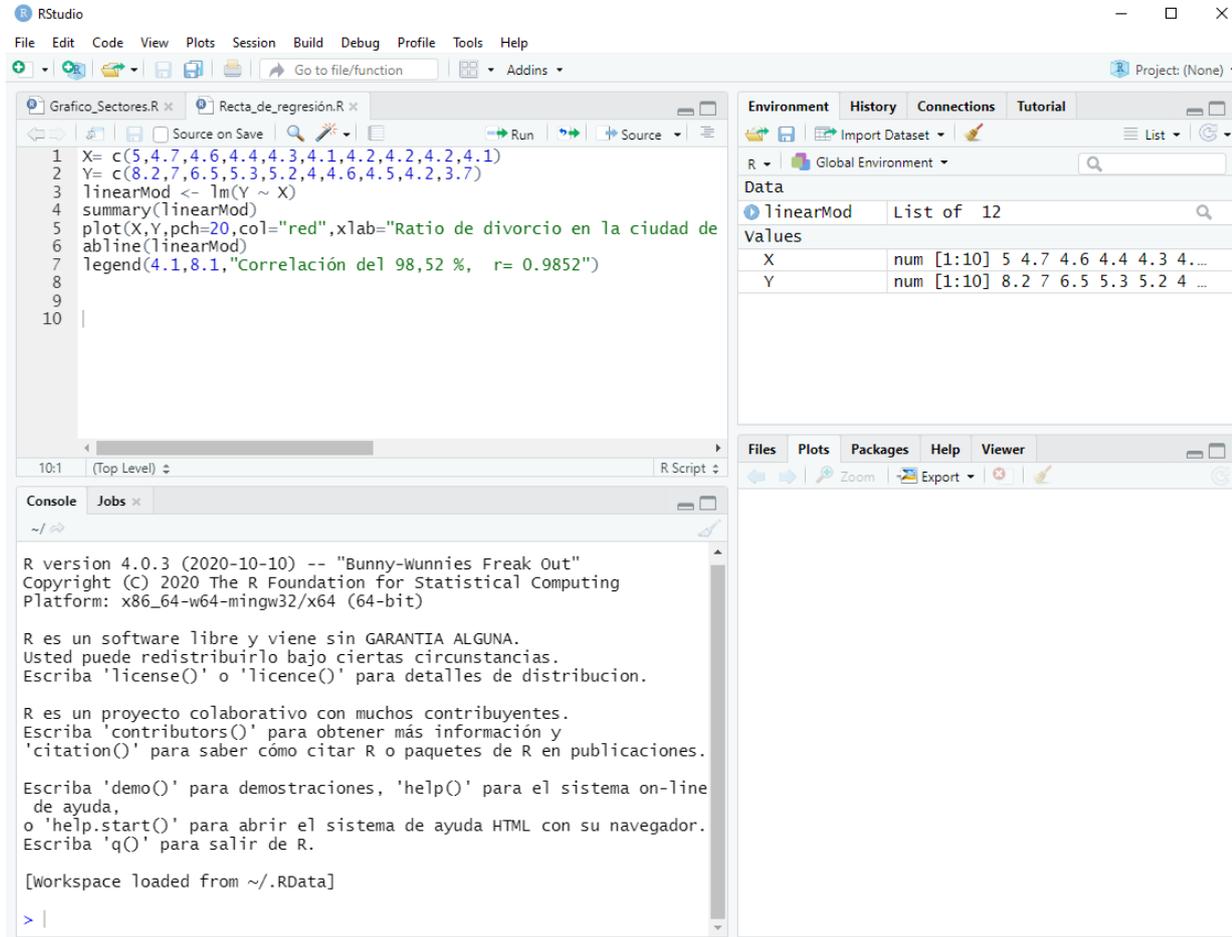


Figura 1.1: Interfaz de RStudio

La información referente a las funciones y paquetes usados en el programa puede consultarse directamente desde la pestaña *Help* en la parte inferior derecha, lo cual facilita su correcto uso.

Para la clasificación de nuestro material (el software R) tomaremos como referencia los criterios descritos en Rico y Moreno (2016) y recogeremos dicha información en las siguientes tablas:

Utilidad	
Amplitud de su utilidad	Muy versátil
Contenido	Álgebra, funciones, probabilidad, estadística.
Nivel educativo	ESO segundo ciclo, Bachillerato, Universidad.
Momento de utilización	Pre-instruccional, co-instruccional, post-instruccional.
Tipo de tarea y actividad que facilita	Mostrar-observar, proponer-manipular, resolver problemas.

Formato	
Soporte	Informático
Accesibilidad	Fácil de encontrar en el mercado, software libre.
Grado de difusión	Muy específico

### 1.3. Fundamentación

Los contenidos de Estadística de la asignatura Matemáticas Aplicadas a las Ciencias Sociales de 1º Bachillerato, descritos en la Orden 15 de Enero de 2021, podemos agruparlos de la siguiente manera:

#### 1. Estadística descriptiva bidimensional:

- Tablas de contingencia
- Distribución conjunta y distribuciones marginales.
- Distribuciones condicionadas.
- Medias y desviaciones típicas marginales y condicionadas.
- Independencia de variables estadísticas.
- Estudio de dependencia de dos variables estadísticas.

#### 2. Representación gráfica:

- Nube de puntos.
- Dependencia lineal de dos variables estadísticas.

### 3. Covarianza y correlación:

- Cálculo e interpretación del coeficiente de correlación lineal.
- Regresión lineal.
- Estimación de variables.
- Predicciones estadísticas y fiabilidad de las mismas.
- Coeficiente de determinación.

Las competencias clave que se trabajan en esta unidad, de las recogidas en la Orden 15 de Enero de 2021, son las siguientes:

- Competencia matemática y competencias básicas en ciencia y tecnología (CMCT).
- Competencia en comunicación lingüística (CCL).
- Competencia digital (CD).
- Competencia de aprender a aprender (CAA).
- Competencias sociales y cívicas (CSC).

No trabajan, sin embargo, las siguiente competencias:

- Sentido de iniciativa y espíritu emprendedor (SEIP).
- Conciencia y expresiones culturales (CEC).

Adicionalmente, los criterios de evaluación que la Orden 15 de Enero de 2021 describe son los siguientes:

1. *Describir y comparar conjuntos de datos de distribuciones bidimensionales, con variables discretas o continuas, procedentes de contextos relacionados con la economía y otros fenómenos sociales y obtener los parámetros estadísticos más usuales, mediante los medios más adecuados (lápiz y papel, calculadora, hoja de cálculo) , valorando la dependencia entre las variables.*

2. *Interpretar la posible relación entre dos variables y cuantificar la relación lineal entre ellas mediante el coeficiente de correlación, valorando la pertinencia de ajustar una recta de regresión y, en su caso, la conveniencia de realizar predicciones, evaluando la fiabilidad de las mismas en un contexto de resolución de problemas relacionados con fenómenos económicos y sociales.*

Criterio que sí aparece en el Bachillerato de Ciencias y que incluiremos a pesar de no estar incluido en Ciencias Sociales:

- *Utilizar el vocabulario adecuado para la descripción de situaciones relacionadas con la estadística, analizando un conjunto de datos o interpretando de forma crítica informaciones estadísticas presentes en los medios de comunicación, la publicidad y otros ámbitos, detectando posibles errores y manipulaciones tanto en la presentación de los datos como de las conclusiones.*

### **ESTADÍSTICA UNIDIMENSIONAL DE REPASO:**

Además del propio tema de estadística bidimensional, repasaremos los conceptos vistos a lo largo de la ESO referentes a estadística unidimensional. Para el repaso de dicho tema trabajaremos el siguiente bloque de contenidos:

- Terminología y conceptos básicos de la Estadística Unidimensional:
  - Individuo, población, muestra y variable estadística.
  - Organización de los datos: gráficos y tablas de frecuencia.
  - Distribución de frecuencias.
  - Parámetros estadísticos de localización, dispersión y posición. Significado y cálculo.

## 2 ANÁLISIS DE CONTENIDO

### 2.1. Organización del contenido

Consideremos el sistema de categorías para el análisis de contenido matemático escolar descrito en Rico y Moreno (2016):

Organizadores para el análisis de contenido		
Ámbitos		
Conceptual	Procedimental	Actitudinal
Primer nivel: unidades de información		
<i>Hechos</i>	<i>Destrezas</i>	<i>Emociones</i>
Términos/notaciones	Operaciones	Seguridad
Convenios	Reglas	Disciplina
Resultados	Algoritmos	Dominio y autoestima
Segundo nivel: abstracción y generalización		
<i>Conceptos</i>	<i>Razonamientos</i>	<i>Moralidad y normas</i>
Extensión	Inductivo	Respeto y aplicación de reglas
Comprensión	Deductivo	Corrección de procedimientos
Analogía	Relacional	Coherencia
Tercer nivel: estructuración		
<i>Estructuras conceptuales</i>	<i>Estrategias</i>	<i>Valores éticos</i>

Para el estudio del tema y su posterior clasificación se han consultado los libros de texto de las editoriales SM (Alcaide et al., 2015) y ANAYA (Jiménez et al., 2020) y diversos estudios realizados sobre el tema en la UGR, entre los que se encuentran Batanero et al. (2017) y Gea et al. (2014).

### 2.1.1. **Ámbito Conceptual**

**Términos:** centro de gravedad, coeficiente de correlación, coeficiente de determinación, coeficiente de regresión, correlación, dependencia, sentido, curvilínea, funcional, lineal, covarianza, desviación típica, diagrama de barra, diagrama de sectores, diagrama de dispersión, distribución conjunta marginal, distribución conjunta condicional, error de estimación, fiabilidad, frecuencia conjunta absoluta, frecuencia relativa, frecuencia condicionada, frecuencia marginal, histograma, incorrelada, independencia, media marginal, método de mínimos cuadrados, nube de puntos, recta de regresión, recta de mínimos cuadrados, tabla de doble entrada, predicción, valor observado, variable dependiente/independiente, variable estadística bidimensional, variación conjunta, varianza marginal.

**Notaciones:**

- Denotaremos por  $\bar{x}$  a la media de una variable  $x$ .
- Denotaremos  $\sigma_x$  a la desviación típica de una variable  $x$ ,  $\sigma_x^2$  a su varianza.
- Denotaremos  $\sigma_{xy}$  a la covarianza de las variable  $x$  e  $y$ .
- Denotaremos  $r$  al coeficiente de correlación.
- Denotaremos  $r^2$  al coeficiente de determinación.
- Denotaremos  $\hat{y}(x_0)$  al valor estimado de  $y$  correspondiente a  $x = x_0$  sobre su correspondiente recta de regresión.

**Resultados:**

- La recta de regresión hace mínimos la suma de los cuadrados de las desviaciones de los puntos de la distribución bidimensional a la recta.
- Existen dos rectas de regresión diferentes.
- Propiedades del coeficientes de correlación:
  - Es adimensional y su valor está comprendido entre -1 y 1.

