Estadística Bidimensional de Matemáticas Aplicadas a las Ciencias Sociales I de 1ºBachillerato con el programa R

Máster Universitario en Formación del Profesorado de Enseñanza Secundaria Obligatoria, Bachillerato, Formación Profesional y Enseñanza de Idiomas.

Presentado por:
D. CARLOS ISMAEL SUÁREZ DÍAZ

Dirigido por:
D. JUAN GERARDO ALCÁZAR ARRIBAS

Alcalá de Henares, a 26 de agosto de 2023

Resumen

En este Trabajo de Fin de Máster se realiza una introducción al uso del programa R y se explican los conceptos esenciales de la estadística bidimensional de Primero de Bachillerato de Ciencias Sociales con el objetivo de realizar una propuesta didáctica sobre estos contenidos. En primer lugar, se presenta el programa R, explicando los conceptos básicos necesarios para comenzar a usarlo, introduciendo, además, ejemplos sencillos que los alumnos pueden practicar para familiarizarse con este entorno. A continuación, se explican los conceptos necesarios de la estadística bidimensional, como el de diagrama de dispersión, correlación lineal y de recta de regresión, con ejemplos de cada uno de ellos. También se da una pequeña introducción a los conceptos de predicción, causalidad y coeficiente de determinación y se detallan las tablas de doble entrada. Finalmente, se detalla la propuesta didáctica previamente mencionada, desarrollando cada una de las sesiones.

Palabras clave

Bidimensional, correlación lineal, diagrama de dispersión, recta de regresión, R, RStudio, tablas doble entrada, excel.

Abstract

In this Final Master Project an introduction to the use of the R program is made and the essential concepts of two-dimensional statistics of the First year of the Social Sciences Baccalaureate are explained with the aim of making a didactic proposal on these contents. First, the R program is presented, explaining the basic concepts necessary to start using it, also introducing simple examples that students can practice to become familiar with this environment. Next, the necessary concepts of two-dimensional statistics are explained, such as the scatterplot, linear correlation, and regression line, with examples of each of them. A short introduction to the concepts of prediction, causality and coefficient of determination is also given, and double entry tables are detailed. Finally, the previously mentioned didactic proposal is detailed, developing each of the sessions.

Keywords

Two-dimensional, linear correlation, scatterplot, regression line, R, RStudio, double-entry tables, excel.

Índice general

Anexos 53

1. Introducción

El presente Trabajo de Fin de Máster persigue trabajar los contenidos de variables bidimensionales de Estadística del curso de primero de bachillerato de Ciencias Sociales, y hacer ejercicios con el programa R.

R es un lenguaje de programación y un entorno diseñado principalmente para el análisis estadístico y la visualización de datos. Es una herramienta poderosa utilizada por científicos de datos, estadísticos, analistas y profesionales en diversas disciplinas para manipular, explorar y modelar datos de manera eficiente.

Una motivación para enseñar al alumnado estos contenidos con el software R es que pueden tener una aproximación de la aplicación real de los contenidos que se ven en clase, además de tener el incentivo de sentir que están haciendo algo más grande, más allá de la asignatura. Steve Jobs en el vídeo "Todos deberían aprender a programar" del canal Cómo Programar (2013) dice: "Todos en este país deberíamos aprender cómo programar una computadora, porque te enseña cómo pensar". Esta frase puede animar al alumnado a aprender con ganas estos contenidos, si sienten que puede servirles para la asignatura de matemáticas y también para otros ámbitos de su vida. Zuckerberg, también en el mismo vídeo, describe a la profesión como "la única que, sentado en una silla, puedes crear algo totalmente nuevo desde cero". Los alumnos que sean más creativos pueden encontrar inspiración en esta frase.

Para realizar esta propuesta, he cogido como referencia la PGA del IES Alonso Quijano (2022), instituto en el que realicé las prácticas del máster. Además, encontré inspiración en el tipo de alumnado que conocí durante las mismas, ya que observé que no habían conocido herramientas de software de matemáticas más allá de GeoGebra, un software que es muy limitado fuera del mundo académico.

En el Capítulo 2, usando como referencia la página web de Coll y Pérez (2016-2017), se explica el lenguaje de programación R, los primeros pasos para que los alumnos puedan ir familiarizándose con el entorno de trabajo y se repasan algunos conceptos básicos del programa.

En el Capítulo 3, usando el libro de Marea Verde de autor Clausell (2022) y el libro de Anaya de autores Colera Jiménez, Oliveira Gonzáles, Colera Cañas, García Pérez, y Alcardo B. (2022), ambos de Matemáticas Aplicadas a las Ciencias Sociales I, además de algunos ejercicios de la página web de Outón Ruiz (s.f.), se describen los conceptos de diagrama de dispersión, correlación lineal, recta de regresión con un ejemplo de R y las tablas de doble entrada con ejemplos en Excel. Las actividades propuestas que competen a este capítulo se presentan en el Anexo I.

En la última sección del Capítulo 3, trabajaremos con el software Excel en lugar de con R. Esto se debe a la dificultad que hay al trabajar con tablas de doble entrada en R para un alumnado recién iniciado en este programa. Excel es una herramienta más útil e intuitiva para que los estudiantes organicen y analicen datos. En el Anexo IV explicamos los comandos básicos de Excel que usamos en la sección de tablas de doble entrada.

En el Capítulo 4, se presenta la propuesta didáctica enmarcándola en el RD 234/2022, 5 de Abril, de Boletín Oficial del Estado (2022), con su correspondiente temporalización, materiales y una guía y descripción de la propuesta didáctica, terminando con los procedimientos de evaluación.

Por último, en las conclusiones, se repasan las ideas principales de lo expuesto en el TFM y se realiza un análisis DAFO para evaluar nuestra propuesta didáctica.

2. El Programa R

2.1. El Lenguaje de Programación R

R es un lenguaje de programación y un entorno para el análisis estadístico y gráfico. Es parte del sistema GNU y se distribuye bajo la licencia GNU GPL, es decir, es software libre y gratuito. Está disponible para Windows, Macintosh y GNU/Linux.

R fue inicialmente creado por R. Ihaka y R.Gentleman de la Universidad de Auckland en 1993 pero, actualmente, el entorno R es el resultado de la colaboración de toda una comunidad de usuarios. A partir de 1997, el desarrollo del código fuente (o base-R) de R es llevado por un grupo de programadores conocido como “The R-core team”. La página web oficial es https://www.r-project.org/, donde podremos encontrar toda la información oficial acerca de R.

R es cada vez más usado, no sólo en la universidad y la docencia, sino también en el mundo de la empresa. Entre las empresas que más lo usan están Google, Facebook, Twitter, Microsoft, IBM, Uber, Ford, Airbnb, American Express.

2.1.1 El Programa RStudio

Nosotros trabajaremos con RStudio, dado que la interfaz de usuario de R no es muy amigable ni versátil. RStudio es un programa que nos permitirá interactuar con R de forma más accesible, además de facilitar muchas de las tareas de programación y análisis de datos en R.

2.1.2 Instalación de R y RStudio

Para la realización de este trabajo, utilizaremos la versión de RStudio 2023.06.1-524 y la versión R-4.3.1, ambas para el sistema operativo Windows.

- Podemos instalar R en su página oficial: https://cran.r-project.org/.
- Podemos descargar RStudio Desktop desde el siguiente enlace: https://posit.co/download/rstudio-desktop/.

2.2. Primeros Pasos con RStudio

Uno de los objetivos de este trabajo es familiarizarse con el entrono de trabajo que proporciona R y RStudio.

2.2.1 Entorno de Trabajo

Si abrimos RStudio, encontraremos una disposición similar a la que se muestra en la Figura 1.

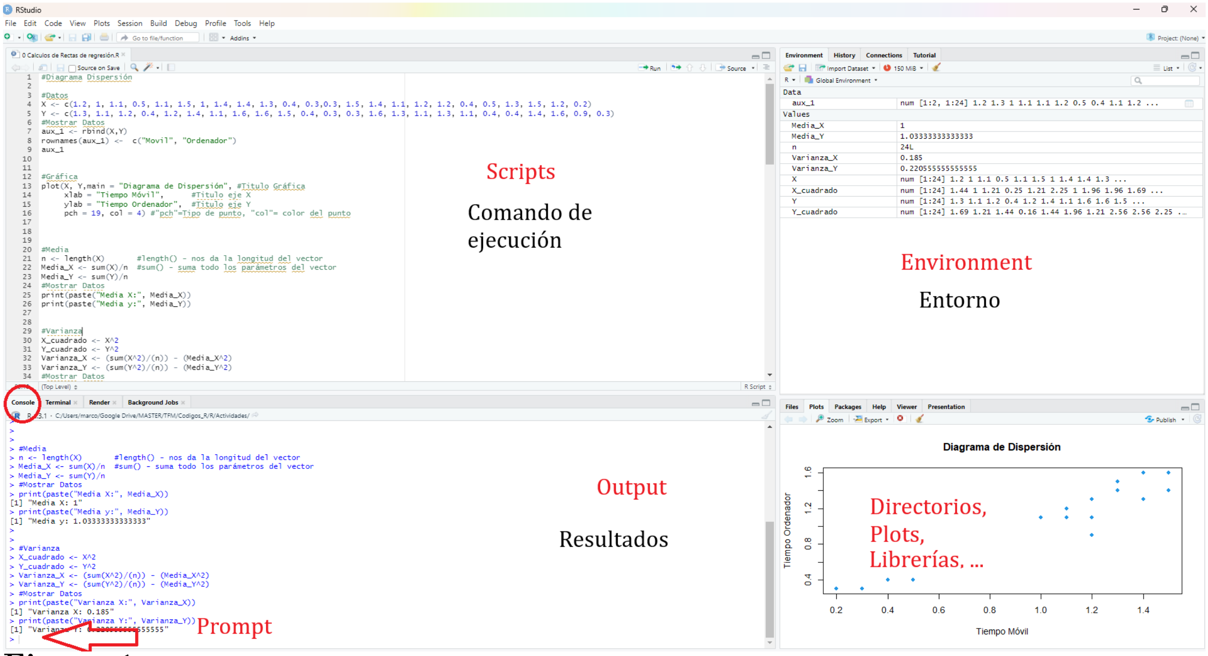


Figura 1

Aspecto general de RStudio.

Una vez abierto RStudio, podemos escribir y ejecutar las órdenes de varias formas:

- Directamente en la consola.
- A través de un script, que tienen como extensión (.R).
- Con ficheros Rmarkdown, que tiene como extensión (.Rmd).

Como se puede observar en la Figura 1, RStudio está dividido en 4 paneles:

- **Panel de la consola**, en la posición inferior izquierda.
- **Panel de comando de ejecución**, en la posición superior izquierda.
- **Panel de entorno**, en la posición superior derecha.
- **Panel de archivos, gráficos, paquetes y ayuda**, en la posición inferior derecha.

Todo esto lo podemos encontrar con más detalle en el Anexo VIII.

2.2.2. Panel de la Consola

El panel de la Consola en RStudio es una interfaz que nos permite interactuar con R mediante comandos en tiempo real. En este panel podemos escribir y ejecutar comandos de R y ver los resultados directamente. Es una parte esencial de RStudio que nos permite probar rápidamente fragmentos de código, experimentar con funciones y realizar análisis de manera interactiva.

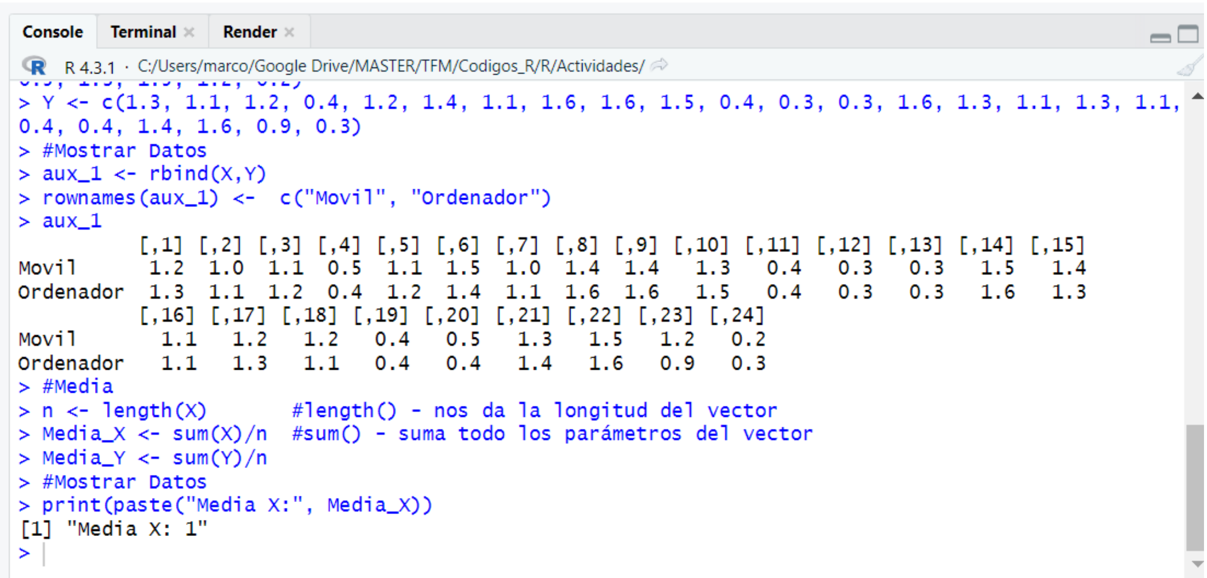


Figura 2

Aspecto del panel de la consola en RStudio.

Sin embargo, el panel de la consola tiene ciertas limitaciones:

- **Interactividad limitada:** Aunque podemos ejecutar comandos uno a uno, no es el entorno más adecuado para desarrollar y mantener código más largo y complejo. La interacción constante con el panel puede ser incómoda para tareas extensas.

- **Historial limitado:** Aunque el historial de comandos nos permite acceder a comandos previamente ejecutados, su capacidad puede ser limitada y no es la mejor opción para mantener un historial de trabajo detallado.

- **Dependencia de la sesión**: Los objetos y variables que creamos en la sesión de la consola no se mantienen cuando cerramos RStudio. Esto puede ser inconveniente si necesitamos mantener el trabajo o los resultados a lo largo del tiempo.

2.2.3. Panel de Comando de Ejecución

En RStudio, el panel de comando de ejecución o “Source” es una parte fundamental de la interfaz que nos permite trabajar con archivos de script en lenguaje R (.R) y archivos Rmarkdown (.Rmd).

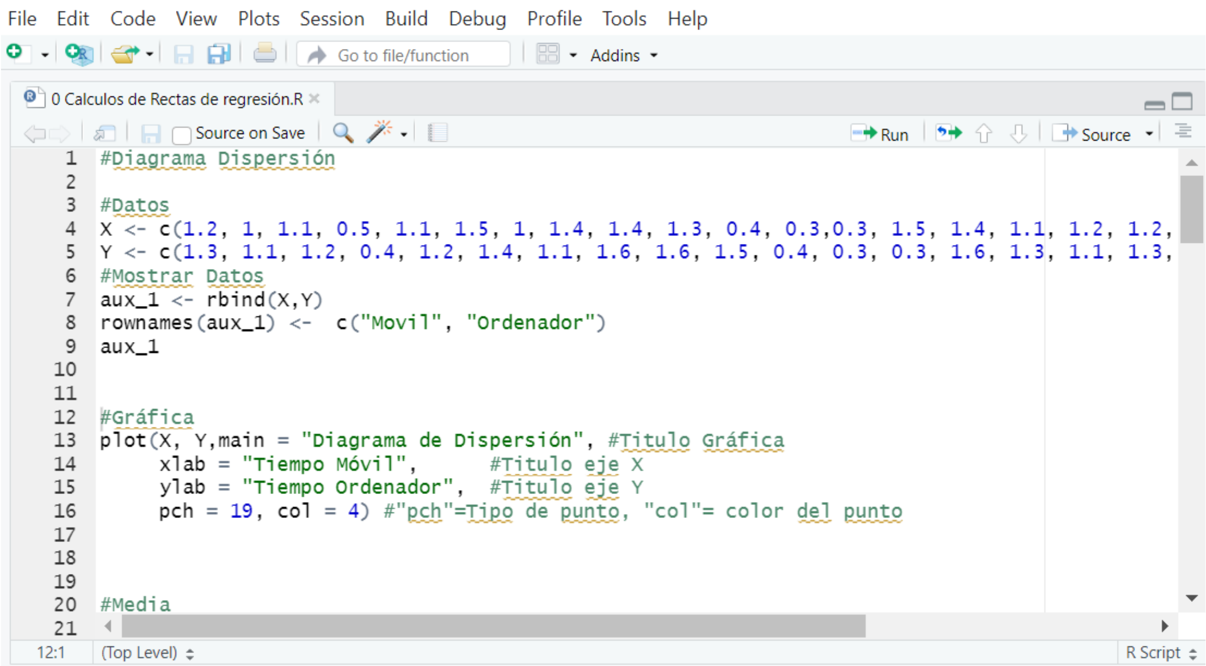


Figura 3

Aspecto del panel de comando de ejecución en RStudio.

Tenemos los siguientes comandos:

- **Nuevo Script / Nuevo Documento Rmarkdown**: Podemos crear un nuevo archivo de script R o un nuevo documento Rmarkdown haciendo clic en los botones correspondientes en la esquina superior izquierda del panel.

- **Abrir Script / Abrir Documento:** Podemos abrir archivos existentes haciendo clic en “File” y luego seleccionando “Open File...” en el menú de la parte superior.

- **Edición:** Una vez que tenemos un archivo abierto en el panel de comando de ejecución, podemos escribir y editar código en él.

- **Ejecución:** Podemos seleccionar una parte del código o todo el archivo y hacer clic en el botón “Run” (Ejecutar), en la esquina superior derecha del panel, para ejecutar el código en la consola R.

- **Guardar**: Podemos guardar los cambios haciendo clic en el icono del disquete en la parte superior del panel.

Script

Un archivo de script (.R) contiene código R que podemos ejecutar directamente en la consola de R. Es útil para escribir y ejecutar comandos o funciones individuales, explorar datos y realizar análisis. Los script son principalmente para tareas de programación y análisis.

Rmarkdown

Un archivo Rmarkdown (.Rmd) combina texto enriquecido y código R en un solo archivo. Podemos mezclar explicaciones, visualizaciones, resultados de código y otros elementos en un formato que es fácil de compartir y presentar. Los documentos Rmarkdown nos permiten crear informes, presentaciones y documentos interactivos con análisis automatizados.

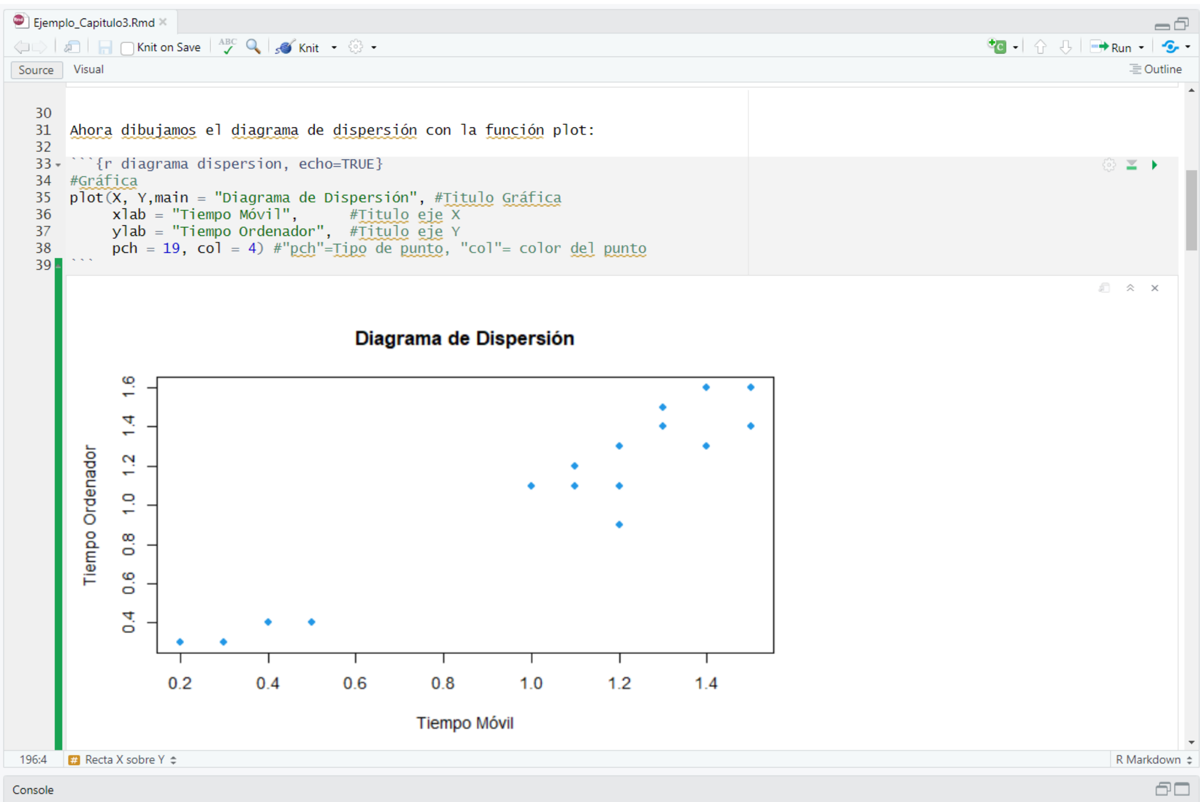


Figura 4

Aspecto del entorno de trabajo de Rmarkdown.

2.2.4. Panel de Entorno

En el panel de entorno o "Environment" se irán registrando los objetos que vayamos creando en la sesión de trabajo. También tenemos la opción de cargar y guardar una sesión de trabajo, importar datos y limpiar los objetos de la sesión. Estas opciones están accesibles a través de la cinta de opciones de la pestaña.

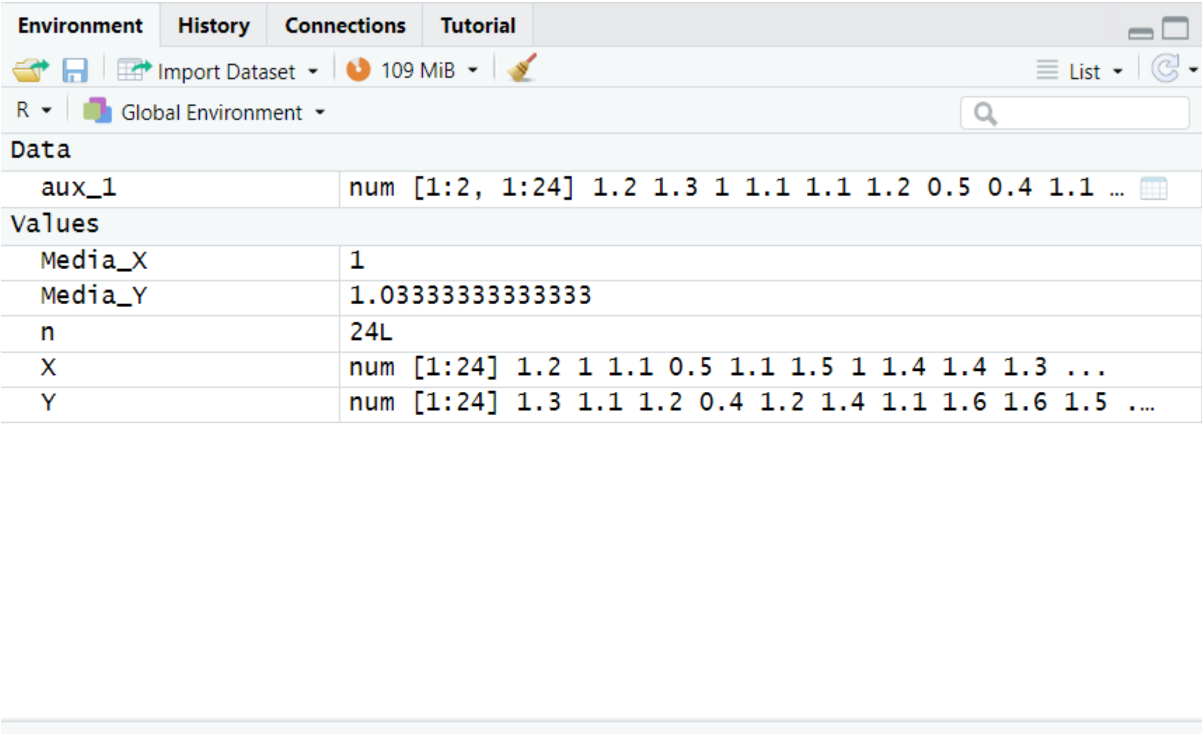


Figura 5

Aspecto del panel de entorno en RStudio.

2.2.5. Panel de Archivos, Gráficos, Paquetes y Ayuda

En este panel cabe destacar las siguientes pestañas, cada una con diferentes opciones:

- **Files**: es una especie de explotador de ficheros.
- **Plots**: donde se visualizan los gráficos que creamos. Entre las opciones disponibles se encuentran:
  1. **Zoom**: para agrandar el gráfico y verlo en otra ventana.

2. **Export**: para exportar/guardar el gráfico. Se puede guardar el gráfico como imagen, pdf o copiarlo al portapapeles.

- **Packages**: proporciona un listado de los paquetes instalados en R y los que han sido cargados en la sesión. A través de las opciones de esta pestaña podemos instalar nuevos paquetes o actualizar los existentes.
- **Help**: Para obtener ayuda sobre una determinada función.

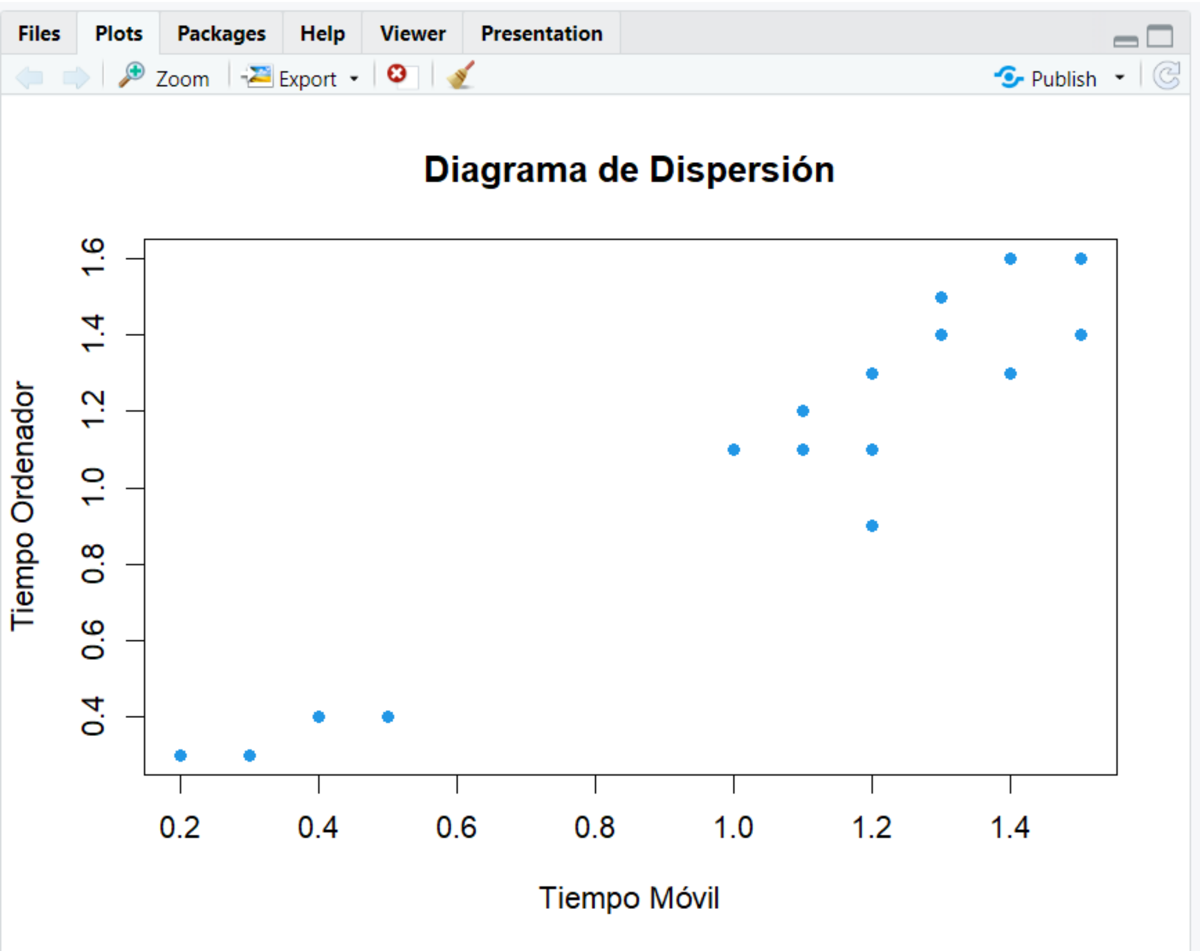


Figura 6

Aspecto del panel de archivos, gráficos, paquetes, ayuda, en RStudio.

2.3 Conceptos Básicos de R

Algunos conceptos fundamentales para empezar con R son:

- **Asignación de variables:** En R, puedes asignar valores a variables utilizando el operador $<-$ o $=$.

```
x <- 10
nombre <- "Juan"
```

- **Imprimir en la consola:** Para imprimir valores en la consola, simplemente hay que escribir el nombre de la variable o el valor. No es necesario utilizar una función de impresión.

```
x
nombre
```

```
[1] 10
[1] "Juan"
```

- **Operaciones aritméticas:** Puedes realizar operaciones matemáticas básicas como suma, resta, multiplicación y división.

```
suma <- 5 + 3
resta <- 10 - 2
multiplicacion <- 4 * 6
division <- 15 / 3
```

- **Vectores:** Los vectores son secuencias de valores del mismo tipo. Puedes crearlos con la función c() (concatenar).

```
numeros <- c(1, 2, 3, 4, 5)
letras <- c("a", "b", "c")
```

- **Funciones predefinidas:** R ofrece muchas funciones predefinidas para realizar diversas operaciones.

```
mean(numeros)   # Calcula la media del vector "numeros"
length(letras)  # Calcula la longitud del vector "letras"
```

Para nuestro trabajo se han creado varios archivos Rmarkdown con el que los alumnos podrán iniciarse en el mundo de la programación. Se ha elegido este formato, dado que, estos archivos son fácilmente reproducibles por el usuario. El primer .Rmd con el que trabajarán los estudiantes se llama Introducción a R y lo encontramos en el Anexo II.

3. Estadística Bidimensional en 1º Bachillerato y R

3.1. Introducción

En este capítulo se explicarán los conceptos de diagrama de dispersión, correlación lineal y de recta de regresión, apoyándonos en ejemplos realizados en R, con los cuales el alumnado podrán seguir la construcción de los diferentes elementos sin perderse gracias a los archivos interactivos creados con Rmarkdown.

Para terminar, también se verán las tablas de doble entrada, pero los ejemplos se realizarán en Excel, pudiendo usar otro software de tratamiento de datos.

3.2. Diagrama de Dispersión

En una población, el estudio simultáneo de dos variables unidimensionales nos puede ayudar a determinar la relación que existe entre ellas. Denominaremos variable X a la variable independiente y variable Y a la variable dependiente. Una **variable bidimensional** es una variable donde cada individuo está definido por el par (X, Y).

Los datos siguen una **distribución bidimensional** cuando cada uno de los individuos llevan asociados los valores de las variables X e Y, y las representamos como (x_i, y_i).

Cuando X e Y de una distribución bidimensional sean **variables cuantitativas** y no estén agrupadas, podemos representar los datos con un **diagrama de dispersión** o con una **nube de puntos**.

3.2.1. Ejemplo: Móvil y Ordenador con R

Se recogen los tiempos de descarga, en segundos, de 24 archivos en un móvil (X) y en un ordenador (Y).

```
          [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13]
Móvil      1.2  1.0  1.1  0.5  1.1  1.5  1.0  1.4  1.4   1.3   0.4   0.3   0.3
Ordenador  1.3  1.1  1.2  0.4  1.2  1.4  1.1  1.6  1.6   1.5   0.4   0.3   0.3

          [,14] [,15] [,16] [,17] [,18] [,19] [,20] [,21] [,22] [,23] [,24]
Móvil       1.5   1.4   1.1   1.2   1.2   0.4   0.5   1.3   1.5   1.2   0.2
```

```
Ordenador  1.6   1.3   1.1   1.3   1.1   0.4   0.4   1.4   1.6   0.9   0.3
```

Obtengamos el diagrama de dispersión a partir de los datos dados:

```r
#Datos
X <- c(1.2, 1, 1.1, 0.5, 1.1, 1.5, 1, 1.4, 1.4, 1.3, 0.4, 0.3,0.3, 1.5,
    1.4, 1.1, 1.2, 1.2, 0.4, 0.5, 1.3, 1.5, 1.2, 0.2)
Y <- c(1.3, 1.1, 1.2, 0.4, 1.2, 1.4, 1.1, 1.6, 1.6, 1.5, 0.4, 0.3, 0.3,
    1.6, 1.3, 1.1, 1.3, 1.1, 0.4, 0.4, 1.4, 1.6, 0.9, 0.3)
#Grafica
plot(X, Y,main = "Diagrama de Dispersion", #Titulo Grafica
     xlab = "Tiempo Movil",       #Titulo eje X
     ylab = "Tiempo Ordenador",   #Titulo eje Y
     pch = 19, col = 4) #"pch"=Tipo de punto, "col"= color del punto
```

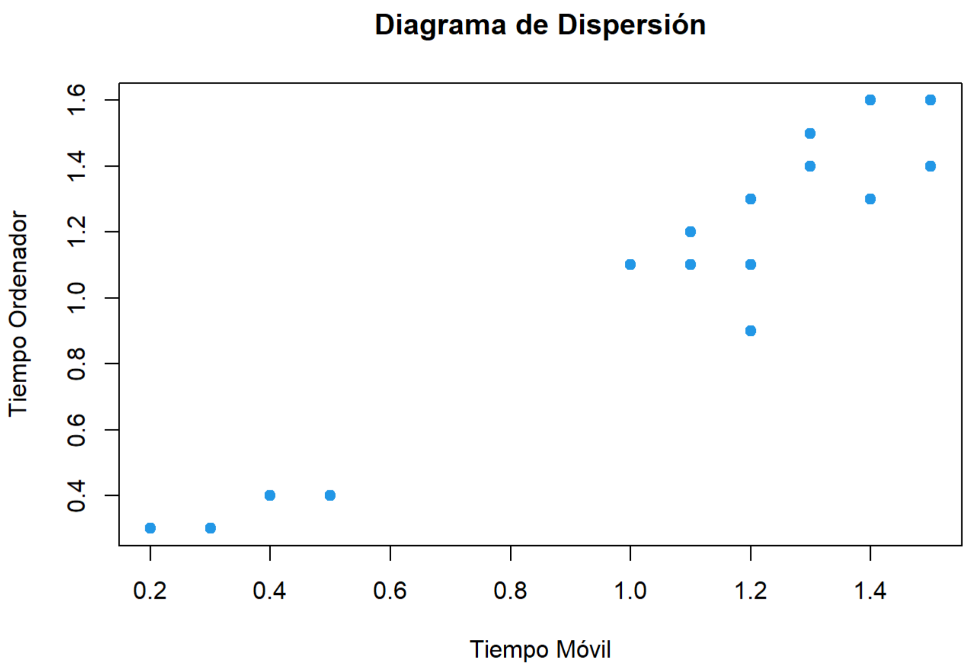


Figura 7

Diagrama de dispersión del ejemplo móvil y ordenador en R.

3.2.2. Idea de Correlación y Regresión

En ciertas ocasiones, cuando observamos un diagrama de dispersión, podemos apreciar una cierta relación entre las variables, a la que llamaremos **correlación**. A priori, para ver la existencia de correlación entre dos variables hemos de determinar si es posible la existencia de una función, la cual se aproxime razonablemente a la nube de puntos y observar qué tan bien se ajusta. Esta función recibe el nombre de **curva de regresión** y podrá ser una recta, una función polinómica, exponencial, entre otras.

Ejemplo: Calculamos la correlación de los datos de 50 vehículos que relacionan el peso en función de los caballos de potencia. Se identifica que los datos se aproximan a una función polinómica de grado 2, por lo que se ha ajustado una curva de regresión de este tipo. Podemos observar en la Figura 8 que los datos se ajustan adecuadamente a la nube de puntos, por lo que podemos pensar que existe correlación entre las variables.

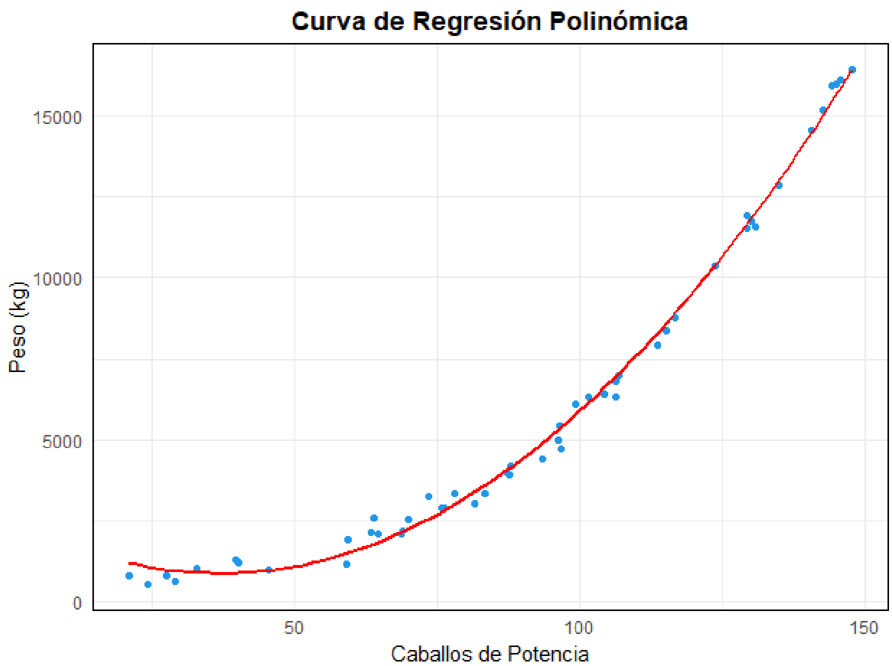


Figura 8

Curva de regresión con datos simulados para el peso del vehículo en función de los caballos de potencia ajustados por una función polinómica de grado 2.

En adelante, nos vamos a dedicar solamente a las curvas de regresión que sean **rectas de regresión**, es decir, intentaremos ajustar los datos a la recta que mejor se ajuste a la relación entre las dos variables. Además, si la relación de las variables es buena, es decir, que la nube de puntos se asemeja a una recta, podremos decir que existe **correlación lineal**. El reto que se nos presenta es medir el grado de correlación lineal y determinar la ecuación de la recta de regresión que mejor se ajuste.