- Si la correlación es perfecta, entonces  $r = 1$  o  $r = -1$ .
- Si la correlación es fuerte,  $|r|$  es próximo a 1.
- Si la correlación es débil,  $|r|$  es próximo a 0.
- Ángulo de las rectas de regresión:
  - Cuando la correlación es casi nula las rectas forman un ángulo próximo a  $90^\circ$ .
  - Si la correlación es fuerte el ángulo que forman ambas rectas es pequeño.
  - Si  $|r|$  es próximo a 1 las rectas son coincidentes.
- La recta de regresión permita estimar valores de la variable dependiente a partir de la independiente.
- La recta de regresión pasa por el centro de gravedad de la distribución.
- Las estimaciones con la recta de regresión son mejores en valores cercanos al punto de medio de la distribución a las obtenidas de valores muy alejados.
- La fiabilidad de la predicción con la recta de regresión aumenta con el tamaño de la muestra.
- El producto de los coeficientes de regresión es  $r^2$ .

**Conceptos:**

- |  |                                     |
|--|-------------------------------------|
| ■ Media, mediana, moda.                | ■ Variable dependiente.             |
| ■ Desviación típica, varianza.         | ■ Variable independiente.           |
| ■ Dependencia funcional.               | ■ Regresión.                        |
| ■ Dependencia estadística o aleatoria. | ■ Modelos no lineales de regresión. |
| ■ Independencia.                       | ■ Rectas de regresión.              |
| ■ Correlación.                         | ■ Coeficientes de regresión.        |
| ■ Covarianza.                          | ■ Coeficiente de determinación.     |
| ■ Coeficientes de correlación lineal.  | ■ Bondad de ajuste.                 |

### 2.1.2. **Ámbito Procedimental**

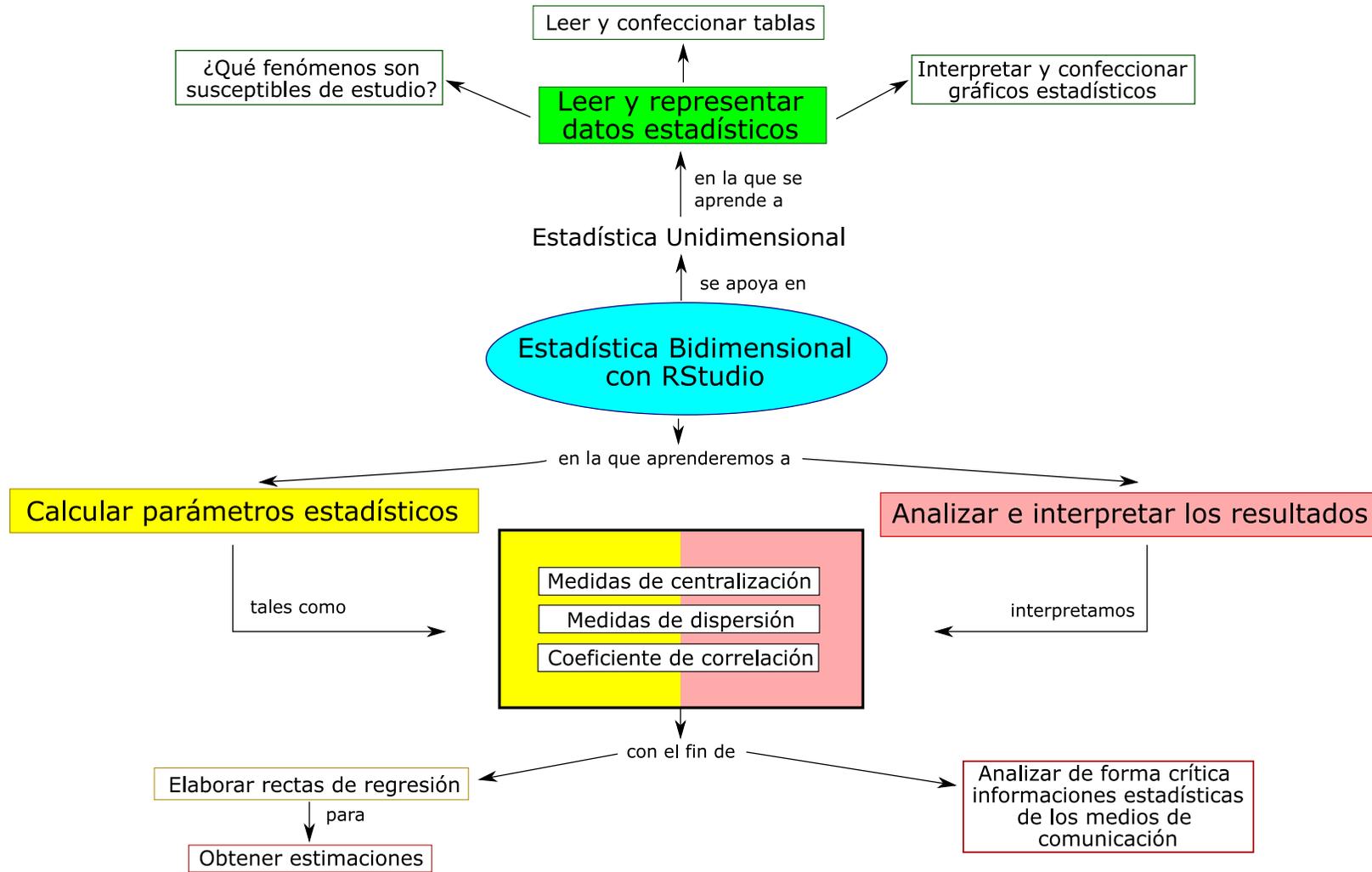
**Procedimientos:**

- Disposición de datos de tablas de frecuencias.
- Cálculo de los parámetros estadísticos a partir de tablas de valores.
- Determinación de la relación que existe entre variables.
- Cálculo del coeficiente de correlación lineal.
- Obtención de la recta de regresión.
- Estimación de una variable a partir de otra.
- Valoración de la bondad del ajuste.
- Estudio conjunto de las rectas de regresión y del coeficiente de correlación para determinar la fiabilidad de las estimaciones.

### 2.1.3. **Ámbito Actitudinal**

- Tener sentido crítico ante las posibles relaciones existentes entre diferentes variables.
- Valorar los métodos estadísticos y gráficos como forma de estudio de las interdependencias entre las realidades sociales.
- Acostumbrarse a contar con varias fuentes de información con el fin de analizar y contrastar una determinada situación.
- Tener gusto por el orden, la precisión y el detalle en la organización y tratamiento de la información.
- Valorar la utilidad de las nuevas tecnologías, en particular R, para crear gráficos y para realizar o comprobar cálculos estadísticos.

## 2.2. Mapa Conceptual



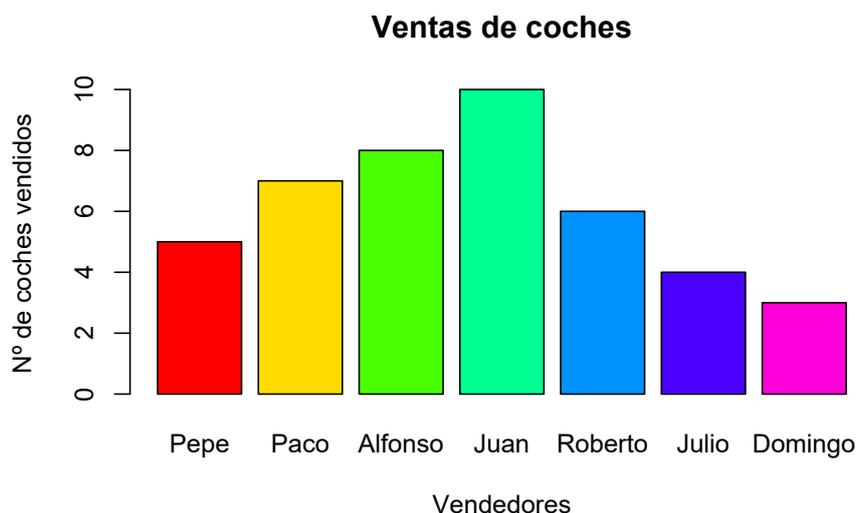
## 2.3. Sistemas de representación

Los conceptos propios del tema de Estadística en 1º de Bachillerato de Ciencias sociales se prestan a ser mostrados mediante cuatro sistemas de representación: verbal, numérico, gráfico y simbólico. A continuación veremos ejemplos de cada uno de dichos sistemas:

- **Verbal:** Una tercera parte de los alumnos de cuarto de la ESO han suspendido, otra tercera parte han sacado un *Bien*, una sexta parte han sacado *Notable* y el resto *Sobresaliente*.
- **Numérico (tablas de valores):** Las ventas, por modelo de coche, en el concesionario han sido las siguientes:

Modelo A	Modelo B	Modelo C	Modelo D	Modelo E
25	40	16	14	55

- **Gráfico:**



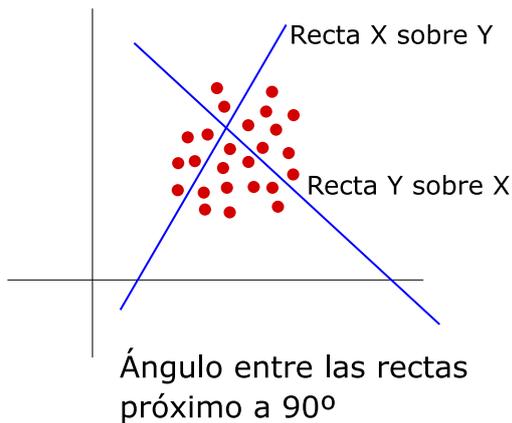
- **Simbólico:** La recta de regresión de la variable X sobre la variable Y viene dada por la expresión :

$$y = 5 + 0,6(x - 6)$$

Un ejemplo muy claro de que un concepto puede verse con distintas representaciones es el concepto de correlación. Veamos varias formas de representar una correlación fuerte entre dos variables.

- Verbal: “Hay correlación fuerte entre X e Y”.

- Gráfica:



- Simbólica:  $|r| \sim 1$

En la siguiente tabla, tomada de [Monroy y Rodríguez \(1998\)](#), podemos ver cómo realizar conversiones entre los distintos sistemas de representación de nuestro tema:

A DE	Situaciones (descripciones verbales)	Tablas	Gráficas	Fórmulas
Situaciones (descripciones verbales)		Medir	Dibujar	Modelar
Tablas	Leer		Marcar puntos	Ajustar
Gráficas	Interpretar	Terminación de una lectura		Ajustar curvas
Fórmulas	Reconocer parámetros	Calcular	Dibujar	

## 2.4. Situaciones

El marco teórico del estudio PISA considera cuatro tipos de situaciones: personales, laborales, sociales y científicas. Veamos ejemplos referentes a nuestro tema de cada uno de los tipos de situaciones:

### 1. Personales:

- Interpretación y análisis de datos de actualidad publicados en los medios de comunicación.
- Recuento y organización de datos.

### 2. Laborales/Educativas:

- Predicción de ganancias al invertir cierta cantidad de dinero.
- Rendimiento de los alumnos en un examen.
- Estudiar si un método de enseñanza es beneficioso o no.
- Estimar el gasto de combustible en un avión tras un número de vuelos.