3.3. Correlación Lineal

El grado de dependencia lineal entre dos variables se mide con el **coeficiente de correlación lineal**, y cuando la dependencia lineal es débil, la recta de regresión carece de interés. Antes de calcular este coeficiente de correlación, veamos algunos parámetros asociados a una distribución bidimensional que nos facilitarán el cálculo de este coeficiente.

3.3.1. Centro de Gravedad

Sea una distribución bidimensional, con una población de tamaño n y con los pares de valores $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$, podemos definir el **centro de gravedad** de la distribución como $(\overline{x}, \overline{y})$, donde $\overline{x}$ es la media de la variable X e $\overline{y}$ es la media de la variable Y:

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n} \qquad \overline{y} = \frac{\sum_{i=1}^{n} y_i}{n}$$

En este apartado recordamos el concepto de desviación típica, que es una medida estadística que cuantifica la dispersión o la variabilidad de un conjunto de datos con respecto a la media. La desviación típica de la variable X la denotamos como σ_x y la hallamos de la siguiente manera:

$$\sigma_x = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n}} = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n} - \overline{x}^2}$$

Siendo σ_x^2 la varianza de la variable X. De igual manera, podemos hallar σ_y:

$$\sigma_y = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \overline{y})^2}{n}} = \sqrt{\frac{\sum_{i=1}^{n} y_i^2}{n} - \overline{y}^2}$$

3.3.2. Covarianza

La **covarianza** es un parámetro estadístico que evalúa la relación entre dos variables de un conjunto de datos. Indica cómo cambian conjuntamente en relación con sus medias. En esencia, la covarianza cuantifica si las dos variables tienden a variar en la misma dirección, es decir, si cuando la variable X aumenta, la variable Y también, entonces nos encontramos con una covarianza positiva. Por otro lado, si las variables tienden a variar en direcciones opuestas, la covarianza será negativa. Si no hay una relación clara entre las variables, la covarianza estará cerca de ser nula. La covarianza la denotaremos como σ_{xy} y se calcula de la siguiente manera:

$$\sigma_{xy} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{n} = \frac{\sum_{i=1}^{n} x_i\, y_i}{n} - \overline{x}\,\overline{y}$$

Es interesante demostrar la igualdad de la fórmula anterior:

$$\sigma_{xy} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{n} = \frac{\sum_{i=1}^{n}(x_i\, y_i - x_i\,\overline{y} - \overline{x}\, y_i + \overline{x}\,\overline{y})}{n} =$$

$$\frac{\sum_{i=1}^{n} x_i\, y_i}{n} - \overline{y}\,\frac{\sum_{i=1}^{n} x_i}{n} - \overline{x}\,\frac{\sum_{i=1}^{n} y_i}{n} + \frac{n\,\overline{x}\,\overline{y}}{n} = \frac{\sum_{i=1}^{n} x_i\, y_i}{n} - \overline{y}\,\overline{x} - \overline{x}\,\overline{y} + \overline{x}\,\overline{y} = \frac{\sum_{i=1}^{n} x_i\, y_i}{n} - \overline{x}\,\overline{y}$$

Es importante destacar que la covarianza puede ser difícil de interpretar en términos absolutos, ya que sus valores dependen de las escalas de las variables. Para comparar la relación entre variables de diferentes escalas, se utiliza el **coeficiente de correlación lineal**.

3.3.3. Coeficiente de Correlación Lineal

El coeficiente de correlación lineal, también conocido como **coeficiente de correlación de Pearson**, es una medida estadística que cuantifica la relación lineal entre dos variables cuantitativas en un conjunto de datos. Indica cómo cambian conjuntamente en relación con sus medias y la fuerza de esa relación. El coeficiente de correlación lineal se denota como r y se calcula mediante la siguiente fórmula:

$$r = \frac{\sigma_{x\,y}}{\sigma_x\,\sigma_y} \qquad \begin{array}{ll} \leftarrow & \text{covarianza} \\ \leftarrow & \text{producto de las desviaciones típicas} \end{array}$$

El coeficiente de correlación r tiene las siguientes propiedades:

- Es adimensional, es decir, no depende de las unidades en las que se expresen los valores de las variables, luego, si se realizan cambios en las unidades, el valor de r no varía.
- Puede tomar valores entre -1 y 1:
    - Un valor de r cercano a 1 indica una **correlación positiva fuerte**, lo que significa que las dos variables tienden a aumentar juntas en una relación lineal positiva.

- Un valor de r cercano a -1 indica una **correlación negativa fuerte**, lo que significa que las dos variables tienden a moverse en direcciones opuestas en una relación lineal negativa.
- Un valor de r cercano a 0 indica una **correlación débil** o la falta de una relación lineal significativa entre las dos variables.

El coeficiente de correlación de Pearson es una herramienta valiosa para comprender y cuantificar la relación lineal entre dos variables. Sin embargo, es importante aclarar que la correlación no implica **causalidad**, es decir, que una variable no tiene por qué ser causa de la otra. Esto lo podemos observar en la Figura 9, donde el aumento de la temperatura global tiene una alta correlación con el descenso del número de piratas, pero esto no quiere decir que si aumenta el número de piratas, la temperatura global baje, ya que no es su causa.

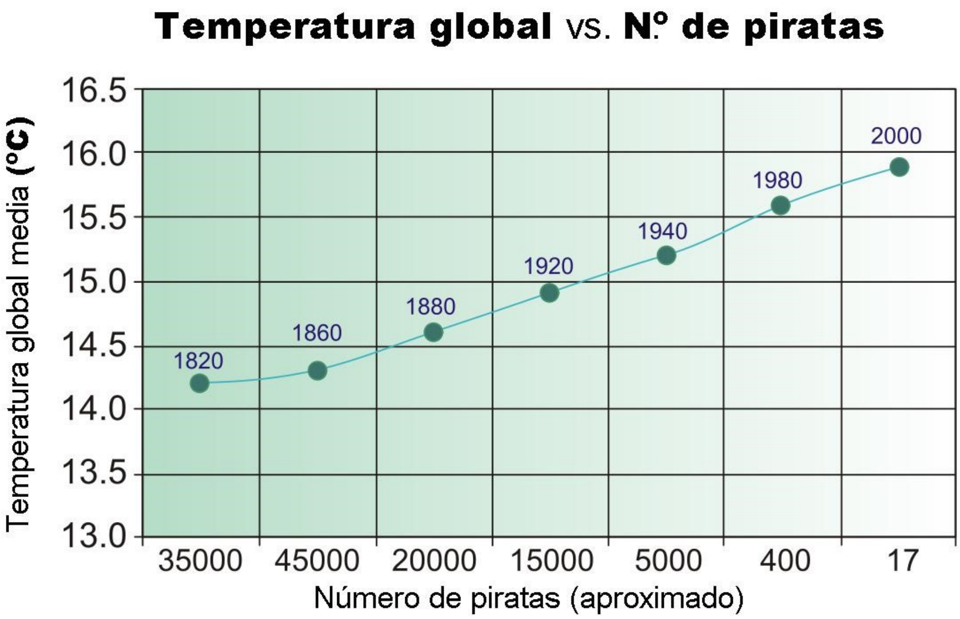


Figura 9

PiratesVsTemp.svg: RedAndr / Osado (CC).

3.3.4. Ejemplo: Móvil y Ordenador con R

Calculemos el coeficiente de correlación de la variable Móvil y Variable Ordenador.

```
#Media
```

```
n <- length(X)          #length() - nos da la longitud de la variable
Media_X <- sum(X)/n     #sum() - suma todo los parametros del variable
Media_Y <- sum(Y)/n
```

```
[1] "Media X: 1"
[2] "Media y: 1.03333333333333"
```

```
#Varianza
X_cuadrado <- X^2
Y_cuadrado <- Y^2
Varianza_X <- (sum(X^2)/(n)) - (Media_X^2)
Varianza_Y <- (sum(Y^2)/(n)) - (Media_Y^2)
```

```
[1] "Varianza X: 0.185"
[1] "Varianza Y: 0.220555555555555"
```

```
#Covarianza
Producto_XY <- X*Y
Covarianza_XY <- (sum(Producto_XY)/n) - (Media_X * Media_Y)
```

```
[1] "Covarianza de XY: 0.195833333333333"
```

```
#Correlacion lineal
Desviacion_X = sqrt(Varianza_X)
Desviacion_Y = sqrt(Varianza_Y)
Correlacion_XY <- Covarianza_XY / (Desviacion_X * Desviacion_Y )
```

```
[1] "Correlación de XY: 0.9694865"
```

El coeficiente de correlación lineal es 0.9695, por lo que deducimos que las variables X e Y tienen una correlación muy fuerte. Además, como $r > 0$, la correlación es positiva.

3.4. Recta de Regresión

El diagrama de dispersión nos permite visualizar la relación entre las variables X e Y. Lo que buscamos ahora es encontrar la recta $y = ax + b$ que "mejor se ajuste" a la nube de puntos $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$, siendo a la **pendiente de la recta** y b la **ordenada en el origen**. También definimos d_i con $i \in [1, n]$ como las diferencias entre las ordenadas de los puntos y de la recta, llamadas residuos:

$$d_i = y_i - ax_i - b$$

Si observamos la Figura 10, lo que queremos encontrar es, entre todas las rectas posibles, la que minimice la suma de las d_i. Es decir, queremos minimizar:

$$\sum_{i=1}^{n} d_i = \sum_{i=1}^{n} y_i - ax_i - b$$

Para que todos los sumandos d_i sean positivos, se utiliza d_i^2, dado que, trabajar con la función valor absoluto resulta incómodo en este caso. Por lo tanto, lo que buscamos minimizar es:

$$\sum_{i=1}^{n} d_i^2 = \sum_{i=1}^{n} (y_i - ax_i - b)^2$$

Para hallar el mínimo de esta función hay que derivar e igualar la derivada a cero. Lamentablemente, esta función tiene dos variables, a y b, y eso obliga a un método de derivación llamado derivación parcial que no está entre los objetivos de este curso.

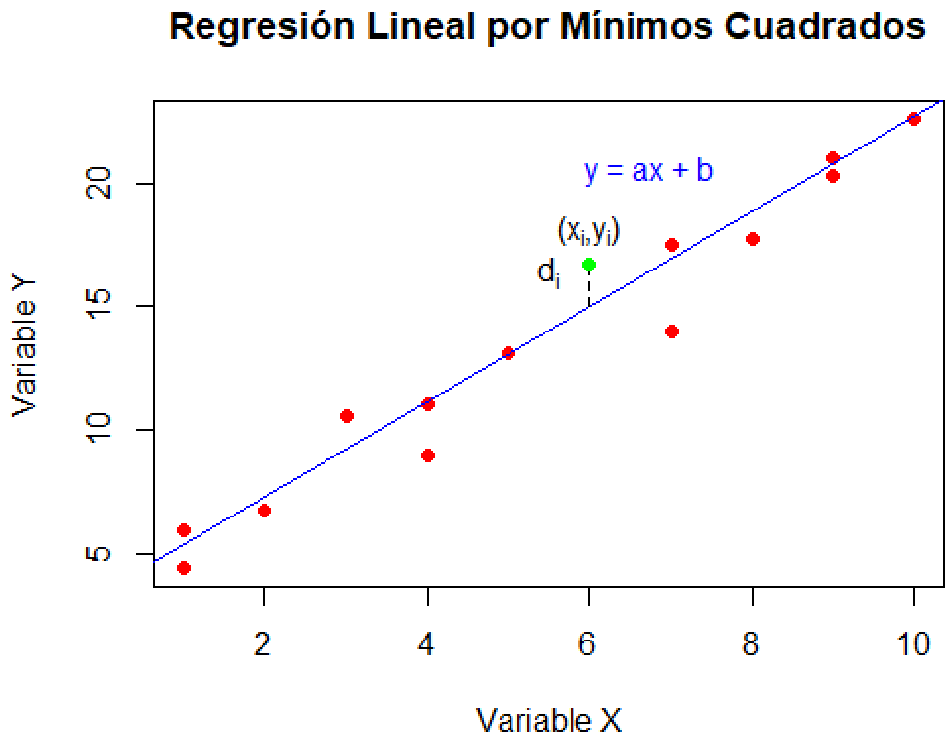


Figura 10

Gráfica de la regresión lineal por mínimos cuadrados.

En cualquier caso, terminamos obteniendo el sistema de ecuaciones:

$$\sum_{i=1}^{n} x_i y_i - a\sum_{i=1}^{n} x_i^2 - b\sum_{i=1}^{n} x_i = 0 \quad \leftarrow \text{derivada respecto } a$$

$$\sum_{i=1}^{n} y_i - a\sum_{i=1}^{n} x_i - nb = 0 \qquad \leftarrow \text{derivada respecto } b$$

Y las soluciones del sistema vienen dadas por:

$$a = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2} \qquad b = \frac{\sum_{i=1}^{n} y_i - a\sum_{i=1}^{n} x_i}{n}$$

Haciendo unas sencillas operaciones podemos simplificar estas fórmulas. Dividimos el numerador y el denominador de la fórmula de a por n^2:

$$a = \frac{\dfrac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n^2}}{\dfrac{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2}{n^2}} = \frac{\dfrac{\sum_{i=1}^{n} x_i\, y_i}{n} - \overline{x}\,\overline{y}}{\dfrac{\sum_{i=1}^{n} x_i^2}{n} - \overline{x}^2} = \frac{\sigma_{x\,y}}{\sigma_x^2}$$

Por lo tanto, a es el coeficiente de correlación. Mientras que en b observamos:

$$b = \frac{\sum_{i=1}^{n} y_i}{n} - a\frac{\sum_{i=1}^{n} x_i}{n} = \overline{y} - a\overline{x} \quad \Rightarrow \quad \overline{y} = a\overline{x} + b$$

Por lo tanto, deducimos que la recta de regresión pasa por el punto $(\overline{x}, \overline{y})$, conocido como centro de gravedad. Con estos elementos podemos escribir la **ecuación de la recta de regresión de Y sobre X** como:

$$y = \overline{y} + \frac{\sigma_{xy}}{\sigma_x}(x - \overline{x})$$

En los libros de texto, a esta recta de regresión y se le conoce como **modelo de regresión lineal** y suele representarse como $\hat{y}$.

3.4.2. Predicción, Causalidad y Coeficiente de Determinación

Como la recta de regresión se ajusta a la nube de puntos y describe su tendencia, se tiene que uno de los propósitos de esta recta suele ser la **predicción**. De esta forma, aproximamos la variable Y para valores de la variable X. El valor de aproximación se denota como $\hat{y}_0$ para un cierto valor x_0 dado. En conclusión, tenemos que $\hat{y}_0$ es el valor estimado o predicho de y_0 correspondiente a un valor x_0, como se ve en la Figura 11.

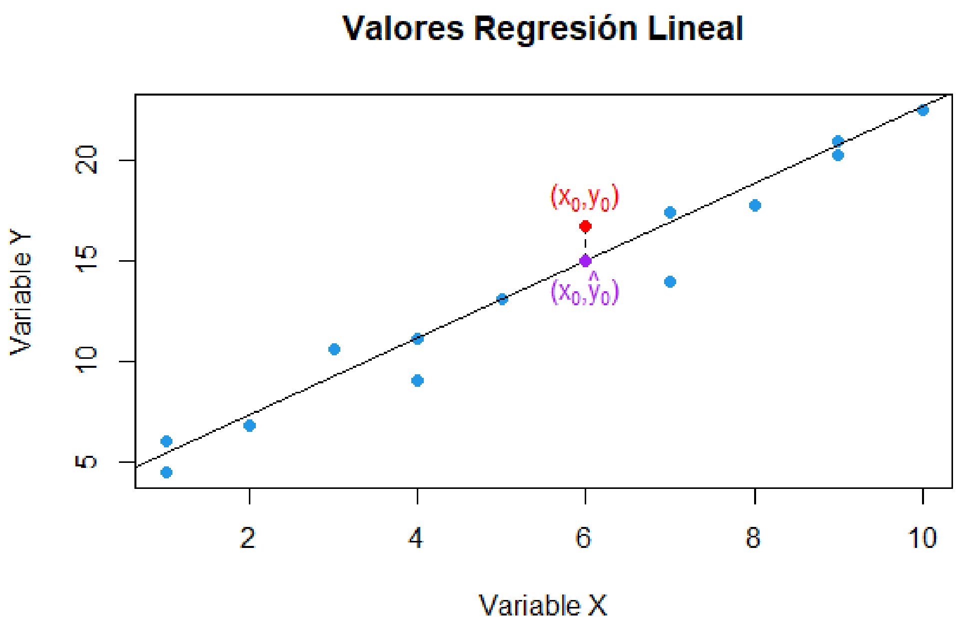


Figura 11

Gráfica de los valores predichos de la regresión lineal.

El **coeficiente de determinación** representa el porcentaje de variabilidad de la variable Y explicada por la recta de regresión, es decir, por su relación con la variable X. El coeficiente de determinación es el cuadrado del coeficiente de correlación:

$$R^2 = r^2 \qquad \text{con} \quad 0 \leq R^2 \leq 1$$

El coeficiente de determinación complementa el coeficiente de correlación al evaluar una predicción realizada a través del modelo lineal de regresión. Este coeficiente refleja el porcentaje de la variabilidad en los datos que la recta de regresión es capaz de explicar.

- $R^2 = 0$: El modelo de regresión no explica ninguna variabilidad en los datos.
- $R^2 = 1$: El modelo de regresión explica perfectamente la variabilidad en los datos.
- $0 < R < 1$: El modelo de regresión explica parte de la variabilidad en los datos. Cuanto más cercano sea a 1, mejor se explican los datos.

Al igual que en el coeficiente de correlación, un coeficiente de determinación alto no implica causalidad entre las variables.

Observación[1]: Un coeficiente de correlación con $r = 0{,}75$ podría explicar que existe dependencia lineal entre las variables. Por otro lado, el coeficiente de determinación con $R^2 = 0{,}5626$ nos indica que la recta de regresión solo es capaz de explicar un 56,26 % de la variabilidad de los datos, y por lo tanto un ajuste de regresión con el modelo lineal no es bueno.

3.4.4. Dos Rectas de Regresión

Hemos visto que la **recta de regresión** Y **sobre** X es:

$$y = \overline{y} + \frac{\sigma_{xy}}{\sigma_x}(x - \overline{x})$$

[1] Normalmente, el coeficiente de correlación lineal se utiliza para cuantificar la intensidad de concordancia que existe entre las variables, mientras que por otro lado, el coeficiente de determinación nos sirve para analizar si el modelo construido, en este caso la recta de regresión, sería un buen modelo de predicción.

De forma similar, pero cambiando el criterio de las diferencias de las d_i, que ahora se hacen respecto a las ordenadas, obtendríamos otra **recta de regresión de X sobre Y**, cuya ecuación es:

$$x = \overline{x} + \frac{\sigma_{x\,y}}{\sigma_y}(y - \overline{y}) \quad \Rightarrow \quad y = \overline{y} + \frac{\sigma_y}{\sigma_{x\,y}}(x - \overline{x})$$

Esta nueva recta pasa también por el centro de gravedad $(\overline{x}, \overline{y})$ y tiene como pendiente $\frac{\sigma_y}{\sigma_{x\,y}}$. El uso de esta recta es poco común, pero sí es interesante el posicionamiento de las dos rectas de regresión.

- Si la correlación es casi nula, las dos rectas se tornan casi perpendiculares.
- Si la correlación es fuerte, el ángulo que forman las dos rectas es pequeño.
- Si $|r|$ es próximo a 1, las rectas son casi coincidentes.

3.4.3. Ejemplo: Móvil y Ordenador con R

Escribimos la ecuación de la **recta de regresión de Y sobre X** como:

$$y = \overline{y} + \frac{\sigma_{x\,y}}{\sigma_x}(x - \overline{x}) = \overline{y} + r(x - \overline{x}) = (\overline{y} - r\overline{x}) + rx$$

Para simplificar la ecuación usaremos:

$$b = (\overline{y} - r\overline{x}) \quad \text{y} \quad a = r$$

$$y = ax + b$$

```r
#Recta de Regresion
a = Correlacion_XY
b = Media_Y -Correlacion_XY * Media_X
#round(valor, numero decimales) sirve para redondear lo decimales
equation <- paste("y =", round(a,3), "* x +", round(b,3))
equation
```

```
[1] "y = 0.969 * x + 0.064"
```

Tenemos que la recta queda $y = 0{,}969x + 0{,}064$. Veamos su gráfica:

```
#Recta de Regresion
plot(X, Y,main = "Recta de Regresion Y sobre X ",
      xlab = "Tiempo Movil",
      ylab = "Tiempo Ordenador",
      pch = 19, col = 4)
abline(b, a, col = "red")
text(0.5, 1, equation, pos = 4, col = "red")
```

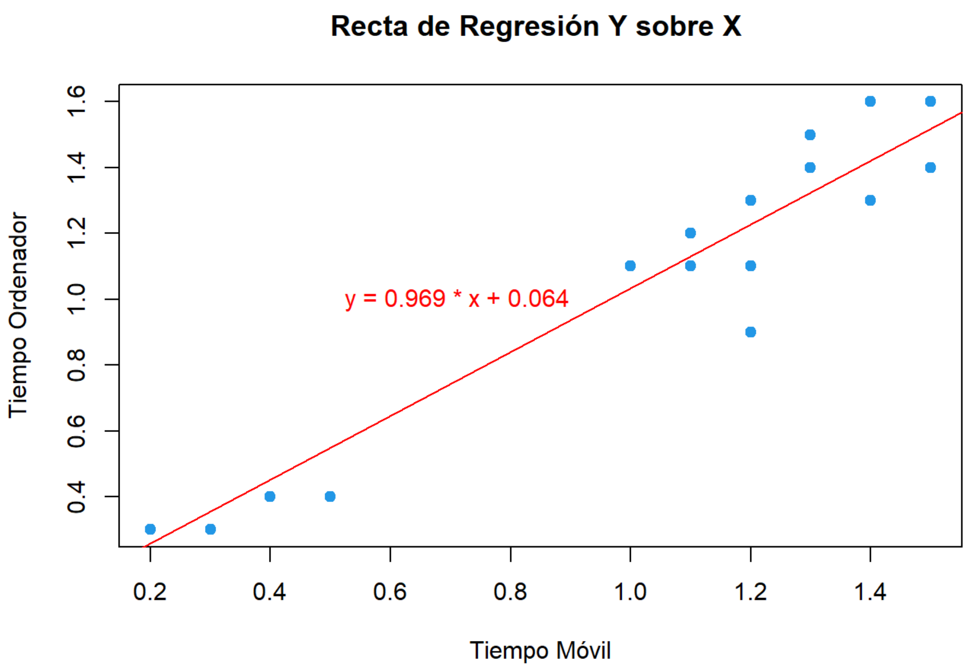


Figura 12

Gráfica de la recta de regresión de Y sobre X con su correspondiente ecuación.

También queremos conocer el tiempo de descarga en el ordenador, sabiendo que en el móvil tardó 0.8 segundos. Es decir queremos hallar y_0 dado un $x = x_0$ en la recta de regresión.

```
#Uso practico de la recta de regresion
x_0 = 0.8
y_0 = a*x_0+b
```

```
[1] (x_0, y_0) = ( 0.8 , 0.839436 )
```

Luego $(x_0, y_0) = (0.8 \ , 0.839)$. Veamos cómo queda este punto en la gráfica:

```r
#Uso practico de la recta de regresion
plot(X, Y,main = "Recta de Regresion Y sobre X ",
      xlab = "Tiempo Movil",
      ylab = "Tiempo Ordenador",
      pch = 19, col = 4)
abline(b, a, col = "red")
points(x_0, y_0, col = "green", pch = 19)
text(0.2, 0.8, equation, pos = 4, col = "red")
text(0.8, 0.7, expression("(x"[0]*",y"[0]*") =" (0.8,0.839)) , pos = 4,
    col = "green")
```

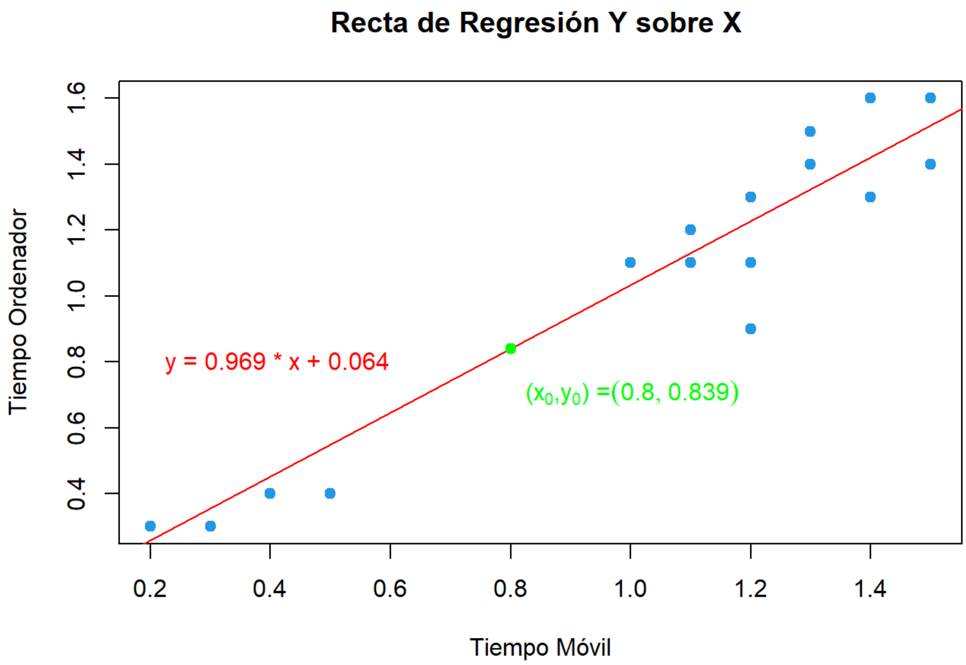


Figura 13

Gráfica de la recta de regresión de Y sobre X calculada con un valor predicho.

```r
#Coeficiente de determinacion
R = Correlacion_XY^2
```

```
[1] 93.9904 %
```

Observamos que el modelo explica el 93.99 % de la variabilidad de los datos. Por lo que consideramos que es un muy buen modelo de regresión.

Para calcular la **recta de regresión X sobre Y** usaremos:

$$x = \overline{x} + \frac{\sigma_{xy}}{\sigma_y}(y - \overline{y}) \quad \Rightarrow \quad y = \overline{y} + \frac{\sigma_y}{\sigma_{xy}}(x - \overline{x}) \quad \Rightarrow \quad y = \frac{\sigma_y}{\sigma_{xy}}x + (\overline{y} - \frac{\sigma_y}{\sigma_{xy}}\overline{x})$$

Sea: $\quad a_x = \dfrac{\sigma_y}{\sigma_{xy}} \quad$ y $\quad b_x = (\overline{y} - \dfrac{\sigma_y}{\sigma_{xy}}\overline{x}) \quad \Rightarrow \quad y = a_x x + b_x$

```
#Recta regresion X sobre Y
a_x = sqrt(Varianza_Y)/Covarianza_XY
b_x = Media_Y - a_x * Media_X
equation_2 <- paste("y =", round(a_x,3), "* x ", round(b_x,3))
```

```
[1] "y = 2.398 * x  -1.365"
```

Tenemos que la recta queda $y = 2{,}398x - 1{,}365$. Veamos su gráfica:

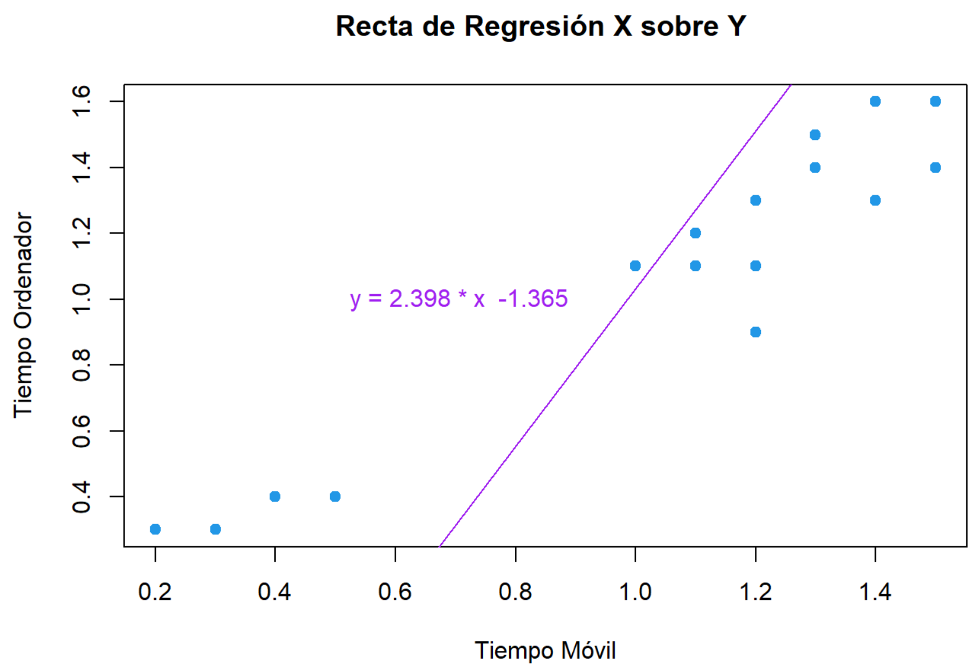


Figura 14

Gráfica de la recta de regresión de X sobre Y con su correspondiente ecuación.

Ahora veamos las dos rectas de regresión en la siguiente gráfica:

```
#Ambas rectas de regresion
```

```
plot(X, Y,main = "Rectas de Regresion ",
      xlab = "Tiempo Movil",
      ylab = "Tiempo Ordenador",
      pch = 19, col = 4)
abline(b_x, a_x, col = "purple")
abline(b, a, col = "red")
points(Media_X, Media_Y, col = "green", pch = 19)
text(0.5, 1, equation, pos = 4, col = "red")
text(0.7, 0.3, equation_2, pos = 4, col = "purple")
text(1, 1, "( x , y )", pos = 4, col = "green")
text(1, 1.08, "  _   _ ", pos = 4, col = "green")
```

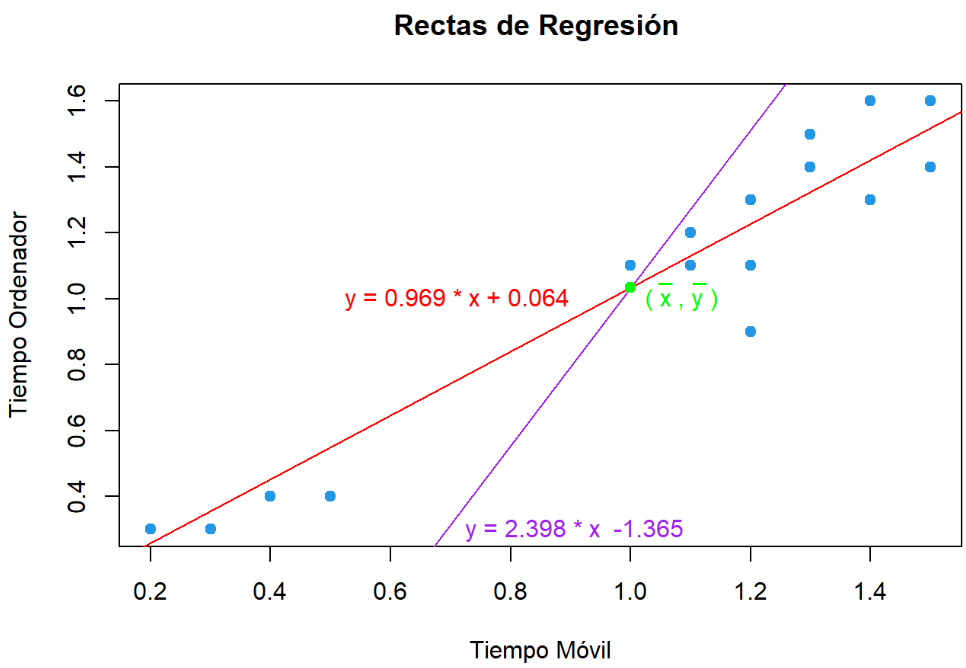


Figura 15

Gráfica de las dos rectas de regresión cortándose en el centro de gravedad.

3.5. Tablas de Doble Entrada

Hemos visto cómo trabajar cuando las variables X e Y son cuantitativas y no agrupadas. En el caso de que la Variable X y/o variable Y, sean **variables cualitativas** o **variables cuantitativas agrupadas**, los datos serán recogidos en **tablas de frecuencias conjuntas**. Estas tablas también se conocen como **tablas de frecuencias**

bidimensionales o **tablas de contingencia**.

Tabla de Frecuencias Absolutas Conjuntas					
$X \setminus Y$	y_1	y_2	...	y_m	$f_{i\cdot}$
x_1	f_{11}	f_{12}	...	f_{1m}	$f_{1\cdot} = \sum_{j=1}^{m} f_{1j}$
x_2	f_{21}	f_{22}	...	f_{2m}	$f_{2\cdot} = \sum_{j=1}^{m} f_{2j}$
...	...	...	...	...	...
x_p	f_{p1}	f_{p2}	...	f_{pm}	$f_{p\cdot} = \sum_{j=1}^{m} f_{pj}$
$f_{\cdot j}$	$f_{\cdot 1} = \sum_{i=1}^{p} f_{i1}$	$f_{\cdot 2} = \sum_{i=1}^{p} f_{i2}$	...	$f_{\cdot m} = \sum_{i=1}^{p} f_{im}$	$\mathbf{n} = \sum_{i=1}^{p} \sum_{j=1}^{m} f_{ij}$

Tabla 1

Tabla de frecuencias absolutas conjuntas.

- Sea X la variable unidimensional con $x_1, ..., x_p$ elementos, y si es posible, ordenados de menor a mayor.
- Sea Y la variable unidimensional con $y_1, ..., y_m$ elementos, y si es posible, ordenados de menor a mayor.
- Sea n el tamaño de la población.
- Sea f_{ij} la frecuencia absoluta conjunta, es decir, el número de veces que aparece el par (x_i, y_j).
- Sea $f_{i\cdot}$ la frecuencia absoluta marginal de X, es decir, el número de veces que aparece "x_i", sin tener en cuenta "y_j".
- Sea $f_{\cdot j}$ la frecuencia absoluta marginal de Y, es decir, el número de veces que aparece "y_j", sin tener en cuenta "x_i".

Ejemplo: Se ha analizado la sangre de 590 pacientes para ver qué medicina había surgido efecto. Cada paciente presentaba un tipo de enfermedad, de las 4 posibles enfermedades, y a cada uno de los pacientes se le había administrado alguno de los tres tipos de medicamentos diferentes. Los datos recogidos se resumen en la siguiente tabla:

Tabla de Frecuencias - Enfermedad y Medicina				
$X \setminus Y$	Medicina A	Medicina B	Medicina C	Total ($f_{i.}$)
Enfermedad A	55	76	89	220
Enfermedad B	76	64	1	141
Enfermedad C	42	3	40	85
Enfermedad D	28	75	41	144
Total ($f_{.j}$)	201	218	171	590

Tabla 2

Ejemplo de tabla de frecuencias absolutas conjuntas.

Observemos que de la Tabla 2 se pueden deducir algunos resultados, como:

- Teniendo la enfermedad A, fueron 89 las personas que mejoraron con la medicina C.
- Teniendo la enfermedad B, fueron 64 las personas que mejoraron con la medicina B.
- En total, 144 personas tenían la enfermedad D.
- En total, 171 personas mejoraron con la medicina C.

3.5.2. Distribuciones Marginales y Condicionadas

Distribuciones Marginales

Si en la Tabla 1 de frecuencias absolutas sumamos por filas y por columnas, obtenemos las frecuencias marginales absolutas de la variable X y de la variable Y, respectivamente. Estas tablas de frecuencias marginales absolutas se comportan como distribuciones unidimensionales.

Marginal X	
X	$f_{i\cdot}$
x_1	$f_{1\cdot}$
x_2	$f_{2\cdot}$
...	...
x_p	$f_{p\cdot}$
	$n = \sum_{i=1}^{p} f_{i\cdot}$

Marginal Y	
Y	$f_{\cdot j}$
y_1	$f_{\cdot 1}$
y_2	$f_{\cdot 2}$
...	...
y_m	$f_{\cdot m}$
	$n = \sum_{j=1}^{m} f_{\cdot j}$

Tabla 3

Tablas de frecuencias marginales absolutas de las variables X e Y.

La suma de las frecuencias absolutas marginales coinciden con la suma de las frecuencias bidimensionales de la tabla de frecuencias absolutas, es decir:

$$n = \sum_{i=1}^{p} f_{i\cdot} = \sum_{j=1}^{m} f_{\cdot j} = \sum_{i=1}^{p} \sum_{j=1}^{m} f_{ij}$$

Ejemplo: Las distribuciones marginales de la Tabla 2 de frecuencias conjuntas del Ejemplo enfermedad y medicina son:

Marginal X	
Enfermedad	Total ($f_{i\cdot}$)
Enfermedad A (x_1)	220
Enfermedad B (x_2)	141
Enfermedad C (x_3)	85
Enfermedad D (x_4)	144
	$n = 590$

Marginal Y	
Y	Total ($f_{\cdot j}$)
Medicina A (y_1)	201
Medicina B (y_2)	218
Medicina C (y_3)	171
	$n = 590$

Tabla 4

Ejemplo de distribuciones marginales absolutas.

Distribuciones Condicionadas

A partir de la tabla de frecuencias bidimensionales, si fijamos el valor de la otra variable, obtendremos otro tipo de distribución unidimensional, conocida como distribución condicionada:

Distribución de X condicionada a $Y = y_j$	
$X\|y_j$	y_j
x_1	f_{1j}
x_2	f_{2j}
...	...
x_p	f_{pj}

Distribución de Y condicionada a $X = x_i$	
$Y\|x_i$	x_i
y_1	f_{i1}
y_2	f_{i2}
...	...
y_m	f_{im}

Tabla 5

Tablas de distribuciones condicionadas de las variables X e Y.

Ejemplo: A partir de la Tabla 2, veamos la distribución de la Enfermedad (X) condicionada a la Medicina C (y_3). Y también, la distribución de la Medicina (Y) condicionada a la Enfermedad B (x_2).

Distribución de Enfermedad condicionada a Medicina C	
Enfermedad	Medicina C
Enfermedad A (x_1)	89
Enfermedad B (x_2)	1
Enfermedad C (x_3)	40
Enfermedad D (x_4)	41

Tabla 6

Ejemplo de distribución de la variable X condicionada a $Y = y_3$.