### 3. Sociales:

- Representar de forma rápida y visual información para su posterior difusión o publicación.
- Estudios sociológicos.

### 4. Científicas:

- Estudios clínicos.
- Determinar si un medicamento es eficaz para tratar una enfermedad.

### 3.1. Expectativas de aprendizaje

#### Foco 1: Leer y representar datos estadísticos

- O1. Distinguir los tipos de fenómenos sociales que son susceptibles de ser tratados estadísticamente.
- O2. Leer y confeccionar tablas estadísticas.
- O3. Interpretar y confeccionar gráficos estadísticos.

#### Foco 2: Calcular parámetros estadísticos

- O4. Hallar las medidas de centralización.
- O5. Hallar las medidas de dispersión.
- O6. Calcular el coeficiente de correlación.
- O7. Calcular las rectas de regresión.

#### Foco 3: Analizar e interpretar de forma crítica los resultados.

- O8. Interpretar las medidas de centralización.
- O9. Interpretar medidas de dispersión.
- O10. Conocer el significado de la correlación entre variables.
- O11. Interpretar el coeficiente de correlación.
- O12. Hacer estimaciones a partir de la recta de regresión.
- O13. Analizar e interpretar de forma crítica informaciones estadísticas presentes en los medios de comunicación, la publicidad y otros ámbitos, detectando posibles errores y manipulaciones tanto en la presentación de los datos como de las conclusiones.

**Objetivo** común a todos los focos

**O14.** Introducir y utilizar el software R en el aula.

En las siguientes tablas mostraremos qué competencias básicas matemáticas, de las recogidas en el informe PISA , se trabajan en mayor medida con cada uno de los objetivos de nuestra Unidad Didáctica.



Figura 3.1: Representación gráfica de los focos de objetivos.

Objetivos	RA	C	M	RP	R	LS	HM
O1. Distinguir los tipos de fenómenos sociales que son susceptibles de ser tratados estadísticamente.	✓						
O2. Leer y confeccionar tablas estadísticas.	✓	✓	✓		✓	✓	✓
O3. Interpretar y confeccionar gráficos estadísticos.	✓	✓	✓		✓		✓
O4. Hallar las medidas de centralización.		✓		✓		✓	✓
O5. Hallar las medidas de dispersión.		✓		✓		✓	✓
O6. Calcular el coeficiente de correlación.		✓		✓			✓
O7. Calcular las rectas de regresión.		✓		✓	✓	✓	✓

Cuadro 3.1: Objetivos y competencias de los focos 1 y 2.

Razonar y argumentar (**RA**), Comunicar (**C**), Matematizar (**M**), Elaborar estrategias para resolver problemas (**RP**), Representar (**R**), Utilizar operaciones y un lenguaje simbólico, formal y técnico (**LS**), Usar herramientas matemáticas (**HM**).

Objetivos	RA	C	M	RP	R	LS	HM
O8. Interpretar las medidas de centralización.	✓			✓			
O9. Interpretar las medidas de dispersión.	✓			✓			
O10. Conocer el significado de la correlación entre variables.	✓						
O11. Interpretar el coeficiente de correlación.	✓			✓			
O12. Hacer estimaciones a partir de la recta de regresión	✓	✓	✓	✓		✓	
O13. Analizar e interpretar de forma crítica informaciones estadísticas presentes en los medios de comunicación, la publicidad y otros ámbitos, detectando posibles errores y manipulaciones tanto en la presentación de los datos como de las conclusiones.	✓	✓				✓	
O14. Introducir y utilizar el software R en el aula.			✓	✓	✓		✓

Cuadro 3.2: Objetivos y competencias del foco 3 y R.

Razonar y argumentar (**RA**), Comunicar (**C**), Matematizar (**M**), Elaborar estrategias para resolver problemas (**RP**), Representar (**R**), Utilizar operaciones y un lenguaje simbólico, formal y técnico (**LS**), Usar herramientas matemáticas (**HM**).

Las competencias que más se trabajan son “Razonar y argumentar” y “Resolver problemas”. Éstas serán potenciadas por el uso de R, ya que podremos dedicar más tiempo a ellas en detrimento de realizar cálculos. En el otro extremo se encuentran las competencias “Matematizar” y “Representar”. Ambas, a pesar de aparecer en pocas ocasiones, serán trabajadas constantemente en los problemas de correlación y regresión, los cuales son el grueso de la unidad.

## 3.2. Limitaciones del aprendizaje

En la siguiente tabla recogemos los principales errores y dificultades derivadas de nuestro tema y los objetivos asociados a ellos. Para la realización de la misma se ha tenido en cuenta mi experiencia personal y algunos de los errores mencionados en el artículo “Conflictos semióticos sobre la correlación y regresión en los libros de texto de Bachillerato” (Gea et al., 2015b).

Errores y dificultades	Objetivos asociados
ED1. Malinterpretar las tablas estadísticas.	O2
ED2. Utilización de gráficos estadístico no adecuados.	O3
ED3. Errores a la hora de confeccionar un gráfico estadístico.	O3
ED4. Pensar que únicamente existe una recta de regresión.	O7, O12
ED5. Obtener una de las rectas de regresión despejando a partir de la otra.	O7
ED6. Lenguaje simbólico inapropiado o impreciso.	O6, O7, O8, O9
ED7. Creer que causalidad y correlación son equivalentes.	O10
ED8. Pensar que un bajo coeficiente de correlación lineal implica que no haya dependencia entre las variables (podría haber, por ejemplo, dependencia curvilínea).	O10, O11
ED9. Utilizar rectas con poco coeficiente de correlación para predecir valores.	O11, O12
ED10. Errores de cálculo a la hora de obtener los parámetros estadísticos.	O4, O5, O6, O7
ED11. Dar por cierta toda aquella información precedente de los medios de comunicación sin cuestionarla.	O2, O3, O13
ED12. Dificultad de la sintaxis de R.	O14

Además de las mencionadas dificultades cabe señalar la propia dificultad por parte del profesor a la hora de usar el software. No todos los docentes poseen los conocimientos tecnológicos suficientes para su uso, por lo cual habría que formar al profesorado involucrado.

### 3.3. Oportunidades de aprendizaje

Con el fin de evitar y corregir los errores y dificultades descritos en la sección anterior se tratará de realizar tareas que:

1. Pongan en evidencia la falta de idoneidad de ciertos gráficos para la representación de algunas informaciones.
2. Necesiten la conversión de datos a otro de los sistemas de representación explicados.
3. Pidan calcular ambas rectas de regresión.
4. Pongan de manifiesto que dependencia y correlación no son lo mismo (pondremos ejemplos donde sea claro que no lo es).
5. Muestren variables que, a pesar de tener una baja correlación lineal, tengan dependencias de otro tipo (logarítmica, cuadrática ...).
6. Muestren gráficos e informaciones incorrectos publicados en medios de comunicación para que los alumnos identifiquen el error o posible manipulación correspondiente.
7. Trabajen la sintaxis de R, por ejemplo pidiendo al alumno que detecte errores en código.

Asimismo, es conveniente que se realicen tareas que involucren el mayor número de objetivos posibles ya que, como se puede comprobar, están íntimamente relacionados. Ésto dará pie a la creación de ejercicios de conexión, gracias a los cuales el alumno podrá apreciar la importancia de todos y cada uno de los objetivos a la hora de resolver los problemas. Finalmente, se intentarán utilizar, siempre que se pueda, datos de actualidad y de interés a la hora de realizar análisis e interpretaciones de informaciones estadísticas.

## 4 ANÁLISIS DE INSTRUCCIÓN

Puesto que este trabajo es una Validación de Material y no una Unidad Didáctica, nos centramos en el uso y papel de R en vez de en la gestión de las propias tareas. Además de ello, incluimos en el anexo tanto las prácticas teóricas en las que se enseñará el funcionamiento de RStudio como algunas de las soluciones a las tareas que mandaremos a los alumnos para asentar los conocimientos que les enseñaremos.

### Tarea 1: Aprendizaje de los conceptos básicos de Rstudio

Esta primera tarea se realizará una vez que se haya introducido en clase el concepto de distribución bidimensional. Para realizar esta tarea necesitaremos llevar al alumnado a un aula de ordenadores, en caso de que el centro disponga de ella, o bien que el alumnado tenga ordenadores (al menos 1 por cada dos alumnos). La duración de la tarea será de una sesión entera y los alumnos aprenderán a instalar RStudio y a usar algunas funciones básicas. A los alumnos se les proveerá de la ficha “Práctica 1: Nociones básicas de RStudio” (véase en el Anexo). Tras finalizar la clase se animará a los alumnos a que instalen RStudio en el ordenador de su casa para así poder practicar de forma autónoma.

### Tarea 2: Estadística descriptiva con RStudio

Esta tarea se realizará una o dos sesiones después de realizar la primera práctica de ordenador. Los contenidos matemáticos necesarios (gráficos estadísticos) ya fueron vistos con anterioridad en el tema dedicado al repaso de la estadística descriptiva unidimensional. Los requerimientos para la tarea son los mismos que en la tarea anterior: al menos un ordenador para cada dos alumnos. Durará, al igual que la anterior práctica, una sesión y el alumno tendrá como guía la ficha “Práctica 2: Estadística descriptiva con Rstudio”(véase Anexo).

### Tarea 3: Gráficos estadísticos erróneos.

Esta tarea se puede realizar una vez que se haya realizado la práctica 2. La tarea persigue, esencialmente, dos objetivos: analizar de forma crítica los gráficos presentes en medios de comunicación y la elaboración de gráficos estadísticos haciendo uso de RStudio. El enunciado de la tarea es el siguiente:

Los siguientes gráficos contienen errores o transmiten sensaciones incorrectas. Identifica dichos fallos y propón una solución para ellos. Finalmente, elige tres de los gráficos erróneos anteriores y corrígelos haciendo uso de RStudio.



Gráfico 1



Gráfico 2

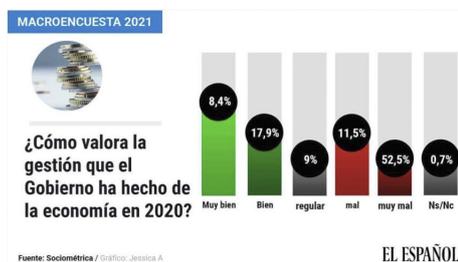


Gráfico 3

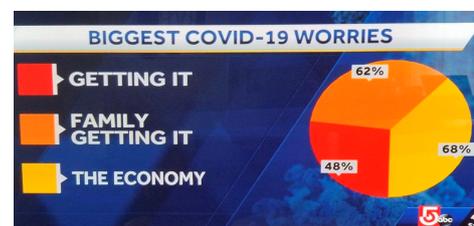


Gráfico 4



Gráfico 5

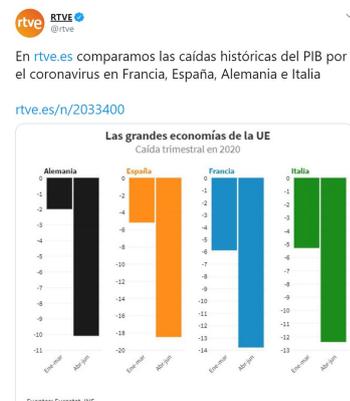


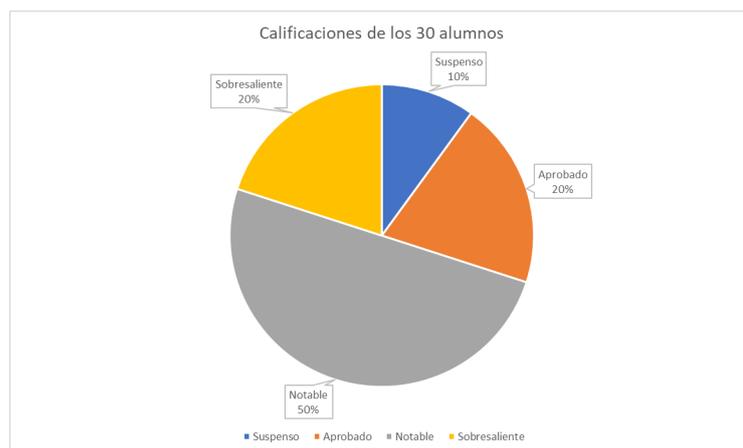
Gráfico 6

Para analizar cada una de las tareas (a excepción de las prácticas de ordenador) usaremos el siguiente tipo de tabla, el cual está inspirado en el visto en **Rico y Moreno (2016)** (página 305). Para las prácticas de ordenador no considero que sea necesario su uso, ya que todas ellas tienen la misma filosofía: aprender a utilizar RStudio, siguiendo la explicación del profesor y realizando pertinentes ejemplos, con el fin de usarlo en los problemas propios del tema de estadística.

<b>Tarea 3.</b>		<b>Duración aprox. : 30-40 min.</b>
<b>Elementos de la tarea</b>	<b>Meta</b>	Trabajo autónomo con RStudio de los alumnos y análisis crítico de informaciones de los medios de comunicación.
	<b>Recursos</b>	Ordenador (al menos 1 por cada 2 alumnos).
	<b>Objetivos trabajados</b>	O3, O13, O14
	<b>Superación de dificultades</b>	ED2, ED3, ED11, ED12
	<b>Situación</b>	Personal.
	<b>Complejidad</b>	Reflexión.
<b>Condiciones</b>	<b>Presentación</b>	Enunciado descrito anteriormente.
	<b>Comunicación (cómo-cuándo)</b>	Se podrá debatir en parejas acerca de los fallos y posibles soluciones.
	<b>Agrupamiento</b>	Por parejas.
<b>Papel de R</b>	Realización de gráficos estadísticos	
<b>Observaciones</b>	En caso de que haya ordenadores para todos se realizará de forma individual. Los alumnos pueden también proponer gráficos vistos en medios de comunicación para ser analizados en clase.	

## Tarea 4: Ejercicio práctico de estadística descriptiva

De nuevo esta tarea corresponde al tema de repaso de estadística descriptiva. Los apartados 1), 3) y 4) son teóricos y los apartados 5 y 6 son los que se prestan a hacer uso de R. El enunciado de la tarea es el siguiente:



1. ¿Qué tipo de gráfico es el anterior?
2. Elabora una tabla de frecuencias absolutas y frecuencias absolutas acumuladas.
3. ¿Cuánta gente tiene notable o sobresaliente?
4. ¿Es posible calcular la nota media de los alumnos a partir del gráfico dado? Si es posible calcúlala, en caso contrario justifica tu respuesta.
5. Elabora un gráfico de barras a partir de los datos de los alumnos.
6. Supongamos que todos los *suspensos* sacaron un 3, todos los *aprobados* un 6, todos los *notables* un 8 y todos los *sobresaliente* un 10. Calcula la media, mediana, desviación típica y varianza de las notas de los alumnos.

Con esta tarea lo que se busca es que el alumno se familiarice con la creación de gráficos estadísticos y el cálculo de parámetros en RStudio a la vez que repasa lo visto en años anteriores. Se recomienda que los cálculos también se hagan a mano.

Pasamos ahora a analizar la tarea usando nuestra tabla:

Tarea 4.		Duración aprox. : 25-30 min.
<b>Elementos de la tarea</b>	<b>Meta</b>	Repasar la estadística descriptiva vista en la ESO a la vez que practicar con RStudio.
	<b>Recursos</b>	Ordenador (al menos 1 por cada 2 alumnos).
	<b>Objetivos trabajados</b>	O2, O3, O4, O5, O14
	<b>Superación de dificultades</b>	ED2, ED3, ED10, ED12
	<b>Situación</b>	Personal
	<b>Complejidad</b>	Reproducción-Conexión.
<b>Condiciones</b>	<b>Presentación</b>	Enunciado ya descrito.
	<b>Comunicación (cómo-cuándo)</b>	Tras realizar la actividad se pondrán en común los resultados y se resolverán las posibles dudas.
	<b>Agrupamiento</b>	Individual-parejas.
<b>Papel de R</b>	- Realización de gráficos estadísticos y cálculo de parámetros.	
<b>Observaciones</b>	<ul style="list-style-type: none"> <li>- La agrupación será individual en las preguntas teóricas y, en aquellas que sea necesario usar R por parejas.</li> <li>- Los cálculos, puesto que son sencillos, es conveniente que se realicen tanto a mano como a ordenador (para verificar).</li> </ul>	

## Tarea 5: Regresión y correlación con RStudio

Esta tarea podrá ser realizada una vez que se vea en clase todo el contenido teórico acerca de la correlación y la regresión de las distribuciones bidimensionales. El objetivo de esta práctica es que el alumnado aprenda a realizar los cálculos haciendo uso del ordenador con el fin de que pueda comprobar que sus cálculos hechos a mano están bien y para agilizar el monótono proceso de calcular los distintos parámetros. Adicionalmente permite al alumnado manipular los rápidamente los datos del problema y poder analizar las consecuencias de dichos cambios. Para este fin se necesitará, como en las dos prácticas anteriores, al menos un ordenador por cada dos alumnos. Además se les dará a los alumnos la ficha “Práctica 3: Regresión y correlación con R” la cual resume los contenidos que se darán durante la sesión (véase Anexo).

## Tarea 6: Correlación y causalidad

La siguiente tarea es la continuación natural de la tarea 5. En ella se pretende que el alumno haga uso de lo visto en la práctica 3 y, además, aprenda que correlación y causalidad no son conceptos equivalentes. Para ello dotaremos al alumno de los siguientes datos<sup>1</sup>:

Y	8.2	7	6.5	5.3	5.2	4	4.6	4.5	4.2	3.7
X	5	4.7	4.6	4.4	4.3	4.1	4.2	4.2	4.2	4.1

Tras ello les haremos las siguientes preguntas:

- Calcula una recta de regresión de los datos y gráficala con RStudio.
- ¿Qué coeficiente de determinación tiene dicha recta?
- A partir de los datos, ¿podemos suponer que X causa Y? ¿Y viceversa?

Cuando los alumnos realicen la actividad obtendrán una recta de regresión con un coeficiente de determinación muy cercano a uno, es decir, existe una correlación

<sup>1</sup>Datos obtenidos de <http://www.tylervigen.com/spurious-correlations>.

muy fuerte entre la variable X y Y. Por tanto, algunos alumnos probablemente asuman que, en consecuencia, una variable causa la otra. Tras debatir ésto con los alumnos el profesor les dirá a los alumnos que la variable X corresponde al ratio de divorcios en la ciudad de Maine y la Y al consumo de margarina, en libras, por habitante. Ahí verán que, evidentemente, las dos variables no guardan ninguna relación. Finalmente, el profesor explicará que si dos variables X e Y están correlacionadas puede ser porque:

1. X causa a Y.
2. Y causa a X.
3. Hay una tercera variable que causa a X e Y.
4. Pura coincidencia.

Tarea 6.		Duración aprox. : 30 min.
<b>Elementos de la tarea</b>	<b>Meta</b>	Poner en uso la práctica 3 a la vez que aprenden que causalidad y correlación no es lo mismo.
	<b>Recursos</b>	Ordenador (al menos 1 por cada 2 alumnos).
	<b>Objetivos trabajados</b>	O2, O6, O11, O14.
	<b>Superación de dificultades</b>	ED7, ED12.
	<b>Situación</b>	Académica.
	<b>Complejidad</b>	Reproducción-Reflexión.
<b>Condiciones</b>	<b>Presentación</b>	Enunciado ya descrito.
	<b>Comunicación (cómo-cuándo)</b>	La actividad consta de dos partes, una en la que los alumnos realizan los cálculos y otra en la que el profesor desvela el significado de las variables para su posterior discusión.
	<b>Agrupamiento</b>	Individual-parejas.
<b>Papel de R</b>	- Cálculo de la recta de regresión y coeficiente de determinación.	

## Tarea 7: Realización de predicciones

En la siguiente tarea usaremos de nuevo lo visto en la práctica 3 y, además, haremos predicciones haciendo uso de la recta de regresión que calcularemos. El enunciado de la tarea es el siguiente:

Se midieron los valores de concentración, en microgramos por centímetro cúbico, de una sustancia A en suero fetal y los valores de su concentración en suero materno. Se obtuvieron los siguientes datos en una muestra de seis embarazadas al final de la gestación.<sup>2</sup>

<b>Concentración suero madre (X)</b>	8	4	12	2	7	9
<b>Concentración suero feto (Y)</b>	6	4	8	1	4	5

- Halla la recta de regresión que permita estimar los valores fetales a partir de los maternos.
- Halla el coeficiente de determinación e interprétalo para estudiar la bondad del ajuste.
- Determina la previsión de sustancia A en el suero fetal que se obtendrá si en el suero materno se han obtenido 10 microgramos por centímetro cúbico.

Como se puede observar en la solución de la tarea (véase Anexo), el cálculo de la recta de regresión y de los coeficientes de correlación y determinación es muy sencillo y rápido. Es por ello que el alumno se centrará más en “saber qué está haciendo” que en calcular los correspondientes parámetros (no obstante, el alumno debe terminar el tema sabiendo calcular a mano los distintos coeficientes y la propia recta de regresión).

---

<sup>2</sup>Enunciado basado en un ejercicio del libro de SM (Alcaide et al., 2015).

Pasemos ahora a analizar la tarea haciendo uso de nuestra tabla:

Tarea 7.		Duración aprox. : 25-30 min.
<b>Elementos de la tarea</b>	<b>Meta</b>	Asentar los conocimientos adquiridos en la tercera práctica a la vez que se trabajan los contenidos de regresión y predicción del tema.
	<b>Recursos</b>	Ordenador (al menos 1 por cada 2 alumnos).
	<b>Objetivos trabajados</b>	O2, O6, O7, O11, O12, O14
	<b>Superación de dificultades</b>	ED9, ED12
	<b>Situación</b>	Científica
	<b>Complejidad</b>	Reproducción
<b>Condiciones</b>	<b>Presentación</b>	Enunciado ya descrito
	<b>Comunicación (cómo-cuándo)</b>	Tras hacer el ejercicio se pondrán en común los resultados y se comentarán los posibles errores.
	<b>Agrupamiento</b>	Individual-parejas.
<b>Papel de R</b>	- Realización de gráficos estadísticos y cálculo de parámetros.	
<b>Observaciones</b>	- La agrupación será individual en las preguntas teóricas y, en aquellas que sea necesario usar R, por parejas. - Los cálculos, puesto que son sencillos, es conveniente que se realicen tanto a mano como a ordenador (para verificar).	