Distribución de Medicina condicionada a Enfermedad B	
Medicina	Enfermedad B
Medicina A (y_1)	76
Medicina B (y_2)	64
Medicina C (y_3)	1

Tabla 7

Ejemplo de distribución de la variable Y condicionada a $X = x_2$.

3.5.3 Independencia Estadística

Dos variables X e Y son **independientes** estadísticamente cuando el comportamiento estadístico de una de ellas no se ve afectado por los valores de la otra variable. Para ver esta independencia, nos apoyaremos en la **tabla de frecuencias relativas conjuntas**, que se calcula de la siguiente manera:

$$F_{ij} = \frac{f_{ij}}{n} \qquad \text{Para todo } i = 1, ..., p \text{ y } j = 1, ..., m$$

$$F_{i.} = \frac{f_{i.}}{n} \qquad \text{Para todo } i = 1, ..., p \qquad \leftarrow \text{Frecuencia Marginal Relativa de X}$$

$$F_{.j} = \frac{f_{.j}}{n} \qquad \text{Para todo } j = 1, ..., m \qquad \leftarrow \text{Frecuencia Marginal Relativa de Y}$$

Por lo que obtendríamos la siguiente tabla:

Tabla de Frecuencias Relativas Conjuntas					
$X \setminus Y$	y_1	y_2	...	y_m	$F_{i\cdot}$
x_1	F_{11}	F_{12}	...	F_{1m}	$F_{1\cdot} = \sum_{j=1}^{m} F_{1j}$
x_2	F_{21}	F_{22}	...	F_{2m}	$F_{2\cdot} = \sum_{j=1}^{m} F_{2j}$
...	...	...	...	...	...
x_p	F_{p1}	F_{p2}	...	F_{pm}	$F_{p\cdot} = \sum_{j=1}^{m} F_{pj}$
$F_{\cdot j}$	$F_{\cdot 1} = \sum_{i=1}^{p} F_{i1}$	$F_{\cdot 2} = \sum_{i=1}^{p} F_{i2}$	...	$F_{\cdot m} = \sum_{i=1}^{p} F_{im}$	$\mathbf{1} = \sum_{i=1}^{p} \sum_{j=1}^{m} F_{ij}$

Tabla 8

Tabla de frecuencias relativas conjuntas.

Si para todos los valores que toman ambas variables las frecuencias relativas conjuntas son iguales al producto de las frecuencias marginales, entonces las variables son independientes[2]. Es decir:

$$Fij = F_{i\cdot} \cdot F_{\cdot j} \qquad \text{Para todo } i = 1, ..., p \text{ y } j = 1, ..., m$$

Ejemplo: A partir de la Tabla 2, veamos si son independientes estadísticamente las variables Enfermedad (X) y Medicina (Y).

Tabla de Frecuencias - Enfermedad y Medicina				
$X \setminus Y$	Medicina A	Medicina B	Medicina C	Total ($f_{i.}$)
Enfermedad A	0,0932	0,1288	0,1508	0,3729
Enfermedad B	0,1288	0,1085	0,0017	0,2390
Enfermedad C	0,0712	0,0051	0,0678	0,1441
Enfermedad D	0,0475	0,1271	0,0695	0,2441
Total ($f_{.j}$)	0,3407	0,3695	0,2898	1

Tabla 9

Ejemplo de tabla de frecuencias absolutas conjuntas.

Como $F_{1\cdot} \cdot F_{\cdot 1} = 0,1270$ y $F_{11} = 0{,}0932$ podemos decir que las variables Enfermedad no son independientes estadísticamente con las variables Medicina.

3.5.4. Cálculo de Coeficiente de Correlación

En las distribuciones bidimensionales, cuando las variables X e Y toman valores cuantitativos y están agrupadas, podemos calcular el coeficiente de correlación de la variable. A partir de la Tabla 1 de frecuencias absolutas podemos obtener:

---

[2] La justificación la podemos encontrar a partir de la independencia de dos sucesos. Es decir, dos sucesos son independientes si la ocurrencia de uno de ellos no modifica la probabilidad del otro $P(A|B) = P(A)$ y $P(B|A) = P(B)$. En consecuencia, si dos sucesos son independientes, $P(A \cap B) = P(A) \cdot P(B)$.

Media de las Variables X e Y:

$$\overline{x} = \frac{\sum_{i=1}^{p} x_i \cdot f_{i\cdot}}{n} \qquad \overline{y} = \frac{\sum_{j=1}^{m} y_i \cdot f_{\cdot j}}{n}$$

Varianza de las Variables X e Y:

$$\sigma_x^2 = \frac{\sum_{i=1}^{p} f_{i\cdot} \cdot (x_i - \overline{x})^2}{n} = \frac{\sum_{i=1}^{p} f_{i\cdot} \cdot x_i^2}{n} - \overline{x}^2$$

$$\sigma_y^2 = \frac{\sum_{j=1}^{m} f_{\cdot j} \cdot (y_j - \overline{y})^2}{n} = \frac{\sum_{j=1}^{m} f_{\cdot j} \cdot y_j^2}{n} - \overline{y}^2$$

Covarianza de las Variables X e Y:

$$\sigma_{xy} = \frac{\sum_{i=1}^{p} \sum_{j=1}^{m} f_{ij}(x_i - \overline{x})(y_j - \overline{y})}{n} = \frac{\sum_{i=1}^{p} \sum_{j=1}^{m} f_{ij} x_i y_j}{n} - \overline{x}\,\overline{y}$$

Coeficiente de Correlación Lineal de las Variables X e Y:

$$r = \frac{\sigma_{x\,y}}{\sigma_x\,\sigma_y} \qquad \begin{array}{l} \leftarrow \ \text{covarianza} \\ \leftarrow \ \text{producto de las desviaciones típicas} \end{array}$$

Con estos elementos también podemos calcular las rectas de regresión:

Recta de regresión de Y sobre X:

$$y = \overline{y} + \frac{\sigma_{x\,y}}{\sigma_x}(x - \overline{x})$$

Recta de regresión de X sobre Y:

$$x = \overline{x} + \frac{\sigma_{x\,y}}{\sigma_y}(y - \overline{y}) \quad \Rightarrow \quad y = \overline{y} + \frac{\sigma_y}{\sigma_{x\,y}}(x - \overline{x})$$

Ejemplo [3]

Se han realizado dos exámenes a 900 alumnos, uno de matemáticas con 4 preguntas y otro de física con 5 preguntas. Se ha recogido el número de preguntas falladas que han obtenido los alumnos en cada examen en la Tabla 10 de frecuencias absolutas. Nos piden calcular el coeficiente de correlación y su recta de regresión.

[3] Este ejemplo tiene dos finalidades: La primera, es que está pensado para ser explicado con Excel, por eso la obtención de las tablas y algunos resultados se encuentran en el Anexo V. La otra, es que el alumno posea un ejemplo de tabla de frecuencia con variables cuantitativas.

Frecuencias Absolutas						
X\Y	**0**	**1**	**2**	**3**	**4**	**5**
0	12	18	19	16	10	6
1	8	41	32	34	37	27
2	15	26	62	37	36	20
3	4	22	53	45	64	69
4	3	19	48	59	30	28

Tabla 10

Tabla de frecuencias absolutas conjuntas con variables cuantitativas.

Empecemos calculando las frecuencias marginales de la tabla de frecuencias:

Frec. Absolutas y Frec. Marginales							
X\Y	**0**	**1**	**2**	**3**	**4**	**5**	$f_{i\cdot}$
0	12	18	19	16	10	6	81
1	8	41	32	34	37	27	179
2	15	26	62	37	36	20	196
3	4	22	53	45	64	69	257
4	3	19	48	59	30	28	187
$f_{\cdot j}$	42	126	214	191	177	150	900

Tabla 11

Tabla de frecuencias absolutas conjuntas y frecuencias marginales, con variables cuantitativas.

Al obtener las frecuencias marginales, podemos estudiar éstas por separado, creando una nueva tabla, la Tabla 12, de distribución marginal de X. La de la variable Y se hace de forma análoga. Al ser estas distribuciones unidimensionales, los parámetros como la media y la varianza son fáciles de obtener, completando la tabla de distribución marginal con los siguientes parámetros: $f_{i\cdot} \cdot x_i$, x_i^2 , $f_{i\cdot} \cdot x_i^2$.

Con estos elementos podemos calcular:

- $\overline{x} = 2090/900 = 2,322$
- $\sigma_x^2 = (6268/900) - 2,322^2 = 1,571$
- $\sigma_x = \sqrt{1,571} = 1,253$

Con la Tabla 13 de distribución marginal Y, también obtenemos:

- $\overline{y} = 2585/900 = 2,872$
- $\sigma_y^2 = (9283/900) - 2,872^2 = 2,064$
- $\sigma_y = \sqrt{1,571} = 1,436$

Distr. Marginal X				
x_i	$f_{i\cdot}$	$f_{i\cdot} \cdot x_i$	x_i^2	$f_{i\cdot} \cdot x_i^2$
0	81	0	0	0
1	179	179	1	179
2	196	392	4	784
3	257	771	9	2313
4	187	748	16	2992
Sumatorio	900	2090		6268

Tabla 12

Tabla de distribución marginal de X, con elementos necesarios para su estudio.

Distr. Marginal Y				
y_j	$f_{\cdot j}$	$f_{\cdot j} \cdot y_j$	y_j^2	$f_{\cdot j} \cdot y_j^2$
0	42	0	0	0
1	126	126	1	126
2	214	428	4	856
3	191	573	9	1719
4	177	708	16	2832
5	150	750	25	3750
Sumatorio	900	2585		9283

Tabla 13

Tabla de distribución marginal de Y, con elementos necesarios para su estudio.

Para hallar la covarianza y facilitar los cálculos, haremos uso de la Tabla 14 donde calcularemos todos los $x_i \cdot y_j \cdot f_{ij}$. Luego:

- $\Sigma x_i \cdot y_j \cdot f_{ij} = 6300$
- $\sigma_{xy} = (6300/900) - (2,322 \cdot 2,872) = 0,330$
- **Coeficiente de correlación:** $r = 0{,}183$

Tabla auxiliar $x_i \cdot y_j \cdot f_{ij}$						
X\Y	**0**	**1**	**2**	**3**	**4**	**5**
0	0	0	0	0	0	0
1	0	41	64	102	148	135
2	0	52	248	222	288	200
3	0	66	318	405	768	1035
4	0	76	384	708	480	560

Tabla 14

Tabla auxiliar que usaremos para calcular $x_i \cdot y_j \cdot f_{ij}$.

Con el coeficiente de correlación hallado, podemos calcular la recta de regresión:

- $y = ax + b$
- $a = 0{,}183$
- $b = 2,872 - 0{,}183 \cdot 2,872$
- **Recta de regresión:** $y = 0{,}183x + 2,477$

4. Propuesta Didáctica

4.1 Contextualización de la Propuesta Didáctica

El trabajo realizado se centra en el estudio de la estadística bidimensional de Primero de Bachillerato de Ciencias Sociales utilizando el software R. Por lo tanto, el marco legislativo que seguimos es el Real Decreto 243/2022, del 5 de abril, por el que se establecen la ordenación y las enseñanzas mínimas de Bachillerato, aprobado por el Ministerio de Educación y Formación Profesional, Boletín Oficial del Estado (2022).

4.1.1 Características del Alumnado

La propuesta está pensada para un curso de 1º de Bachillerato de 30 estudiantes. Por la experiencia obtenida en el prácticum, se asume que los estudiantes son poco hábiles realizando tareas con el ordenador, por lo que las características mínimas para que el alumnado pueda aprender lo que se pretende enseñar con esta propuesta didáctica son:

- Un conocimiento básico de conceptos matemáticos propios del curso de 1º Bachillerato.
- Cierta habilidad para interpretar gráficos y comprender patrones.
- Cierta capacidad de pensamiento lógico a la hora de programar.
- Paciencia y perseverancia. El proceso de aprendizaje de la programación y la estadística puede ser iterativo, y se requiere tiempo y comprensión para llegar a conclusiones sólidas.

4.1.2. Objetivos

Objetivos Generales:

- Comprender los conceptos fundamentales de estadística bidimensional, incluyendo variables, distribuciones y correlaciones y poder aplicarlos al mundo real.
- Aplicar métodos estadísticos bidimensionales para analizar y visualizar conjuntos de datos, identificando patrones y relaciones entre variables.

- Interpretar resultados estadísticos bidimensionales y utilizarlos para tomar decisiones informadas en situaciones prácticas.
- Utilizar RStudio para realizar análisis estadísticos bidimensionales y generar visualizaciones significativas.

Objetivos Específicos:

- Comprender la noción de correlación lineal y cómo medir la fuerza y dirección de la relación entre dos variables.
- Aprender a ajustar una recta de regresión a un conjunto de datos bidimensional y usarla para predecir valores futuros.
- Analizar tablas bidimensionales para describir relaciones entre variables.
- Utilizar gráficos de dispersión y otros tipos de visualizaciones para representar la relación entre dos variables.

4.1.3. Competencias Específicas y Criterios de Evaluación

Las competencias específicas y criterios de evaluación que competen a nuestra propuesta didáctica viene recogida en la Tabla 15. Esta tabla hace referencia al Anexo VI de este trabajo, donde se detalla cada punto y que nos marca el RD243/2022 Boletín Oficial del Estado (2022).

Com. Esp.	Cri. Eva.	Des. Ope.
2.	2.1. , 2.2.	STEM1, STEM2, CD3, CPSAA4, CC3, CE3.
3.	3.1. ,3.2	CCL1, STEM1, STEM2, CD1, CD2, CD3, CD5, CE3.
4.	4.1.	STEM1, STEM2, STEM3, CD2, CD3, CD5, CE3.
5.	5.1.,5.2.	STEM1, STEM3, CD2, CD3, CCEC1.
7.	7.1.,7.2.	STEM3, CD1, CD2, CD5, CE3, CCEC4.1, CCEC4.2.
8.	8.1.,8.2.	CCL1, CCL3, CP1, STEM2, STEM4, CD2, CD3, CCEC3.2.
9.	9.1.,9.2.	CP3, STEM5, CPSAA1.1, CPSAA1.2, CPSAA3.1, CPSAA3.2, CC2, CC3, CE2.

Tabla 15

Tabla resumen de la competencias específicas, criterios de evaluación y descriptores operativos.

4.1.4. Contenidos (Saberes Básicos)

Los saberes básicos que competen a nuestra propuesta se recogen de forma resumida en la Tabla 16. Esta tabla hace referencia al Anexo VII de este trabajo, donde se detalla cada punto y que nos marca el RD243/2022 Boletín Oficial del Estado (2022).

A. Sentido numérico	
1, 2, 3	1.1, 2.2, 3.1
B. Sentido de la medida	
2	2.1
C. Sentido algebraico	
1,2,4,5	1.1, 2.1, 4.1, 4.3, 5.1, 5.2
D. Sentido estocástico	
1	1.1, 1.2, 1.3, 1.4
E. Sentido socioafectivo	
1, 2, 3	1.1, 1.2, 2.2, 3.1, 3.2

Tabla 16

Tabla resumen de los saberes básicos.

4.2. Temporalización de la Propuesta Didáctica

En la propuesta hemos tenido en cuenta que las clases de 1º de bachillerato suelen durar en torno a 50 minutos. Por lo que se ha pensado que la duración de la propuesta didáctica durará unas 10 sesiones, más una sesión extra para imprevistos, tal y como viene recogido en la Tabla 17.

Temporalización Sesiones	
Sesión 1,2 y 3	Introducción a R
Sesión 4	Diagramas de dispersión
Sesión 5	Correlación lineal
Sesión 6 y 7	Recta de regresión
Sesión 8 y 9	Tablas de doble entrada
Sesión 10	Prueba final con R
Sesión 11	Sesión extra

Tabla 17

Tabla temporalización de sesiones con contenido.

4.3. Materiales

Los materiales necesarios para realizar esta propuesta didáctica en el aula son:

- Pizarra.
- Aula con proyector o pizarras interactivas.
- Ordenadores con software R y RStudio.
- Internet.

4.4. Guía y Descripción de la Propuesta Didáctica

La metodología que se usará en la propuesta será una parte expositiva, siendo estas clases magistrales, y otra parte en la que los estudiantes trabajarán en parejas resolviendo los problemas propuestos.

Los archivos Rmarkdown con los que trabajaremos con los estudiantes son:

- Introducción a R - Parte 1 (ver Anexo II).
- Introducción a R - Parte 2, (ver Anexo II).
- Ejemplo Bidimensional Móvil y Ordenador (ver Anexo III).

Las fichas de actividades se encuentran en el Anexo I y son:

- Introducción a R - Hoja 1 de Problemas.
- Actividades con R - Hoja 2 de Problemas.
- Actividades con Excel - Hoja 3 de Problemas .
- Prueba Final con R.

El archivo de Excel que utilizaremos para mostrar a los estudiantes, que contiene dos ejemplos, se puede encontrar en el Anexo V.

Ahora vamos a describir lo que haremos sesión por sesión:

Sesión 1- Introducción a R

La sesión está dividida en cuatro partes:

1. Primera parte (10 minutos): Veremos el vídeo de YouTube, Cómo Programar (2013), y así motivar al alumnado.
2. Segunda parte (10 minutos): Explicaremos qué es R y RStudio.
3. Tercera parte (15 minutos): Explicaremos la instalación de R y RStudio. Se les proporcionará también el tutorial de descarga (Aguilar Rojas, Fabiola (2022)).
4. Cuarta parte (15 minutos): Los alumnos abrirán RStudio y podrán trabajar con el archivo Introducción a R - Parte 1 para que se familiaricen con el programa.

El objetivo final de esta sesión es que entiendan la importancia de aprender un lenguaje de programación y que sean capaces de instalar R y RStudio en sus ordenadores particulares. Por último, que empiecen a familiarizarse con RStudio.

Sesión 2 - Introducción a R

La sesión está dividida en tres partes:

1. Primera parte (10 minutos): Para resolver dudas y posibles problemas que hayan tenido los estudiantes.

2. Segunda parte (15 minutos): Para terminar de ver el archivo de Introducción a R - Parte 1 y la mitad del archivo Introducción a R - Parte 2.

3. Tercera parte (25 minutos): Por parejas, empezarán a trabajar la ficha Introducción a R - Hoja 1 de Problemas.

El objetivo de esta sesión es que cojan soltura en reproducir los archivos Rmarkdown y creen su primer script con el que puedan trabajar y familiarizarse con el lenguaje de programación R.

Sesión 3 - Introducción a R

La sesión está dividida en tres partes:

1. Primera parte (10 minutos): Para resolver dudas y posibles problemas que hayan tenido los estudiantes.

2. Segunda parte (15 minutos): Para terminar de ver el archivo de Introducción a R - Parte 2.

3. Tercera parte (25 minutos): Por parejas, trabajarán la ficha Introducción a R - Hoja 1 de Problemas.

El objetivo al terminar esta sesión es que sepan hacer operaciones básicas en R, trabajar con vectores en R y realizar gráficas sencillas.

Sesión 4 - Diagrama de Dispersión

La sesión está dividida en tres partes:

1. Primera parte (15 minutos): Explicar de manera teórica la sección de diagramas de dispersión y la idea de correlación y regresión

2. Segunda parte (15 minutos): Mostrar la sección de diagrama de dispersión del archivo Rmarkdown Ejemplo Bidimensional Móvil y Ordenador en RStudio.

3. Tercera parte (15 minutos): Por parejas, trabajarán la ficha Actividades con R - Hoja 2 de Problemas.

El objetivo de esta sesión es que sepan la definición de distribución bidimensional y empezar a introducir los conceptos de curva de regresión, recta de regresión y correlación lineal. También que empiecen a identificar patrones de manera visual con ayuda en los diagramas de dispersión.