## Tarea 8: ¿Cómo varía la media y la desviación típica si modificamos los datos de partida?

El objetivo de la siguiente tarea es que el alumno compruebe de manera computacional las siguientes propiedades.

- Si la media de unos datos,  $X$ , es  $\mu$ , entonces la media de  $X + b$  es  $\mu + b$ .
- Si la varianza de unos datos,  $X$ , es  $\sigma^2$ , entonces la varianza de  $X + b$  sigue siendo la misma, es decir  $\sigma^2$ .
- Si la media de unos datos,  $X$ , es  $\mu$ , entonces la media de  $a * X$  es  $a * \mu$ .
- Si la varianza de unos datos,  $X$ , es  $\sigma^2$ , entonces la varianza de  $a * X$  es  $a^2 * \sigma^2$ .

Para ello se le dará al alumno un script (véase en el Anexo la Tarea 8) en el cual podrá modificar tanto los datos como los parámetros  $a$  y  $b$ . Al ejecutarlo se darán cuenta de que las propiedades mencionadas anteriormente son ciertas.

Dependiendo del nivel de aprendizaje del alumnado (tanto de la teoría como de programación) se podrán realizar las siguientes variantes:

1. Los alumnos usan el script dado y no se demuestra teóricamente.
2. Los alumnos usan el script dado y se demuestra teóricamente.
3. Los alumnos realizan ellos mismos el script y no se demuestra teóricamente.
4. Los alumnos realizan ellos mismos el script y demostrarlo teóricamente.

La dificultad de realizar el script por parte del alumno es media-baja.

Esta es la tabla de análisis correspondiente a esta tarea:

<b>Tarea 8. Duración aprox. : 20-40 min. (según variante)</b>		
<b>Elementos de la tarea</b>	<b>Meta</b>	Aprender propiedades sobre la media y la varianza de una variable haciendo uso de Rstudio.
	<b>Recursos</b>	Ordenador (al menos 1 por cada 2 alumnos).
	<b>Objetivos trabajados</b>	O4, O5, O8, O9, O14
	<b>Superación de dificultades</b>	E10, E12
	<b>Situación</b>	Educativa
	<b>Complejidad</b>	Reproducción
<b>Condiciones</b>	<b>Presentación</b>	Enunciado ya descrito
	<b>Comunicación (cómo-cuándo)</b>	Antes de mostrar las propiedades a probar se preguntará al alumno para que trabajen su intuición.
	<b>Agrupamiento</b>	Individual-parejas.
<b>Papel de R</b>	- Realizar los cálculos para comprobar que las propiedades se cumplen independientemente de los parámetros y de los datos.	
<b>Observaciones</b>	<p>- Si se considera necesario pueden graficarse datos, media y varianza para que lo comprueben tanto de forma gráfica como numérica.</p> <p>- También puede plantearse la tarea como explicación, es decir, es únicamente el profesor el que maneja el script y lo proyecta a los alumnos explicándolo. Esto reduciría los recursos necesarios a un único ordenador con proyector.</p>	

## Tarea 9: Importación de grandes cantidades de datos con RStudio

Hasta el momento hemos aprendido a graficar y a hacer cálculos haciendo uso de datos que hemos introducido de forma manual. Al hacerlo de forma manual no hemos podido introducir grandes cantidades de datos y, en consecuencia, no hemos utilizado todo el potencial que nos ofrece RStudio. En esta cuarta práctica, cuyo guión se puede consultar en el Anexo, el alumno aprenderá a importar grandes cantidades de datos para su posterior tratamiento. El ejemplo que trataremos serán los datos de población de los países del mundo desde el año 1800 hasta 2019. En primer lugar, seleccionaremos las poblaciones de España y Argentina y luego haremos la serie temporal de ambos países. Del mismo modo se podría haber estudiado, por ejemplo, si hay correlación entre las poblaciones de ambos países ya que, una vez importados los datos, podemos hacer con ellos todo aquello que eramos capaces en las prácticas anteriores. Para practicar éstos conocimientos recién aprendidos se les propondrá a los alumnos realizar la tarea 10.

## Tarea 10: Internet y escolarización

La tarea consistirá en darles a los alumnos una base de datos en la cual se relaciona el porcentaje de personas con acceso a Internet y la edad media a la que abandonan la escuela en los países del mundo. El objetivo de los alumnos es importar dichos datos haciendo uso de lo visto en la práctica anterior y estudiar si ambas variables tienen o no correlación. Para ello calcularán el coeficiente de correlación y dibujarán tanto los datos como la recta de regresión que escojan. Además de ello se les pedirá que saquen una conclusión acerca de los resultados. Finalmente, se les recomendará a los alumnos la página web <https://ourworldindata.org/>, de la cual hemos obtenido todos los archivos .csv, para que sigan practicando.

<b>Tarea 10. Duración aprox. :</b>		
<b>Elementos de la tarea</b>	<b>Meta</b>	Afianzar lo aprendido en la tarea anterior.
	<b>Recursos</b>	Ordenador (al menos 1 por cada 2 alumnos).
	<b>Objetivos trabajados</b>	O7,O14
	<b>Superación de dificultades</b>	ED12
	<b>Situación</b>	Personal
	<b>Complejidad</b>	Reproducción-Reflexión
<b>Condiciones</b>	<b>Presentación</b>	Enunciado ya descrito
	<b>Comunicación (cómo-cuándo)</b>	Tras hacer el ejercicio se discutirá en clase las conclusiones obtenidas.
	<b>Agrupamiento</b>	Individual-parejas.
<b>Papel de R</b>	Cálculo de la recta de regresión e importación de datos.	

## Tarea 11: Creación de informes con RStudio haciendo uso del paquete RMarkdown

La tarea consistirá en enseñar a los alumnos cómo generar informes con RMarkdown con el objetivo de que puedan presentar la tarea final haciendo uso de este paquete. Para la explicación utilizaremos cualquier de los programas ya realizados, el cual copiaremos y pegaremos un archivo .Rmd (extensión de los archivos RMarkdown). Tras esto, les enseñaremos cómo tienen que escribir en el archivo para diferenciar entre lo que es código de R y lo que no y, finalmente, les señalaremos el botón que tienen que utilizar para que RStudio les genere automáticamente el documento. Como ya se ha comentado en el trabajo, todos los archivos presentes en el Anexo han sido generados haciendo uso de este paquete, por lo cual servirían para ejemplificar a los alumnos el resultado de un posible informe.

Tarea 11.		Duración aprox. : 20-30 min.
<b>Elementos de la tarea</b>	<b>Meta</b>	Generar informes para presentar las tareas .
	<b>Recursos</b>	Ordenador (al menos 1 por cada 2 alumnos).
	<b>Objetivos trabajados</b>	O14
	<b>Superación de dificultades</b>	ED12
	<b>Situación</b>	Educativa y laboral
	<b>Complejidad</b>	Reproducción-Reflexión
<b>Condiciones</b>	<b>Presentación</b>	Enunciado ya descrito
	<b>Comunicación (cómo-cuándo)</b>	El profesor explicará el uso del paquete mientras los alumnos practican.
	<b>Agrupamiento</b>	Individual-parejas.
<b>Papel de R</b>	Generación de informes con código R.	

## 4.1. Atención a la diversidad

La siguiente y última tarea, la cual será objeto de evaluación como veremos en el siguiente capítulo, estará estructurada en tres niveles de complejidad. De tal forma, todos los alumnos serán capaces de realizarla independientemente de su nivel y daremos pie a que aquellos con más nivel puedan seguir desarrollando sus conocimientos.

La tarea en cuestión consiste en investigar sobre un tema propuesto por el profesor (o por el alumno si el profesor da el visto bueno), hacer un informe sobre el mismo, realizar gráficos que ilustren los datos y sacar conclusiones haciendo uso de los conocimientos tratados a lo largo del tema.

Ejemplos de temas posibles:

- “Horas de estudio diarias y resultados en el examen de Geografía ”
- “¿Realizar deporte ayuda a mejorar el rendimiento académico? ”
- “¿Influye la cantidad de tiendas de IKEA en el número de premios Nobel de un país?”

Este trabajo se realizará en grupos de 4 personas, los cuales tendrán que exponerlo a sus compañeros una vez realizado el trabajo.

Para ilustrar la tarea realizaremos un ejemplo de dicha tarea con el tema “Estudio sobre la relación entre las notas en matemáticas y lengua de los alumnos de 1ºA ”:

Los alumnos recibirán las notas de cada uno de los alumnos, sin nombres, en Lengua y Matemáticas. Una vez que los tengan tendrán que realizar una tabla en la cual recojan toda esta información y realizar gráficos para ilustrar dichos datos. Tras ello, tendrán que aventurarse a conjeturar si hay relación o no entre los datos y, en caso de que la haya, decir de qué tipo. Para ello tendrán que hacer uso de todas los conocimientos vistos en el tema y podrán utilizar como apoyo todo lo aprendido con RStudio (gráficos, cálculo de rectas de regresión ...).

Veamos los distintos niveles de complejidad establecidos:

- **Nivel 1 de complejidad:** Los alumnos que no tengan suficiente dominio con Rstudio realizarán dicha tarea a mano, aunque se les permitirá eventualmente hacer uso del mismo para realizar los cálculos pertinentes.
- **Nivel 2 de complejidad:** Aquellos alumnos con cierto dominio en Rstudio realizarán los cálculos y gráficos haciendo uso del mismo, los cuales incluirán en su informe.
- **Nivel 3 de complejidad:** En caso de que dominen todos los conocimientos que se explicarán de RStudio los alumnos tendrán que realizar el informe haciendo uso del paquete *RMarkdown*, generando un pdf con toda la información, gráficos y cálculos.

## 5 ANÁLISIS EVALUATIVO

A la hora de evaluar la unidad se tendrá en cuenta el examen que se realice al final de la misma y una serie de tareas que se realizarán a lo largo de las sesiones. La nota del examen tendrá un ponderación del 70% en la nota final y las tareas un 30%.

Una de estas tareas es la vista en el apartado de atención a la diversidad en la cual, como ya se ha explicado, le propondremos a cada grupo de alumnos una serie de temas para que investiguen de forma autónoma, obtengan conclusiones estadísticas relevantes y realicen un informe.

La heterogeneidad en el nivel de dominio de las nuevas tecnologías podría conllevar que un alumno, a pesar de dominar los contenidos estadísticos del tema, no pudiese realizar la tarea de forma adecuada. Es por ello que se considerarán 3 niveles de complejidad, en lo referente al dominio tecnológico, de forma que cada uno de los alumnos podrá escoger aquel que mejor se adecue a sus destrezas. Ciertamente es que, como se podrá ver a continuación en la rúbrica de la tarea escrita, la elección de niveles de complejidad inferior conllevará una reducción en la nota destinada al uso de la tecnología, sin embargo el porcentaje destinada al mismo no es especialmente significativo (15%).