Sesión 5 - Correlación Lineal

La sesión está dividida en tres partes:

1. Primera parte (20 minutos): Explicar de manera teórica la sección de correlación lineal.

2. Segunda parte (15 minutos): Mostrar la sección de correlación lineal del archivo Rmarkdown Ejemplo Bidimensional Móvil y Ordenador en RStudio.

3. Tercera parte (15 minutos): Por parejas, trabajarán la ficha Actividades con R - Hoja 2 de Problemas.

El objetivo de esta sección es que los estudiantes vean cómo se calcula la correlación lineal de manera teórica y las propiedades de este coeficiente. También queremos que aprendan a calcular el coeficiente en R y sepan interpretar los resultados.

Sesión 6 - Rectas de Regresión

La sesión está dividida en tres partes:

1. Primera parte (25 minutos): Explicar de manera teórica la recta de regresión y su cálculo por mínimos cuadrados.

2. Segunda parte (15 minutos): Mostrar la sección de Recta de Regresión del archivo Rmarkdown Ejemplo Bidimensional Móvil y Ordenador en RStudio.

3. Tercera parte (10 minutos): Por parejas, trabajarán la ficha Actividades con R - Hoja 2 de Problemas.

El objetivo de esta sección es que los estudiantes vean cómo se deduce de manera teórica la ecuación de la recta de regresión, los elementos que la componen y sepan hallarla. También buscamos que los estudiantes sepan dibujarla e interpretarla.

Sesión 7 - Rectas de Regresión

La sesión está dividida en tres partes:

1. Primera parte (20 minutos): Explicar de manera teórica lo que significa predicción, causalidad y el coeficiente de determinación. También cómo hallar la recta de predicción X sobre Y.

2. Segunda parte (15 minutos): Mostrar la sección de predicción y causalidad, coeficiente de determinación y Recta x sobre Y del archivo Rmarkdown Ejemplo Bidimensional Móvil y Ordenador en RStudio.

3. Tercera parte (15 minutos): Por parejas, trabajarán la ficha Actividades con R - Hoja 2 de Problemas.

El objetivo de esta sección es que los estudiantes sepan hallar valores de predicción con la recta de regresión Y sobre X. La diferencia entre predicción y causalidad. Conocer la existencia de otro coeficiente de determinación. Por último, que sepan hallar la recta X sobre Y y las propiedades que tiene y comparar con la otra recta de regresión.

Sesión 8 - Tablas de Doble Entrada

La sesión está dividida en tres partes:

1. Primera parte (20 minutos): Explicar la notación utilizada en las tablas de doble entrada, las distribuciones marginales y condicionadas, y la independencia estadística de las variables.

2. Segunda parte (15 minutos): Mostrar el primer ejemplo del archivo de Excel e ir indicando los pasos seguidos para su obtención.

3. Tercera parte (15 minutos): Por parejas, trabajarán en la ficha Actividades con Excel - Hoja 3 de Problemas.

El objetivo es que los estudiantes aprendan sobre las tablas de doble entrada desarrollando la capacidad de organizar y analizar datos de manera efectiva, identificando patrones y relaciones entre dos conjuntos de variables.

Sesión 9 - Tablas de Doble Entrada

La sesión está dividida en tres partes:

1. Primera parte (15 minutos): Explicar la obtención de coeficiente de correlación y la recta de regresión cuando se da una tabla de doble entrada.

2. Segunda parte (20 minutos): Mostrar el segundo ejemplo del archivo de Excel e ir indicando los pasos seguidos para su obtención.

3. Tercera parte (15 minutos): Por parejas, trabajarán en la ficha Actividades con Excel - Hoja 3 de Problemas.

El objetivo es que los estudiantes sepan realizar los cálculos necesarios para obtención del coeficiente correlación y recta de regresión cuando tengan una tabla de doble entrada.

Al finalizar esta sesión, se informará a los estudiantes de que habrá una prueba final. Podrán llevar un hoja escrita por una cara, con las fórmulas escritas y con código de R que crean necesario, sin ejemplos. Si todo ha ido sin imprevistos, la sesión extra podría servir como clase de repaso y aclarar dudas.

Sesión 10 - Prueba Final con R

La prueba final consta de 2 ejercicios:

- El primero, son 6 actividades parecidas a las que han ido trabajando a lo largo de las sesiones.

- El segundo, son 4 cuestiones teóricas sencillas para ver si realmente han comprendido el temario.

El objetivo de esta prueba es ver si los estudiantes han alcanzado los objetivos esperados y evaluar el desarrollo obtenido a lo largo de las sesiones. También verificar si esta propuesta didáctica necesita una propuesta de mejora.

4.5. Procedimiento de Evaluación

Los instrumentos de evaluación que usaremos serán:

- **Pruebas orales:** en las que se pida a los estudiantes que expliquen con sus propias palabras los conceptos de la estadística bidimensional, como por ejemplo, el significado del coeficiente de correlación, así como que resuelvan problemas que involucren estas ideas.

- **Pruebas escritas:** en las que se pida a los estudiantes que expliquen con sus propias palabras los conceptos de estadística bidimensional, así como que resuelvan problemas que involucren estas ideas.

- **Observación:** es una técnica útil para evaluar la capacidad de los estudiantes para identificar y trabajar con los conceptos la estadística bidimensional y el programa R.

- **Producción:** se asignan tareas que involucren la resolución de problemas que impliquen la aplicación de los conceptos de estadística bidimensional, tanto individuales como en parejas.

5. Conclusiones

En la realización de este trabajo, se ha tenido en mente a los alumnos de 4º de la ESO y 1º de Bachillerato del IES Alonso Quijano y su trato con los ordenadores. La experiencia de trabajar con ordenadores por parte del alumnado era casi nula, por lo que ha supuesto un reto intentar diseñar una propuesta realista, la cual fuese accesible para ellos, y que a la vez abarcara el temario propuesto.

A lo largo de este trabajo, se ha explorado cómo la estadística bidimensional y el programa R se combinan para proporcionar una comprensión profunda de la relación entre dos variables y cómo estas relaciones pueden traducirse en información valiosa, desde el cálculo de coeficientes de correlación y la recta de regresión hasta la creación de gráficos útiles que nos permiten la visualización de los datos y extraer conclusiones significativas.

La enseñanza de la estadística bidimensional puede ser una oportunidad para introducir a los estudiantes en el pensamiento crítico, el análisis de datos y la toma de decisiones basadas en evidencia. El uso de R permite a los alumnos experimentar con conceptos estadísticos de manera práctica y visual, lo que facilita una comprensión más profunda y duradera.

Para proporcionar una visión más completa de este trabajo, vamos a realizar un análisis DAFO.

Análisis Interno

Fortalezas:

- **Aprendizaje Activo:** La combinación de ejercicios en R y estadística bidimensional promueve un enfoque de aprendizaje activo donde los estudiantes pueden explorar, analizar y visualizar datos por sí mismos.

- **Preparación para un Mundo de Datos:** Introducir a los estudiantes en herramientas como R les proporciona habilidades relevantes en una era donde la capacidad de analizar datos es cada vez más crucial.

- **Preparación para Estudios Superiores:** Introducir a los estudiantes en el uso de herramientas como R en etapas tempranas puede prepararlos para enfrentar análisis de datos más complejos en sus futuras carreras.

Debilidades:

- **Dificultad de Aprendizaje en R:** La introducción de una plataforma de programación como R puede resultar desafiante para algunos alumnos, ya que requerirá tiempo para familiarizarse con la sintaxis y las funcionalidades.

- **Falta de Adaptabilidad Educativa:** Tanto profesores como estudiantes pueden mostrar resistencia al cambio en el enfoque educativo, especialmente si están acostumbrados a métodos de enseñanza tradicionales.

Análisis Externo

Oportunidades:

- **Innovación Educativa:** La introducción de R como una herramienta para la enseñanza puede ser una oportunidad para innovar en el proceso educativo, fomentando un aprendizaje interactivo y práctico.

- **Desarrollo de Pensamiento Crítico:** La enseñanza de estadística bidimensional puede fomentar la habilidad de los estudiantes para evaluar y cuestionar la información presentada en diferentes contextos.

Amenazas:

- **Limitaciones Tecnológicas en los Centros:** La disponibilidad de ordenadores y acceso a internet puede ser limitada en algunos centros educativos, lo que podría dificultar la implementación efectiva de la propuesta.

- **Brecha Digital:** La falta de acceso a tecnología en hogares puede crear una brecha en el aprendizaje, ya que algunos estudiantes podrían tener menos exposición a prácticas digitales.

Análisis DAFO		
	Fortalezas	**Debilidades**
Análisis Internos	▪ Aprendizaje activo ▪ Preparación para un mundo de datos ▪ Preparación para estudios superiores	▪ Dificultad de aprendizaje en R ▪ Falta de adaptabilidad educativa
	Oportunidadess	**Amenazas**
Análisis Externo	▪ Innovación educativa ▪ Desarrollo de pensamiento crítico	▪ Limitaciones tecnológicas en los centros ▪ Brecha digital

Tabla 18

Tabla de Análisis DAFO.

Referencias

Aguilar Rojas, Fabiola. (2022). *DESCARGAR E INSTALAR R Y RSTUDIO PARA WINDOWS 2022.* Descargado de https://youtu.be/E5KzCLn1EsI

Boletín Oficial del Estado. (2022, abril). *Real Decreto 243/2022, de 5 de abril, por el que se establecen la ordenación y las enseñanzas mínimas del Bachillerato.* Descargado de https://www.boe.es/eli/es/rd/2022/04/05/243/con

CC BY SA Posit Software, PBC. (2023, julio). *RStudio IDE: cheetsheet.* Descargado de https://rstudio.github.io/cheatsheets/rstudio-ide.pdf

Clausell, I. (2022). *Matematicas Aplicadas a las Ciencias Sociales I, 1ºBachillerato.* Aputes Marea Verde. Descargado de https://www.apuntesmareaverde.org.es/grupos/mat/Bachillerato/BS1%2006%20Estadistica.pdff

Colera Jiménez, J., Oliveira Gonzáles, M. J., Colera Cañas, R., García Pérez, R., y Alcardo B., A. (2022). *Matematicas Aplicadas a las Ciencias Sociales I, Bachillerato.* Grupo Anaya,S.A.

Coll, V., y Pérez, P. J. (2016-2017). *Introducción a R.* Descargado de https://www.uv.es/vcoll/nota-informativa.html

Cómo Programar. (2013). *Todos deberían aprender a programar.* Descargado de https://youtu.be/Y1HHBXDL9bg

IES Alonso Quijano. (2022). *Programación General Anual.* Descargado de http://cloud.educa.madrid.org/index.php/s/h9s95E8w8NKLmOY

Outón Ruiz, J. M. (s.f.). *Distribuciones bidimensionales.* IES Fuerte de Cortadura. Descargado de https://selectividad.intergranada.com/Bach/mate1ccnn/Clase/Tema_10.pdf

Anexos

Anexo I. Fichas de Actividades Propuestas

Fichas de Actividades Propuestas

Introducción a R - Hoja 1 de Problemas

1. Resuelve con R los siguientes ejercicios:

1. $\frac{15}{2} - 4^2$
2. $\frac{8}{2} \cdot (4 + 3^2)$
3. $\frac{1125}{125} - 10^{\frac{9}{3}}$
4. $\frac{\sqrt{81}}{9} - 81^{2/3}$
5. $\frac{100}{6} \cdot (\frac{100}{20} - \frac{100}{50})^2$

2. Resuelve con R los siguientes ejercicios:

Sea x = 5, y =2 y z=3

1. $t_1 = x^2 - 4$
2. $t_2 = x^{z-3}$
3. $t_3 = (x + 3) * (y + 5)$
4. $t_4 = x^{y^z}$
5. $t_5 = \sqrt{xy^z}$
6. $t_6 = \sqrt{z^6} - \sqrt{x^y}$

3. Resuelve con R los siguientes ejercicios:

sea x = (31,19,9,7,6,5,5,5) y = (30,16,12,11,11,9,10,5)

1. n =longitud del vector x

2. sumaX = suma de todas las variables del vector x
3. sumaY = suma de todas las variables del vector y
4. mediaX = sumaX/n
5. mediaY = sumaY/n
6. cuadradoX = eleva lo elementos de x^2
7. cuadradoY = eleva lo elementos de y^2
8. VarX = ((Suma de todos los elementos de cuadradoX) / n)- mediaX
9. VarY = ((Suma de todos los elementos de cuadradoY) / n)- mediaY
10. SdX = raíz cuadrada de VarX
11. SdY = raíz cuadrada de VarY
12. XY = Producto de xy
13. sumXY = Producto escalar de xy

4. Resuelve con R los siguientes ejercicios:

Continuación del ejercicio anterior.

1. Gráfica de diagrama de puntos de x e y
2. Poner titulo "Hola Mundo"
3. Poner nombre a los ejes. eje x : "variable x", eje y : "variable y"
4. Cambiar color el color de los puntos a morado
5. Añadir recta y = 0.959x + 2.571 de color azul (**abline()**)
6. Añadir el nombre de la recta (**text()**), si es posible, texto de color azul.
7. Añadir el punto (mediaX,mediaY) de color rojo
8. Añadir el nombre del punto, "centro de gravedad", si es posible, texto de color rojo

Actividades con R - Hoja 2 de Problema

1. De un muelle se cuelgan pesos y obtenemos los siguientes alargamientos:

```
Peso gr (X)           0 10.0 30 60 90 120.0 150 250.0  0.0  10
Alargamiento cm (Y)   0  0.5  1  3  5   6.5   8  10.2 12.5  18
```

Calcular:

1. El coeficiente de correlación, ¿Qué nos indica?
2. La recta de regresión Y sobre X
3. Sí el peso es de 50 gramos, ¿Cuál es el alargamiento esperado del muelle?
4. Dibujar el Diagrama de dispersión.
5. Dibujar la recta de regresión Y sobre X.
6. Dibujar el punto en el cual encontramos el alargamiento esperado, si el peso es de 70 gramos.

2. Se toman la estatura de 12 madres y 12 hijas, y se recogen en la siguiente tabla:

```
estatura madres en cm (X) 166 168 165 156 170 167 154 169 167 158 172 175
estatura hijas en cm (Y)  168 170 168 160 171 165 157 172 165 159 172 174
```

Calcular:

1. El coeficiente de correlación, ¿Qué nos indica?
2. La recta de regresión Y sobre X
3. Si una madre mide 160 cm, ¿cuál es la estatura estimada de la hija?
4. Dibujar el Diagrama de dispersión.
5. Dibujar la recta de regresión Y sobre X.
6. Dibujar el punto en el cual encontramos la estatura de la hija, si la madre mide 174cm de altura.

3. Se recogen los pesos y las alturas de 11 jugadores de un equipo de fútbol.

peso Kg (X)	80	80	77	68	85	80	74	79	76	73	78
altura cm (Y)	187	185	184	173	189	183	177	189	180	176	182

Calcular:

1. La recta de regresión Y sobre X
2. La recta de regresión X sobre Y
3. Dibujar el Diagrama de dispersión.
4. Dibujar la recta de regresión Y sobre X, y la recta de regresión X sobre Y
5. Señalar el punto de corte
6. Si un jugador pesa 75Kg ¿cuál es su valor esperado de altura? Dibújalo en la gráfica
7. Si un jugador mide 180cm ¿cuál es su valor esperado de peso? Dibújalo en la gráfica

Actividades con Excel- Hoja 3 de Problemas

1. Se desea investigar el ganado caprino y el ganado ovino de un país. En la tabla de doble entrada adjunta se presentan los resultados de un estudio de 100 explotaciones ganaderas. Se proporcionan las frecuencias conjuntas del número de cabezas (en miles) de cabras X y ovejas Y que poseen las explotaciones.

X\Y	**0**	**1**	**2**	**3**	**4**
0	4	6	9	4	1
1	5	10	7	4	2
2	7	8	5	3	1
3	5	5	3	2	1
4	2	3	2	1	0

1. Halla las medias, varianzas y desviaciones típicas marginales.

2. Halla el número medio de ovejas condicionado a que en la explotación hay 2000 cabras.

3. Halla el número medio de cabras que tienen aquellas explotaciones que sabemos que no tienen ovejas.

4. Halla la covarianza y el coeficiente de correlación entre ambas variables.

2. Indicar si la variables X e Y son independientes.

X\ Y	**1**	**2**	**3**
5	1	10	5
10	2	20	10
15	4	40	20

3. Un sociólogo afirma que las mujeres se casan más jóvenes que los hombres. Para apoyar dicha afirmación presenta la siguiente tabla obtenida en una encuesta a 50 parejas, donde H representa la edad de los hombres y M la de las mujeres

X\Y	**[15,20)**	**[20,25)**	**[25,30)**	**[30,35)**
[15,20)	1	4	3	0
[20,25)	1	5	7	3
[25,30)	0	2	8	8
[30,35)	0	0	2	6

1. ¿Es correcta la afirmación del psicólogo? Razona la respuesta

2. ¿Qué edad se puede esperar para un hombre casado con una mujer de 25 años?

3. Estudia la fiabilidad de la predicción del apartado anterior.

Prueba Final con R

1. Se han recogido la potencia en caballos de vapor y la velocidad máxima que alcanzan de nueve modelos diferentes de automóviles:

Cv (X)	92	90	102	95	130	145	150	322	295
Km/h (Y)	180	175	175	170	185	193	188	265	252

Calcular:

1. El coeficiente de correlación, ¿Qué nos indica?
2. La recta de regresión Y sobre X
3. Si el automóvil tiene 120Cv, ¿Cuál es la estimación de la velocidad máxima alcanzada?
4. Dibujar el Diagrama de dispersión.
5. Dibujar recta de regresión Y sobre X.
6. Dibujar el punto en el cual encontramos la velocidad máxima esperada, si tiene 200Cv.

2. Contesta las siguientes preguntas:

1. ¿Es posible que la covarianza y el coeficiente de correlación tengan signos distintos?
2. ¿Qué indica el punto de corte de las recta de regresión Y sobre X, y la recta X sobre Y, de una distribución bidimensional?
3. **Verdadero o Falso.** Si |r| es cercano a 1. Entonces r indica solamente que la correlación es fuertemente positiva.
4. **Verdadero o Falso.** Si |r| es cercano a 0. Entonces r indica solamente que la correlación es débil.

Razona todas la respuestas.

Anexo II. Introducción a R

Introducción a R - Parte 1

Introducción a R - Parte1

Carlos Suárez Díaz

2023-08-16

Operaciones Básicas

Veamos como podemos usar R como una calculadora

Suma y Resta

```
3+4
```

```
## [1] 7
```

```
82-64
```

```
## [1] 18
```

```
15+78-18
```

```
## [1] 75
```

Producto y División

```
3*4
```

```
## [1] 12
```

```
82/64
```

```
## [1] 1.28125
```

```
15*78/5
```

```
## [1] 234
```

Potencias y raíces cuadradas

```
#Potencias
3^3
```

```
## [1] 27
```

```
2^7
```

```
## [1] 128
```

```
#Raíz cuadrada
sqrt(9)
```

```
## [1] 3
```

```
sqrt(35)
```

```
## [1] 5.91608
```

Jerarquía de operaciones

Es importante recalcar que R sigue las jerarquía de operaciones, y que si queremos hacer una operación que no sea prioritaria sobre otra en la jerarquía, usaremos ().

```
3*5+3
```

```
## [1] 18
```

```
3*(5+3)
```

```
## [1] 24
```

```
#sin paréntesis, se opera según la jerarquía
5 ^ 2 + 30 / sqrt(25)
```

```
## [1] 31
```

```
#con paréntesis respetando la jerarquía
(5^2) + (30 / sqrt(25))
```

```
## [1] 31
```

```
#con paréntesis sin respetar a la jerarquía
((5^2)+30)/ sqrt(25)
```

```
## [1] 11
```

Otras Operaciones interesantes

```
#logaritmo neperiano
log(10)
```

```
## [1] 2.302585
```

```
#número pi
pi
```

```
## [1] 3.141593
```

```
pi*10
```

```
## [1] 31.41593
```

Definir variable

Podemos definir variable como x o y de la forma “x <- 5” o “y = 8”.

```
#definir variable
x = 5
y = 8
z = 3
```

Definidas las variables, podemos hacer operaciones con ellas:

x = 5 y = 8 z = 3

```
#definir variable
x+y
```

```
## [1] 13
```

```
z-x
```

```
## [1] -2
```

```
x*y
```

```
## [1] 40
```

```
9*x/z
```

```
## [1] 15
```

```
x^2
```

```
## [1] 25
```

```
x^z
```

```
## [1] 125
```

Podemos definir estas operacioenes en otras variables:

```
#definir variable
t1 <- x+y
t2 <- z-x
t3 <-9*x/z
```

En la siguiente clase se detallas más operaciones.

Resolver Problemas

Practicar con la hoja 1 de problemas ejercicios 1 y 2

Introducción a R - Parte 2

Introducción a R - Parte 2

Carlos Suárez Díaz

2023-08-16

Vectores en R

R trabaja con carácter vectorial, es decir, que opera componente a componente.

Función c()

Para definir un vector utilizamos la función c().

En Madrid, Barcelona, Valencia y Málaga tienen: 35,28,17,13 centros comerciales y 27,24,18,25 bancos respectivamente.

```
comercial <- c(35,28,17,13) #número de centros comerciales por ciudad
banco <- c(27,24,18,25) #número de bancos por ciudad
comercial #para ver los resultados
```

```
## [1] 35 28 17 13
```

```
banco
```

```
## [1] 27 24 18 25
```

Hemos creado dos vectores, el vector “comercial” y el vector “banco”.

Normalmente los vectores no pueden llevar espacios y suelen denominarse con letras o abreviaturas.

```
x <- c(7,8,3,4)
y <- c(1,2,9,6)
x
```

```
## [1] 7 8 3 4
```

```
y
```

```
## [1] 1 2 9 6
```

Operaciones básicas con vectores de R

Como hemos dicho R trabaja con carácter vectorial, es decir, elemento a elemento.

Las operaciones básicas, como suma, resta, producto, división, potencias y raíces, también las hace elemento a elemento.

```
x <- c(7,8,3,4)
y <- c(1,2,9,6)
x+y
```

```
## [1]  8 10 12 10
```

```
x-y
```

```
## [1]  6  6 -6 -2
```

```
x*y
```

```
## [1]  7 16 27 24
```

```
x/y
```

```
## [1] 7.0000000 4.0000000 0.3333333 0.6666667
```

```
x^2
```

```
## [1] 49 64  9 16
```

```
sqrt(x)
```

```
## [1] 2.645751 2.828427 1.732051 2.000000
```

Si además definimos una nueva variable con operaciones de otras variables que sean vectores, se crea una variable vectorial con estas operaciones y podemos utilizarlas posteriormente.

```
t1 <- x+y
t2 <- x-y
t3 <- t1 * t2
t1
```

```
## [1]  8 10 12 10
```

```
t2
```

```
## [1]  6  6 -6 -2
```

```
t3
```

```
## [1]  48  60 -72 -20
```

Cuidado que se puede redefinir una misma variable

```
t3
```

```
## [1]  48  60 -72 -20
```

```
t3 <- x^2
t3
```

```
## [1] 49 64  9 16
```

Funciones útiles para trabajar con vectores:

Función length()

nos devuelve la longitud de un vector

```
x <- c(1,2,4,5,7,8)
y <- c(8,9,1)
n1 <- length(x) #Longitud del vector x
n2 <- length(y) #Longitud del vector y
n1
```

```
## [1] 6
```

```
n2
```

```
## [1] 3
```

Función sum()

suma todos los elementos de un vector

```
x <- c(1,2,4,5,7,8)
y <- c(8,9,1)
sumax <- sum(x) #suma todos los elementos del vector x
sumay <- sum(y) #suma todos los elementos del vector y
sumax
```

```
## [1] 27
```

```
sumay
```

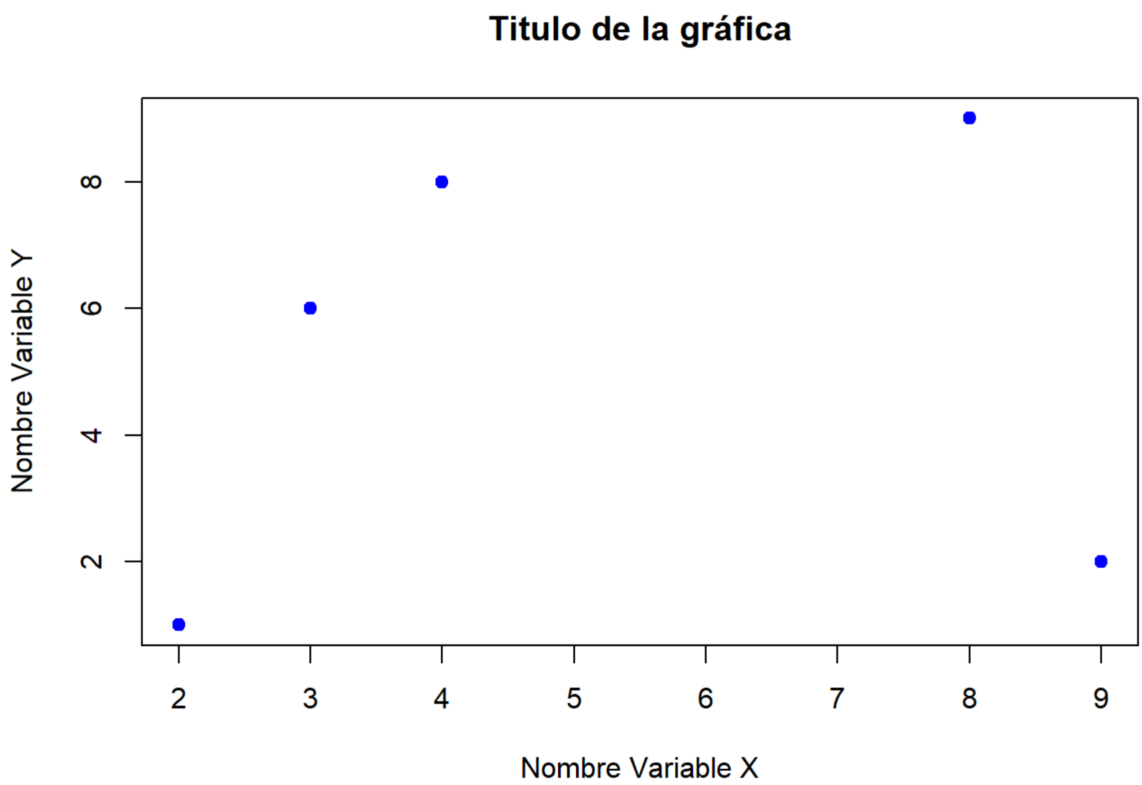
```
## [1] 18
```

Gráficas en R

Función Plot()

Nosotros trabajaremos con los diagramas de puntos. Para ellos usaremos la función plot(variable X, variable y, …)

```
x <-c(2,3,4,8,9)
y <-c(1,6,8,9,2)
plot(x, y,
     main = "Titulo de la gráfica",  #Titulo Gráfica
     xlab = "Nombre Variable X",     #Titulo eje X
     ylab = "Nombre Variable Y",     #Titulo eje Y
     pch = 19,                       #Tipo de punto
     col = "blue")                   #color de los puntos
```

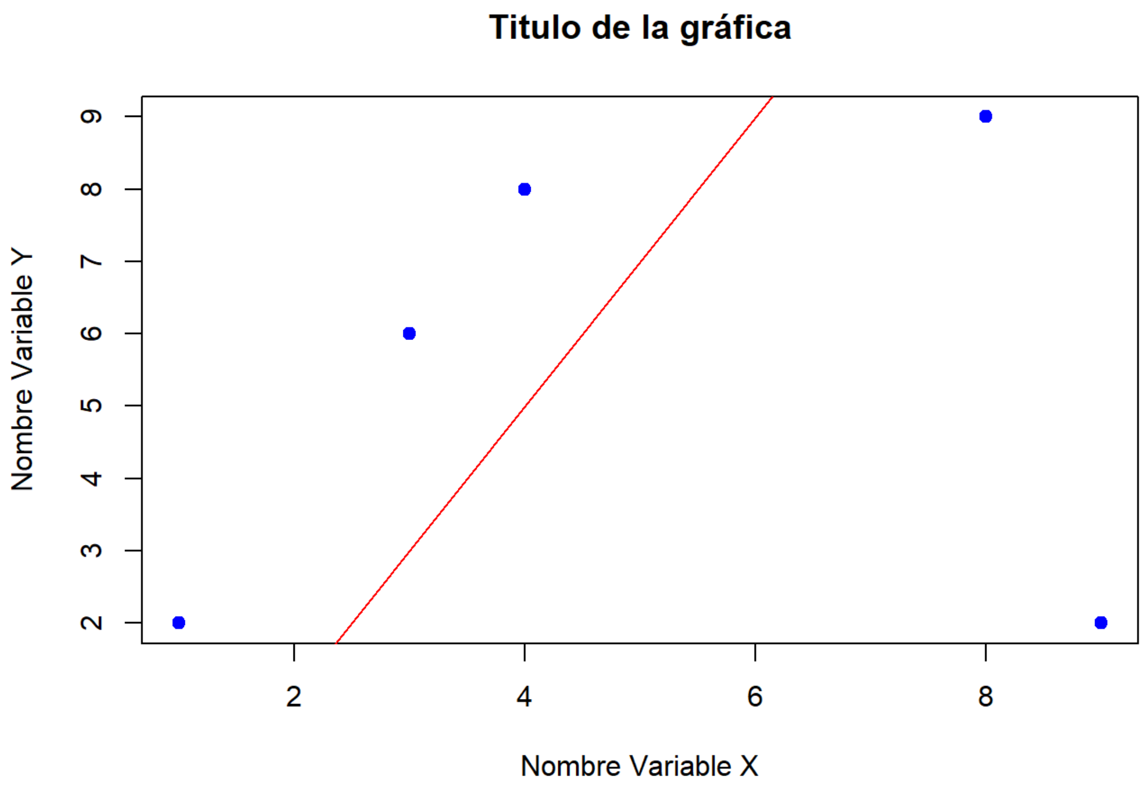


Función Abline()

Sirve para añadir rectas a una gráfica. abline() va seguido del Plot() necesariamente. Si queremos dibujar la recta y=ax+b donde a es la pendiente y b la ordenada en el origen, escribimos abline(ordendana en origen,pendiente,…).

Dibujemos la recta y=2x+3

```
x <-c(1,3,4,8,9)
y <-c(2,6,8,9,2)

#Sea la recta y=2x-3
b=-3        #término independiente
a=2        #pendiente
plot(x, y,
     main = "Titulo de la gráfica",  #Titulo Gráfica
     xlab = "Nombre Variable X",     #Titulo eje X
     ylab = "Nombre Variable Y",     #Titulo eje Y
     pch = 19,                       #Tipo de punto
     col = "blue")                   #color de los puntos en inglés o con números
abline(b, a, col = "red")
```

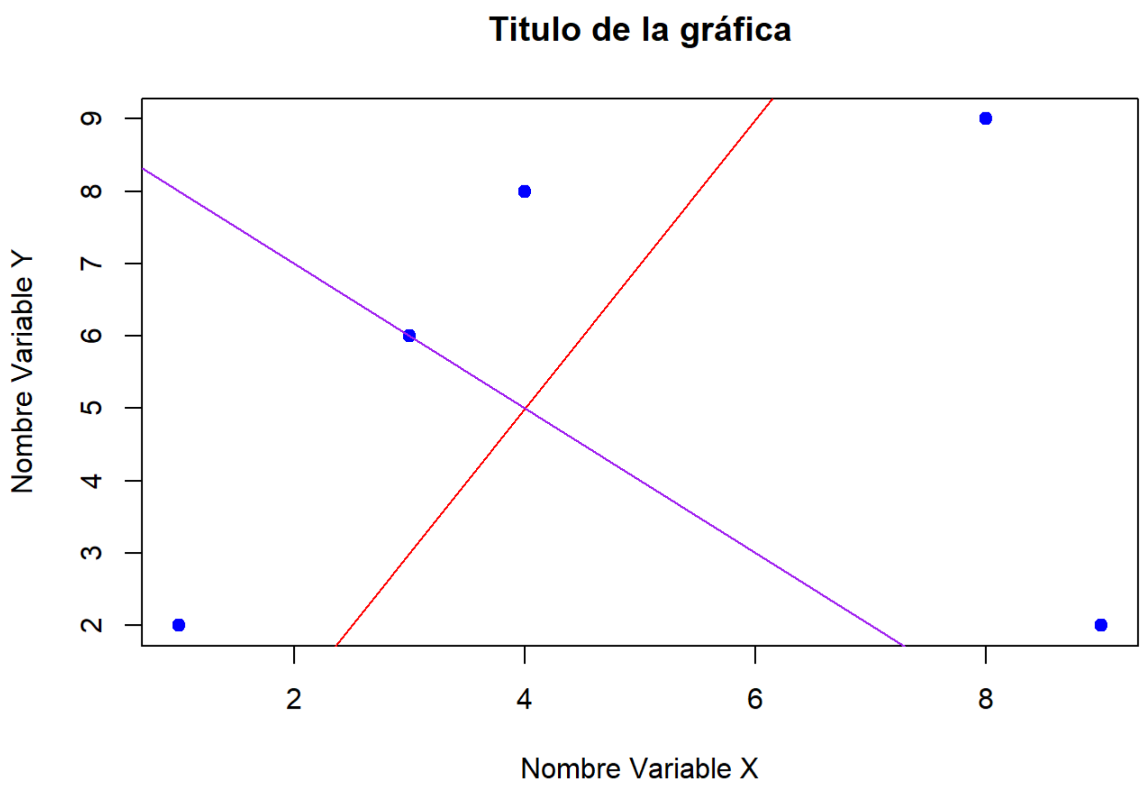


Si seguimos poniendo ablines podemos añadir tantas rectas como queramos. Dibujemos las rectas:

y=2x-3

y=-x+9

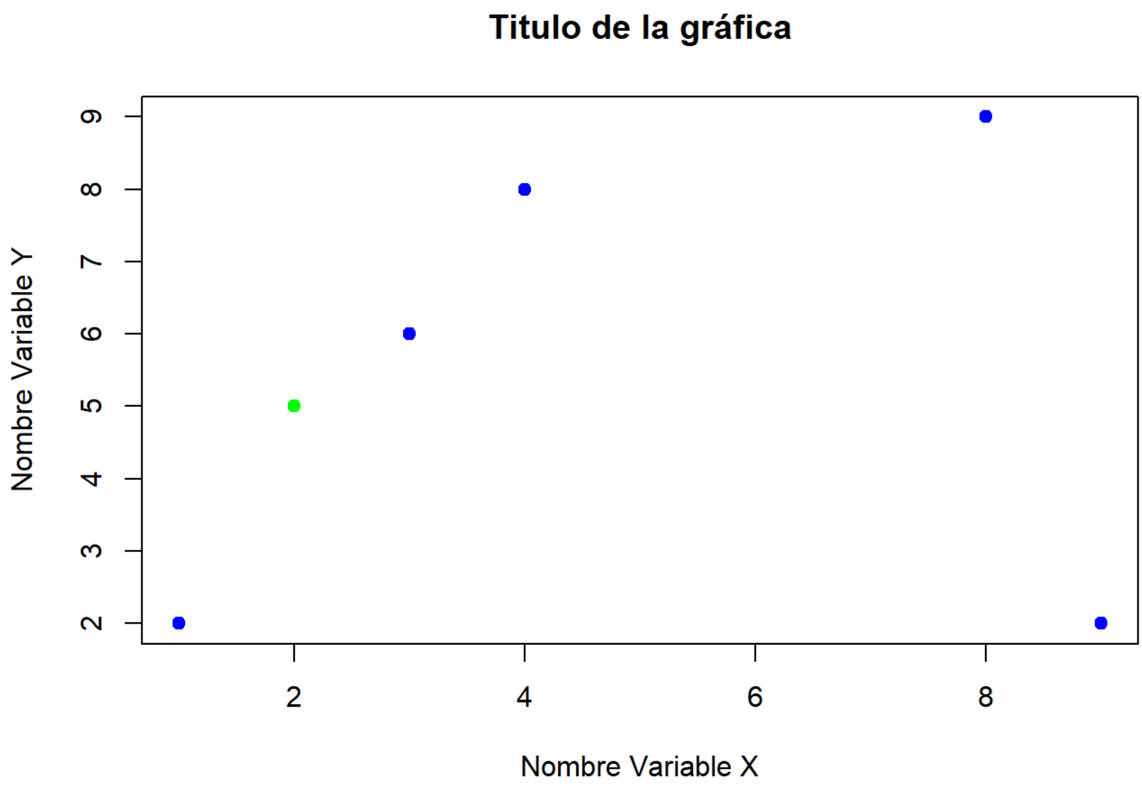
```
x <- c(1,3,4,8,9)
y <- c(2,6,8,9,2)
#Sea la recta1: y=2x-3
b1 <- -3        #término independiente
a1 <- 2         #pendiente

#Sea la recta 2: y=-x+9
b2 <- 9        #término independiente de la recta 2
a2 <- -1        #término independiente de la recta 2

plot(x, y,
     main = "Titulo de la gráfica",  #Titulo Gráfica
     xlab = "Nombre Variable X",     #Titulo eje X
     ylab = "Nombre Variable Y",     #Titulo eje Y
     pch = 19,                       #Tipo de punto
     col = "blue")                   #color de los puntos en inglés o con números
abline(b1, a1, col = "red")
abline(b2, a2, col = "purple")
```



Función points()

Sirve para añadir puntos a las gráficas. points() va seguido del Plot() necesariamente. Si queremos dibujar un punto P(p1,p2) de coordenadas, usamos point(p1,p2,…).

```
x <-c(1,3,4,8,9)
y <-c(2,6,8,9,2)
#Sea P(2,5)
p1=2        #coordenada x del punto
p2=5        #coordenada y del punto
plot(x, y,
     main = "Titulo de la gráfica",  #Titulo Gráfica
     xlab = "Nombre Variable X",     #Titulo eje X
     ylab = "Nombre Variable Y",     #Titulo eje Y
     pch = 19,                       #Tipo de punto
     col = "blue")                   #color de los puntos en inglés o con números
points(p1, p2,
       col = "green",                #Tipo de punto
       pch=19)                       #color de los puntos en inglés o con números
```



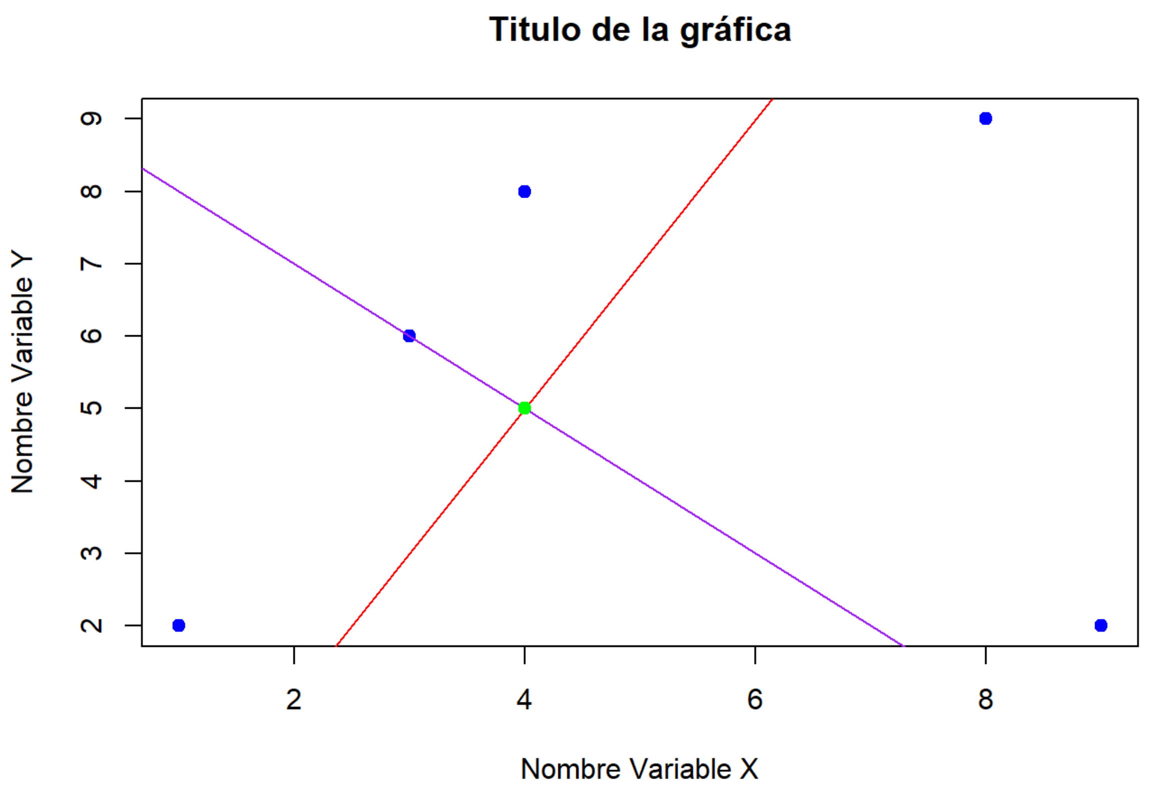
Podemos usar abline y point seguidos. Si resolvemos las dos rectas anteriores:

y=2x-3

y=-x+9

Tenemos que se cortan el (4,5).