Finalmente, como se comenta en la tarea, los alumnos tendrán que exponer el trabajo al resto de sus compañeros, trabajando así la competencia lingüística (CLL). Tras la exposición tanto el resto de sus compañeros como el profesor les harán preguntas acerca de su trabajo. Para la evaluación de la exposición oral se seguirá también una rúbrica.

En caso de contar con un alumno de altas capacidades se le podría introducir el uso de la regresión exponencial y proponerle una tarea en la cual los datos no ajusten bien linealmente pero sí exponencialmente. De esta forma reforzaría el **ED8** a la vez que aprende contenido adicional.

<b>RÚBRICA TAREA DIVERSIDAD (TRABAJO ESCRITO)</b>				
<b>Categoría/Criterio</b>	<b>Niveles de Logro</b>			
	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>
<b>Comprensión de datos (30%)</b>	No interpreta los datos	Interpreta parte de los datos.	Interpreta la totalidad de los datos.	Interpreta la totalidad de los datos y saca conclusiones sobre los mismos.
<b>Cálculos (25%)</b>	La mayoría de los cálculos son incorrectos.	Hay algunos cálculos incorrectos.	Los cálculos son correctos.	Los cálculos son correctos y están expresados adecuadamente.
<b>Presentación (10%)</b>	Desordenado y sucio.	Desordenado en ciertas partes.	Ordenado y limpio.	Ordenado, limpio y vistoso.
<b>Uso de la tecnología (15%)</b>	No hace uso de ella.	Uso anecdótico.	Hace uso adecuado de ella.	Genera el informe y los cálculos haciendo uso de ella.
<b>Justificaciones (20%)</b>	No justifica.	Los justificaciones son escasas o vagas.	Justificaciones adecuadas.	Justificaciones concisas y muy convenientes.

<b>RÚBRICA TAREA DIVERSIDAD (EXPOSICIÓN)</b>			
<b>Categoría/Criterio</b>	<b>Niveles de Logro</b>		
	<b>1</b>	<b>2</b>	<b>3</b>
<b>Conocimiento y preparación del tema (20%)</b>	No muestra control sobre el tema.	Domina el tema pero no responde adecuadamente a las preguntas.	Domina el tema y contesta de forma precisa.
<b>Estructura y orden (30%)</b>	Exposición desorganizada y confusa.	Exposición organizada, no capta totalmente la atención.	Exposición organizada, capta la atención.
<b>Uso de vocabulario estadístico (30%)</b>	No usa el vocabulario estadístico propio de la unidad.	Utiliza el vocabulario estadístico en la mayoría de las ocasiones.	Utiliza el vocabulario estadístico en toda la exposición de forma precisa.
<b>Aspectos técnicos (formas, tiempo, diapositivas, interacciones) (20%)</b>	Carencias importantes.	Diseño correcto.	Presentación sobresaliente.

## 6 CONCLUSIONES

El uso de del software R, en particular RStudio, no es un tema nuevo. Trabajos como los de [Armero \(2019\)](#), [Briz y Serrano \(2018\)](#), [Reuelta \(2017\)](#) y [Alemany \(2015\)](#) ya proponen su uso para la enseñanza de distintas ramas de las matemáticas: álgebra, análisis y la propia estadística. En cierta forma, este trabajo surgió inspirado en [Alemany \(2015\)](#), que proponía la enseñanza de la estadística de primero de bachillerato de ciencias sociales aunque, su caso, a partir de proyectos.

La idea que perseguía el trabajo, la cual creo que se ha alcanzado, era explicar dicha Unidad Didáctica como se explicaría normalmente pero introduciendo de forma natural el software R tal y como se hace, por ejemplo, con GeoGebra en los bloques de geometría. Introducir el uso de RStudio, como se ha comentado a lo largo del trabajo, tiene una serie de dificultades añadidas, sin embargo considero que los beneficios compensan lo suficiente como para que merezca la pena.

Gracias a Rstudio hemos podido, a partir de grandes cantidades de datos, confeccionar numerosos gráficos, calcular parámetros estadísticos, obtener las correspondientes rectas de regresión y, haciendo uso de ellas, realizar predicciones. Finalmente, gracias al paquete RMarkdown, hemos sido capaces de exportar todos estos gráficos y cálculos a formato PDF, con el fin de poder guardar y compartir nuestro trabajo.

Si conseguimos que el alumnado consiga manejo con RStudio podremos usarlo en el tema de probabilidad, como se puede ver, por ejemplo, en [Reuelta \(2017\)](#), ya que es muy útil para realizar simulaciones de experimentos probabilísticos. En consecuencia, sería muy interesante, en mi opinión, que el tema de estadística bidimensional no fuese simplemente un tema en el cual se pueda utilizar RStudio, sino el primero de una serie de temas de estadística y probabilidad que se traten con este software. Además de ello, como ya se comentó, mejoraremos el pensamiento computacional del alumno, lo cual le ayudará a la hora de razonar y, adicionalmente, a programar de forma más

eficiente en caso de que lo necesitase en la universidad o en su futuro profesional.

Habría sido muy favorable para el trabajo haber podido explicar estos contenidos, haciendo uso de R, durante las prácticas en el centro, ya que se podrían haber visto de forma más clara tanto los errores y dificultades como las virtudes de trabajar con este software. Sin embargo, no ha sido posible puesto que únicamente he podido dar clase en 1º de la ESO.

- Alemaný, N. (2015). *La estadística de 1º de Bachillerato a través de proyectos y el software R* [Trabajo fin de máster].
- Armero, C. (2019). *Algoritmos matemáticos en R: aplicaciones a la docencia y al cálculo científico* [Trabajo fin de Grado].
- Batanero, C., Gea, M. M., López-Martín, M. M., y Arteaga, P. (2017). Análisis de los conceptos asociados a la correlación y regresión en los textos de bachillerato. *Didacticae: Revista de Investigación en Didácticas Específicas*(1), 60–76.
- Briz, Á., y Serrano, Á. (2018). Aprendizaje de las matemáticas a través del lenguaje de programación R en Educación Secundaria. *Educación matemática*, 30(1), 133–162.
- Consejería de Educación y Deporte. (2021). *Orden de 15 de enero de 2021, por la que se desarrolla el currículo correspondiente a la etapa de Bachillerato en la comunidad autónoma de andalucía, se regulan determinados aspectos de la atención a la diversidad y se establece la ordenación de la evaluación del proceso de aprendizaje del alumnado*.
- Galindo, A. (2017). Didáctica con R. Menos cuentas y más pensamiento crítico. *Pensamiento Matemático*, 7(1), 53–73.
- Gea, M. M., Batanero, C., Arteaga, P., Cañadas, G. R., y Contreras, J. M. (2014). Análisis del lenguaje sobre la correlación y regresión en libros de texto de bachillerato. *Suma*, 76, 37–45.
- Gea, M. M., Batanero, C., López-Martín, M. M., y Contreras, J. M. (2015a). Los recursos tecnológicos en la estadística bidimensional en los textos españoles de bachillerato. *Tecné, Episteme y Didaxis: TED*(38).
- Gea, M. M., López-Martín, M. M., y Roa, R. (2015b). Conflictos semióticos sobre la correlación y regresión en los libros de texto de bachillerato. *Avances de Investigación en Educación Matemática*(8).
- Monroy, O., y Rodríguez, L. (1998). Diseño, desarrollo y evaluación de situaciones problemáticas de estadística. *Una empresa docente*.

Revuelta, A. (2017). *R como herramienta de apoyo a la enseñanza de Estadística en Bachillerato* [Trabajo fin de Grado].

Rico, L., y Moreno, A. (2016). *Elementos de didáctica de la matemática para el profesor de secundaria*. Pirámide.

# **Anexos**

# Práctica 1: Nociones básicas de RStudio

## Cargar los datos

Una de las primeras cosas que debemos aprender es a crear vectores con los datos que vayamos a utilizar. Veamos un ejemplo:

```
Datos <- c(1,2,3,4,5,6,7,8,9,10)
```

La flecha indica que los datos se guardarán en el vector de nombre “Datos”, para crear vectores tendremos que escribir `c()` y dentro del paréntesis escribir los datos separados por comas.

Para ver los datos que hemos guardado basta con escribir el nombre del vector como sigue:

```
Datos
```

```
## [1] 1 2 3 4 5 6 7 8 9 10
```

Supongamos ahora que tenemos que vamos a trabajar con datos bidimensionales, al igual que estamos haciendo en clase. Para ello tendremos que crear dos vectores, uno con la primera coordenada de los datos y otro con la segunda.

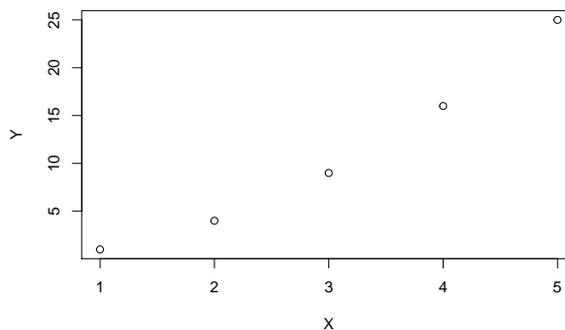
Para verlo más claro veamos un ejemplo de como guardaríamos los datos  $(1,1),(2,4),(3,9),(4,16),(5,25)$ :

```
X= c(1,2,3,4,5)  
Y= c(1,4,9,16,25)
```

## Graficar los datos

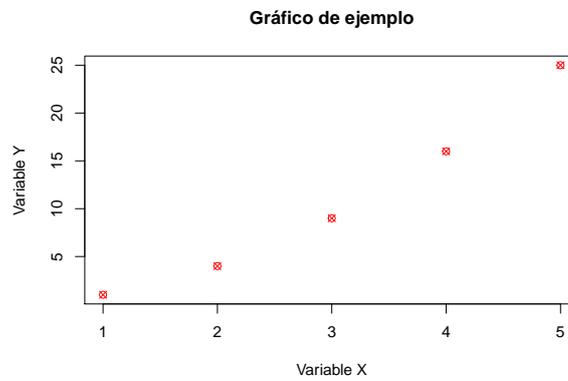
Una vez guardados podremos usarlos para distintos propósitos, por ejemplo graficarlos. Para ello usaremos la función `plot`:

```
plot(X,Y)
```



Esta función tiene numerosas opciones, entre las que se encuentra cambiar el color y el tipo de puntos, incluir una leyenda, cambiar el título de los ejes... Para aprender como utilizar todas estas opciones basta con buscar “plot” en la pestaña de ayuda o bien escribiendo en la terminal “`? plot`”, ahí encontrareis con detalle la sintaxis de la función. Veamos un ejemplo de lo que podemos llegar a hacer con la función `plot`:

```
plot(X,Y,col="red",pch=13,main="Gráfico de ejemplo",xlab="Variable X", ylab="Variable Y")
```



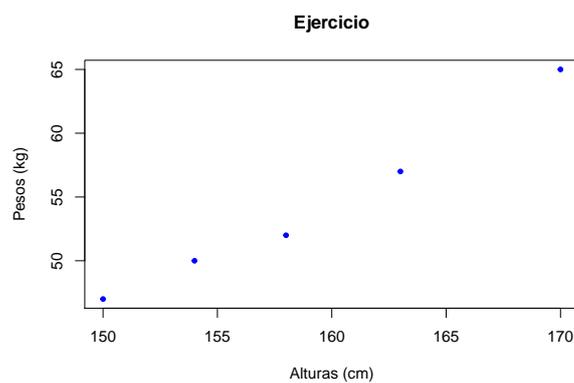
## Ejercicio práctico

Crea dos vectores con las siguientes pesos y alturas y gráficalos:

Pesos (kg)	47	50	52	57	65
Alturas (cm)	150	154	158	163	170

## Solución

```
Pesos <- c(47,50,52,57,65)
Alturas <- c(150,154,158,163,170)
plot(Alturas, Pesos, col="blue", pch=20, main="Ejercicio",xlab="Alturas (cm)",
     ylab="Pesos (kg)")
```



## Práctica 2: Estadística descriptiva con RStudio

### Parámetros estadísticos

RStudio es uno de los principales software estadísticos en la actualidad. Uno de sus puntos fuertes es que gracias a él podemos calcular fácilmente múltiples parámetros estadísticos como la media, varianza, desviación típica... En primer lugar cargaremos los datos del ejercicio de la práctica anterior, los cuales usaremos para aprender a calcular dichos parámetros:

```
Pesos <- c(47,50,52,57,65)
Alturas <- c(150,154,158,163,170)
```

Veamos ahora como se calculan media, varianza y desviación típica de un conjunto de datos (haremos el ejemplo con los pesos):

```
mean(Pesos) # Calcula la media
```

```
## [1] 54.2
```

```
var(Pesos) # Calcula la varianza
```

```
## [1] 49.7
```

```
sd(Pesos) # Calcula la desviación típica
```

```
## [1] 7.049823
```

Además de estos parámetros podemos obtener el mínimo, el máximo, media y cuartiles de los datos haciendo uso de la función summary:

```
summary(Pesos)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      47.0   50.0   52.0   54.2   57.0   65.0
```

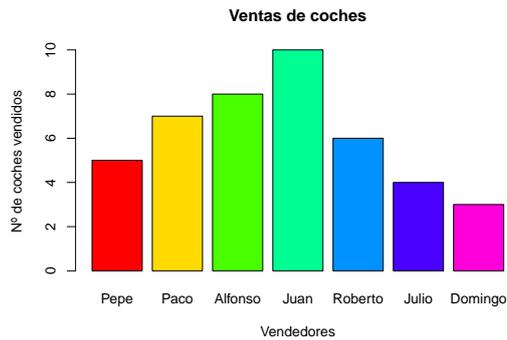
### Gráficos estadísticos

Otro de los principales puntos fuertes de RStudio es el apartado gráfico, gracias a él podremos crear de forma muy sencilla gráficos estadísticos tales como diagramas de barras, gráficos de sectores, gráficos de caja y bigotes...

Veamos una serie de ejemplos:

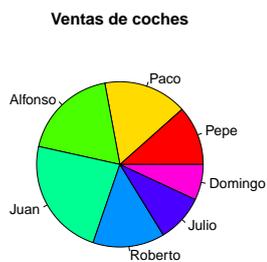
#### Gráfico de barras

```
Datos <- c(5,7,8,10,6,4,3)
Nombres <- c("Pepe","Paco","Alfonso","Juan","Roberto","Julio","Domingo")
barplot(Datos, names.arg = Nombres,main="Ventas de coches",xlab = "Vendedores",
        ylab= "Nº de coches vendidos",col=rainbow(7))
```



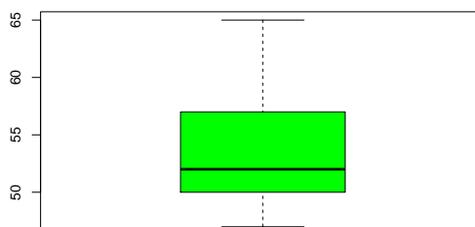
## Gráfico de sectores

```
Datos <- c(5,7,8,10,6,4,3)
Nombres <- c("Pepe","Paco","Alfonso","Juan","Roberto","Julio","Domingo")
pie(Datos, labels=Nombres,main="Ventas de coches",col=rainbow(7))
```



## Caja y bigotes

```
Pesos <- c(47,50,52,57,65)
boxplot(Pesos,col="green")
```



Todos estos gráficos tienen muchas más opciones, las cuales pueden ser consultadas haciendo uso de la pestaña de ayuda de Rstudio.

## Práctica 3: Regresión y correlación en R

En esta práctica aprenderemos a calcular la recta de regresión y los coeficientes de determinación de una distribución bidimensional.

Para ejemplificar el proceso tomaremos los siguientes datos de ejemplo:

```
X <- c(1,2,3,5,7,9,10)
Y <- c(0,3,5,6,8,9,14)
```

A continuación calcularemos la recta de regresión de Y sobre X (para realizar la de X sobre Y basta con intercambiar los papeles de ambas variables en los sucesivos pasos). Para ello haremos uso de la función lm :

```
regresion= lm (formula = Y~X)
```

Mostraremos en pantalla ahora los datos de dicha recta:

```
summary(regresion)
```

```
##
## Call:
## lm(formula = Y ~ X)
##
## Residuals:
##      1      2      3      4      5      6      7
## -1.16732  0.60506  1.37743 -0.07782 -0.53307 -1.98833  1.78405
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.06031     1.07557  -0.056  0.957455
## X           1.22763     0.17351   7.075  0.000873 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.487 on 5 degrees of freedom
## Multiple R-squared:  0.9092, Adjusted R-squared:  0.891
## F-statistic: 50.06 on 1 and 5 DF,  p-value: 0.0008726
```

Los datos que nos interesan de los anteriores son los que se encuentran en la sección Coefficients y el coeficiente de determinación ( Multiple R-squared ), en particular:

(Intercept)	-0.06031
X	1.22763

y

Multiple R-squared :	0.9092
----------------------	--------

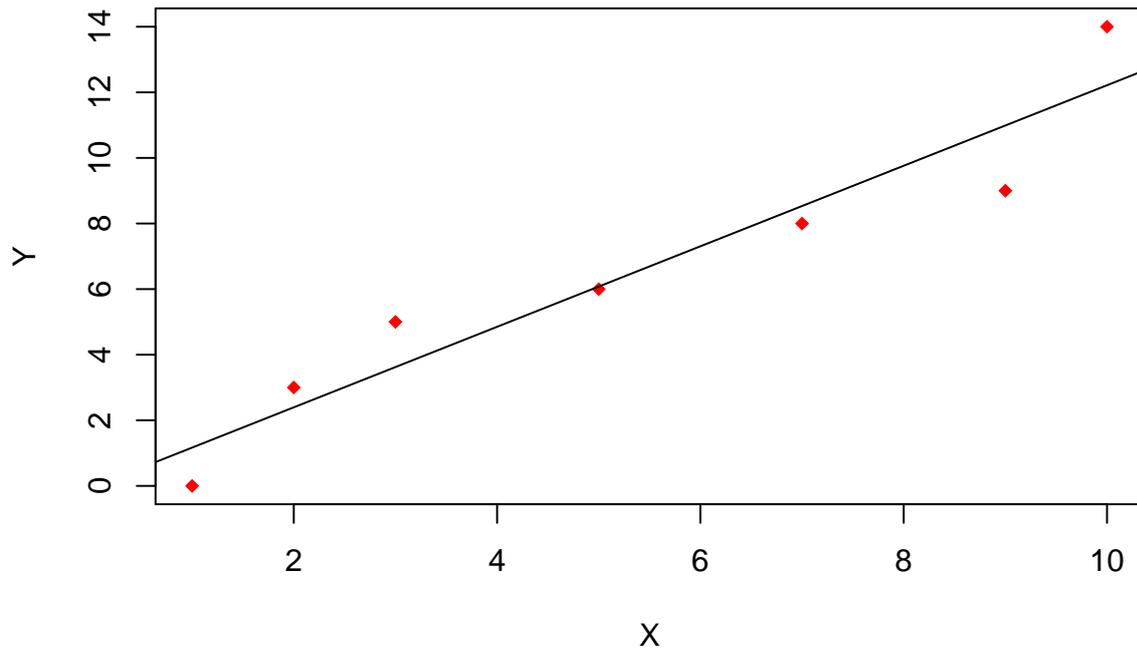
Estos datos nos dicen que la recta de regresión de Y sobre X es:

$$y = -0.06031 + 1.22763x$$

y que el coeficiente de determinación de la recta es 0.9092.

Finalmente pintaremos nuestros datos y la recta de regresión:

```
plot(X,Y,pch=18, col="red")  
abline(regresion)
```



Si quisieramos calcular el coeficiente de correlación una posible forma de hacerlo sería calculando la raíz cuadrada al valor del coeficiente de determinación y poniendo el resultado con signo positivo o negativo dependiendo de si la correlación es positiva o negativa (se ve claramente en el dibujo). En este caso:

```
CoefCorrelacion = sqrt(0.9092)  
CoefCorrelacion
```

```
## [1] 0.9535198
```

con signo positivo ya que la correlación es positiva.

## Tarea 7

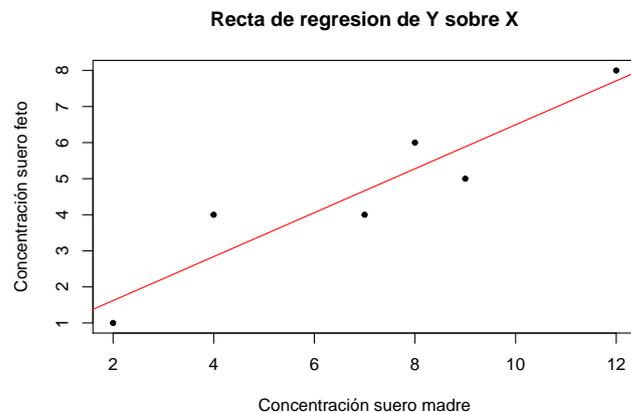
Cargamos, en primer lugar, los datos del problema

```
X<- c (8,4,12,2,7,9)
Y<- c (6,4,8,1,4,5)
```

Calculamos a continuación la recta de regresión lineal y la dibujamos:

```
regresion<- lm(formula = Y~X )
summary(regresion)
```

```
##
## Call:
## lm(formula = Y ~ X)
##
## Residuals:
##      1      2      3      4      5      6
## 0.7240 1.1615 0.2865 -0.6198 -0.6667 -0.8854
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.4010     0.9119   0.440  0.68280
## X              0.6094     0.1181   5.162  0.00669 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9444 on 4 degrees of freedom
## Multiple R-squared:  0.8695, Adjusted R-squared:  0.8368
## F-statistic: 26.65 on 1 and 4 DF,  p-value: 0.006689
plot(X,Y,pch=20,main="Recta de regresion de Y sobre X",xlab="Concentración suero madre",
      ylab="Concentración suero feto")
abline(regresion,col="red")
```



En consecuencia la recta de regresión es:

$$y = 0.4010 + 0.6094x,$$

y su coeficiente de determinación es 0.8368.

Puesto que el coeficiente de determinación tiene un valor cercano a 1, concluimos que el ajuste tiene buena bondad.

Hagamos ahora la estimación sustancia A en suero fetal sabiendo que se obtuvieron 10 microgramos por centímetro cúbico en el suero materno:

```
Estimacion= regresion$coefficients[1]+regresion$coefficients[2]*10  
Estimacion
```

```
## (Intercept)  
##      6.494792
```

En consecuencia, estimamos que habrá 6.494792 microgramos por centímetro cúbico de sustancia A en el suero fetal.

## Tarea 8: ¿Cómo varía la media y la desviación típica si modificamos los datos de partida?

### Objetivo de la tarea

En la siguiente tarea comprobaremos computacionalmente que:

- Si la media de unos datos,  $X$ , es  $\mu$ , entonces la media de  $X + b$  es  $\mu + b$ .
- Si la varianza de unos datos,  $X$ , es  $\sigma^2$ , entonces la varianza de  $X + b$  sigue siendo  $\sigma^2$ .
- Si la media de unos datos,  $X$ , es  $\mu$ , entonces la media de  $a * X$  es  $a * \mu$ .
- Si la varianza de unos datos,  $X$ , es  $\sigma^2$ , entonces la varianza de  $a * X$  es  $a^2 * \sigma^2$ .

### Datos tomados (modificables)

Para hacer los cálculos tomaremos los siguientes datos:

```
X <- c(2,5,3,5,6,3,6,4,7,23,12) #podemos variar los datos
a= 10 # parámetro que podemos variar
b= 3 # parámetros que podemos variar
```

Los datos anteriores pueden modificarse a placer, al ejecutar el programa se cambio todo automáticamente.

### Media y varianza de nuestros datos de partida ( $X$ ):

En primer lugar calculemos la media y la varianza de  $X$ :

```
Media=mean(X)
Media

## [1] 6.909091
Varianza= var(X)
Varianza

## [1] 35.69091
```

En consecuencia tenemos que la media de  $X$  es 6.9090909 y la varianza 35.6909091

### Media y varianza de $X + b$ :

Veamos que ocurre con la media de  $X + b$ :

```
Media_b=mean(X+b)
Media_b

## [1] 9.909091
```

Se cumple lo que queríamos comprobar, ya que  $\mu + b = 6.9090909 + 3 = 9.9090909$ .

Veamos que ocurre con la varianza de  $X + b$ :

```
Varianza_b= var(X+b)
Varianza_b
```

```
## [1] 35.69091
```

La varianza, como se puede observar, es la misma  $35.6909091 = 35.6909091$

## Media y varianza de $a * X$ :

Veamos que ocurre con la media de  $a * X$ :

```
Media_a=mean(a*X)
Media_a
```

```
## [1] 69.09091
```

Se cumple lo que queríamos comprobar, ya que  $a * \mu = 10 * 6.9090909 = 69.0909091$ .

Veamos que ocurre con la varianza de  $a * X$ :

```
Varianza_a= var(a*X)
Varianza_a
```

```
## [1] 3569.091
```

Se cumple lo que queríamos comprobar, ya que  $a^2 * \sigma^2 = (10)^2 * 35.6909091 = 3569.0909091$ .

## Práctica 4 : Importación de grandes cantidades de datos con Studio

Hasta la fecha hemos trabajado con datos de poca envergadura ya que teníamos que ingresarlos de forma manual. En esta práctica aprenderemos a importar datos desde un archivo .csv, los cuales son archivos de texto en los que los datos están separados por comas (CSV = “Comma Separated Values”).

Una de las ventajas que tiene RStudio es que podemos, incluso, obtener los datos desde un link sin necesidad de descargarlo, sin embargo trabajaremos con un archivo previamente descargado para no necesitar conexión a internet. Tomaremos un archivo obtenido de la página <https://ourworldindata.org/> el cual recoge las poblaciones de los distintos países desde el año 1800.

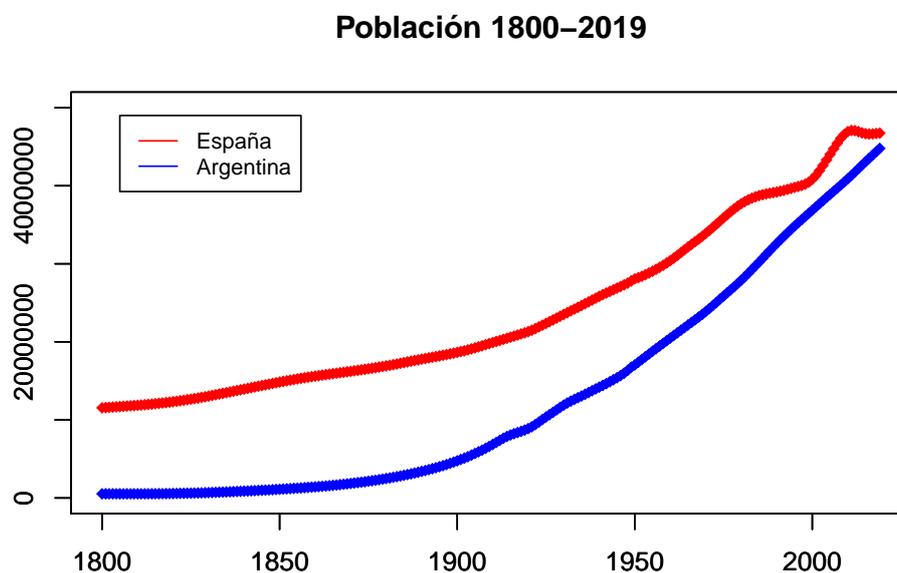
Para leer el archivo .csv haremos uso de la función `read.csv`, si los datos están separados por “punto y coma” en vez de comas será necesario usar la función `read.csv2`.

```
Datos <- read.csv(file="Poblacion.csv")
```

Tras ello tomaremos los datos de España y de Argentina haciendo uso de la función `subset`:

```
Spain <- subset (Datos, Entity== "Spain")  
Argentina <- subset (Datos, Entity== "Argentina")
```

Finalmente, haciendo uso de los conocimientos vistos en prácticas anteriores podremos graficar la población de España (en rojo) y de Argentina (en azul) a lo largo de los años:



Esto es simplemente un ejemplo de lo que se puede hacer, una vez que tenemos los datos importados podemos realizar con ellos todo lo que hemos aprendido en las prácticas anteriores.

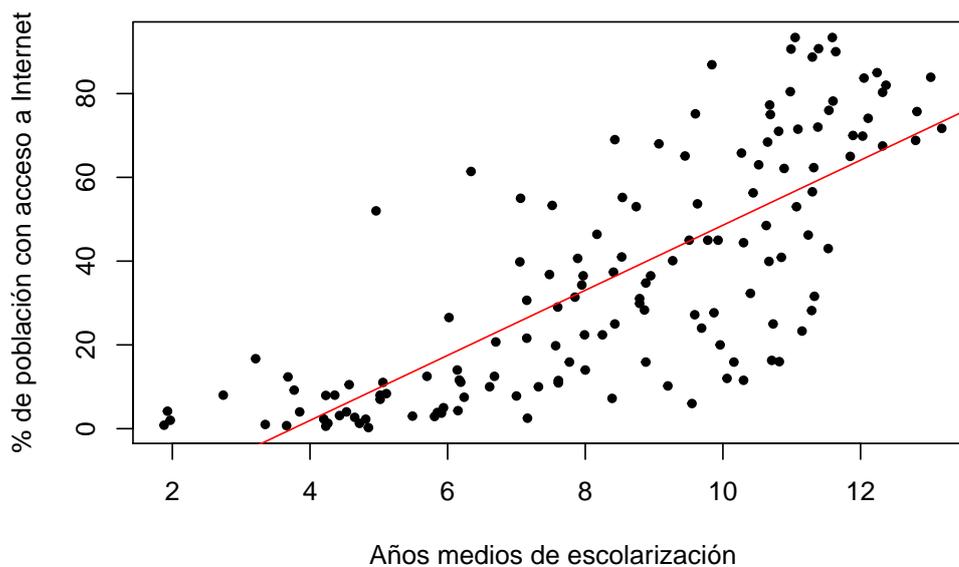
## Tarea 10: Relación entre Internet y la escolarización

En la siguiente tarea le daremos al alumno un archivo .csv obtenido de <https://ourworldindata.org/grapher/correlation-between-internet-usage-and-mean-years-of-schooling> con los datos de diversos países acerca del porcentaje de población con acceso a internet y los años medios de escolarización. Tras ésto le diremos al alumno que responda a las siguientes preguntas:

- Calcula la recta de regresión de los datos correspondientes a 2010.
- Indica el coeficiente de determinación y de correlación de la recta.
- ¿Qué conclusiones extraes? ¿Como consecuencia de un alto uso de Internet aumenta la edad media de escolarización?

### Solución de la tarea:

```
Datos <- read.csv("Internet.csv")
Datos2010<- subset(Datos, Year=="2010")
Cole <- Datos2010$Educational.Attainment..average.years.of.total.education...Barro.Lee.Education.Database
Internet<-Datos2010$Percentage.of.Individuals.using.the.Internet..ICT...2015.
regresion <- lm(Internet~Cole)
plot(Cole, Internet,xlab="Años medios de escolarización",
      ylab="% de población con acceso a Internet", pch=20)
abline(regresion,col="red")
```



```
summary(regresion)
```

```
##
## Call:
## lm(formula = Internet ~ Cole)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -39.352 -11.891  -1.544  13.082  42.581
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -29.1118     4.7882  -6.08 1.05e-08 ***
## Cole         7.7684     0.5419  14.33 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 18.16 on 142 degrees of freedom
## (98 observations deleted due to missingness)
## Multiple R-squared:  0.5913, Adjusted R-squared:  0.5885
## F-statistic: 205.5 on 1 and 142 DF,  p-value: < 2.2e-16
```

En consecuencia tenemos que la recta de regresión es:

$$y = -29.1118 + 7.7684x$$

donde  $y$  es el porcentaje de la población con acceso a internet y  $x$  los años medios de escolarización.

El coeficiente de determinación es 0.59013 y el coeficiente de regresión es  $\sqrt{0.59013} = 0.7681992$  (tomamos el valor positivo por ser la pendiente de la recta positiva)

Puesto que el coeficiente de correlación, en valor absoluto, es próximo a 1 tenemos que hay alta correlación entre ambas variables. Sin embargo, no podemos afirmar que un alto uso de Internet aumente la edad media de escolarización (correlación  $\neq$  causalidad). Es probable que haya otra variable (la riqueza del país) que cause tanto a  $X$  como a  $Y$ .