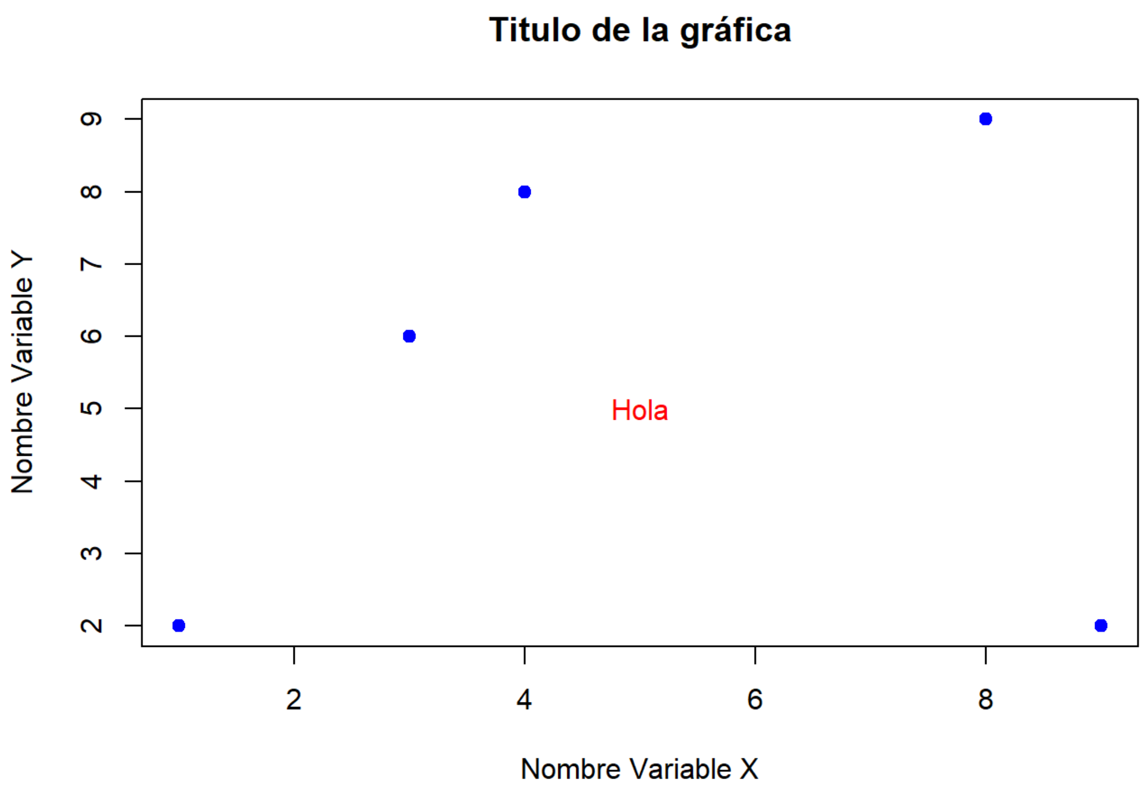
```
x <-c(1,3,4,8,9)
y <-c(2,6,8,9,2)

#Sea la recta1: y=2x-3
b1 <- -3         #término independiente
a1 <- 2          #pendiente

#Sea la recta 2: y=-x+9
b2 <- 9         #término independiente de la recta 2
a2 <- -1         #término independiente de la recta 2

#Sea punto de corte (5,5)
p1=4         #coordenada x del punto
p2=5         #coordenada y del punto


plot(x, y,
     main = "Titulo de la gráfica",  #Titulo Gráfica
     xlab = "Nombre Variable X",     #Titulo eje X
     ylab = "Nombre Variable Y",     #Titulo eje Y
     pch = 19,                       #Tipo de punto
     col = "blue")                   #color de los puntos en inglés o con números
abline(b1, a1, col = "red")
abline(b2, a2, col = "purple")
points(p1,p2,col="green",pch=19)
```



Función text()

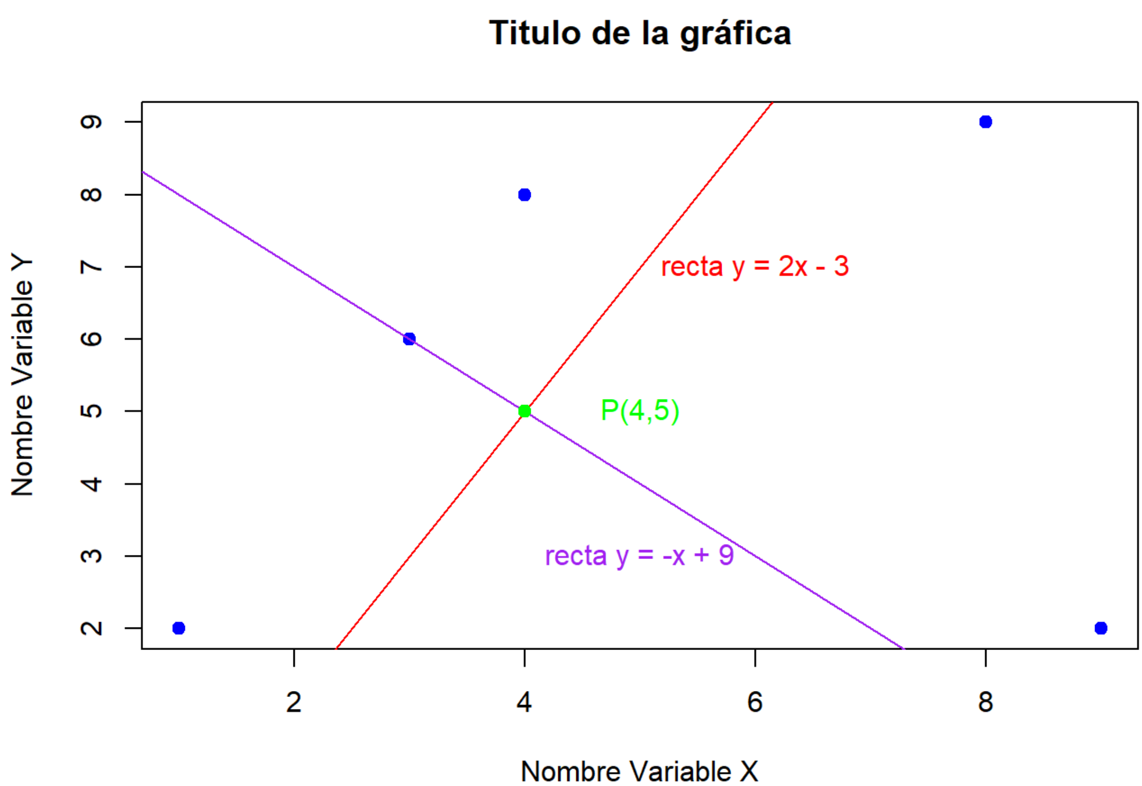
Sirve para colocar texto en una gráfica. text() va detrás de un plot() necesariamente. si queremos escribir un texto en una gráfica solamente (coordenada_x,coordenada_y, “texto en comillas”, col= “color”).

```
x <-c(1,3,4,8,9)
y <-c(2,6,8,9,2)


plot(x, y,
     main = "Titulo de la gráfica",  #Titulo Gráfica
     xlab = "Nombre Variable X",     #Titulo eje X
     ylab = "Nombre Variable Y",     #Titulo eje Y
     pch = 19,                       #Tipo de punto
     col = "blue")                   #color de los puntos en inglés o con números
text(5, 5, "Hola",col="red")
```



También se suele usar para poner nombres a lo elementos de la gráfica, cuando hay muchos.

```
x <-c(1,3,4,8,9)
y <-c(2,6,8,9,2)

#Sea la recta1: y=2x-3
b1 <- -3        #término independiente
a1 <- 2         #pendiente

#Sea la recta 2: y=-x+9
b2 <- 9        #término independiente de la recta 2
a2 <- -1        #término independiente de la recta 2

#Sea punto de corte (5,5)
p1=4        #coordenada x del punto
p2=5        #coordenada y del punto


plot(x, y,
     main = "Titulo de la gráfica",    #Titulo Gráfica
     xlab = "Nombre Variable X",       #Titulo eje X
     ylab = "Nombre Variable Y",       #Titulo eje Y
     pch = 19,                         #Tipo de punto
     col = "blue")                     #color de los puntos en inglés o con números
abline(b1, a1, col = "red")            #Recta 1
text(6,7,"recta y = 2x - 3",col="red")
abline(b2, a2, col = "purple")         #Recta 2
text(5,3,"recta y = -x + 9",col="purple")
points(p1,p2,col="green",pch=19)       #Punto de corte
text(5,5,"P(4,5)",col="green")
```



Resolver problemas

Practicar con la hoja 1 de problemas ejercicios 3 y 4

Anexo III. Ejemplo Bidimensional Móvil y Ordenador

Ejemplo Bidimensional Móvil y Ordenador

Carlos Suárez Díaz

2023-08-14

Diagrama Dispersión

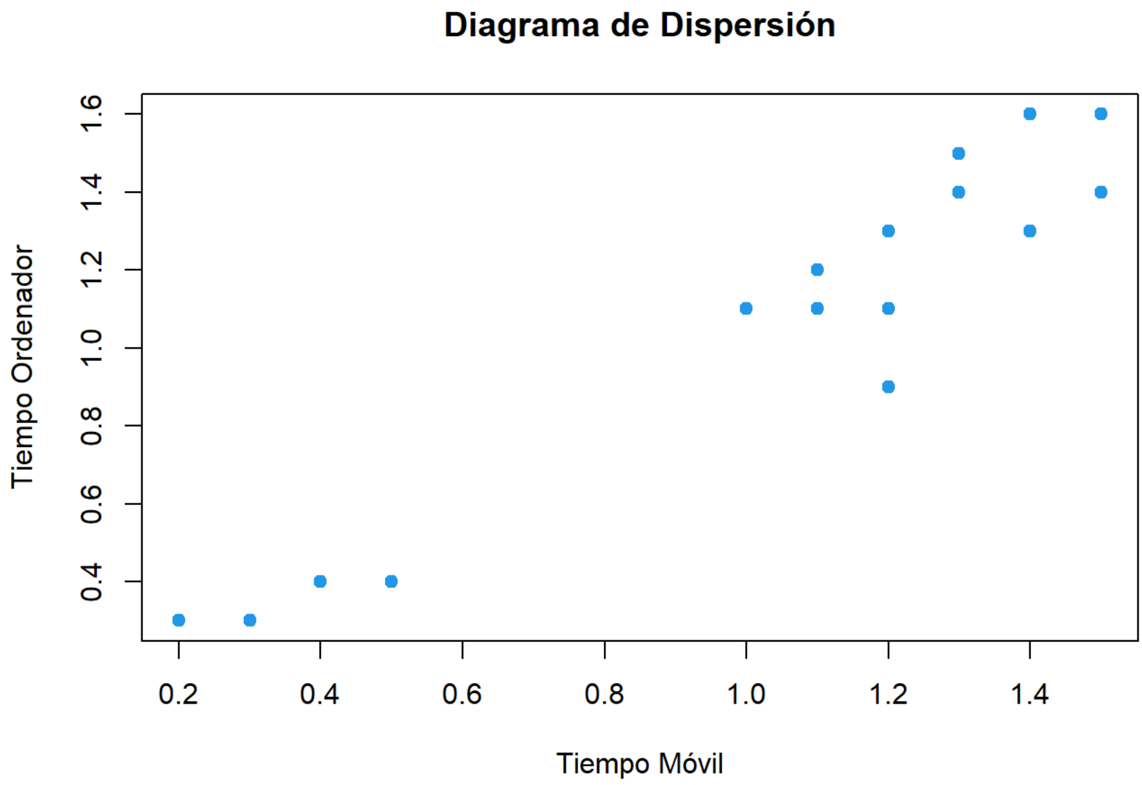
Se recogen el tiempo de descarga, en segundos, de 24 archivos en un móvil (X) y en un ordenador (Y).

```
#Datos
X <- c(1.2, 1, 1.1, 0.5, 1.1, 1.5, 1, 1.4, 1.4, 1.3, 0.4, 0.3,0.3, 1.5, 1.4, 1.1, 1.2, 1.2,
0.4, 0.5, 1.3, 1.5, 1.2, 0.2)
Y <- c(1.3, 1.1, 1.2, 0.4, 1.2, 1.4, 1.1, 1.6, 1.6, 1.5, 0.4, 0.3, 0.3, 1.6, 1.3, 1.1, 1.3,
1.1, 0.4, 0.4, 1.4, 1.6, 0.9, 0.3)


#Mostrar Datos
aux_1 <- rbind(X,Y)
rownames(aux_1) <-  c("Movil", "Ordenador")
aux_1
```

```
##           [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13]
## Movil      1.2  1.0  1.1  0.5  1.1  1.5  1.0  1.4  1.4   1.3   0.4   0.3   0.3
## Ordenador  1.3  1.1  1.2  0.4  1.2  1.4  1.1  1.6  1.6   1.5   0.4   0.3   0.3
##           [,14] [,15] [,16] [,17] [,18] [,19] [,20] [,21] [,22] [,23] [,24]
## Movil       1.5   1.4   1.1   1.2   1.2   0.4   0.5   1.3   1.5   1.2   0.2
## Ordenador   1.6   1.3   1.1   1.3   1.1   0.4   0.4   1.4   1.6   0.9   0.3
```

Ahora dibujamos el diagrama de dispersión con la función plot:

```
#Gráfica
plot(X, Y,main = "Diagrama de Dispersión", #Titulo Gráfica
     xlab = "Tiempo Móvil",      #Titulo eje X
     ylab = "Tiempo Ordenador",  #Titulo eje Y
     pch = 19, col = 4) #"pch"=Tipo de punto, "col"= color del punto
```



Correlación Lineal

Lo que buscamos es:

$$r = \frac{\sigma_{xy}}{\sigma_x \, \sigma_y} \qquad \begin{array}{l} \leftarrow \quad \text{covarianza} \\ \leftarrow \quad \text{producto de las desviaciones típicas} \end{array}$$

Para ello empecemos buscando los siguientes elementos:

Centro de Gravedad

$\overline{x}$ es la media de la variable X e $\overline{y}$ es la media de la variable Y:

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n} \qquad \overline{y} = \frac{\sum_{i=1}^{n} y_i}{n}$$

```
#Media
n <- length(X)       #length() - nos da la longitud del vector
Media_X <- sum(X)/n  #sum() - suma todo los parámetros del vector
Media_Y <- sum(Y)/n

#Mostrar Datos
print(paste("Media X:", Media_X))
```

```
## [1] "Media X: 1"
```

```
print(paste("Media y:", Media_Y))
```

```
## [1] "Media y: 1.03333333333333"
```

Varianza

$$\sigma_x^2 = \frac{\sum_{i=1}^n (x_i - \overline{x})^2}{n} = \frac{\sum_{i=1}^n x_i^2}{n} - \overline{x}^2 \qquad \sigma_y^2 = \frac{\sum_{i=1}^n (y_i - \overline{y})^2}{n} = \frac{\sum_{i=1}^n y_i^2}{n} - \overline{y}^2$$

```
#Varianza
X_cuadrado <- X^2
Y_cuadrado <- Y^2

Varianza_X <- (sum(X^2)/(n)) - (Media_X^2)
Varianza_Y <- (sum(Y^2)/(n)) - (Media_Y^2)

#Mostrar Datos
print(paste("Varianza X:", Varianza_X))
```

```
## [1] "Varianza X: 0.185"
```

```
print(paste("Varianza Y:", Varianza_Y))
```

```
## [1] "Varianza Y: 0.220555555555555"
```

Covarianza

$$\sigma_{xy} = \frac{\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y})}{n} = \frac{\sum_{i=1}^n x_i \, y_i}{n} - \overline{x}\,\overline{y}$$

```
#Covarianza

Producto_XY<- X*Y
Covarianza_XY <- (sum(Producto_XY)/n) - (Media_X * Media_Y)

#Mostrar Datos
print(paste("Covarianza de XY:", Covarianza_XY))
```

```
## [1] "Covarianza de XY: 0.195833333333333"
```

Coeficiente de Correlación Lineal

$$r = \frac{\sigma_{x\,y}}{\sigma_x\,\sigma_y} \qquad \begin{array}{ll} \leftarrow & \text{covarianza} \\ \leftarrow & \text{producto de las desviaciones típicas} \end{array}$$

```
#Correlación lineal

Desviacion_X = sqrt(Varianza_X)
Desviacion_Y = sqrt(Varianza_Y)
Correlacion_XY <- Covarianza_XY / (Desviacion_X * Desviacion_Y )
Correlacion_XY
```

```
## [1] 0.9694865
```

```
#Mostrar Datos
cat("Correlación de XY:", Correlacion_XY)
```

```
## Correlación de XY: 0.9694865
```

El coeficiente de correlación lineal es 0.9695, por lo que deducimos que las variables X e Y tiene una correlación muy fuerte, además, como $r > 0$, la correlación es positiva.

Recta de Regresión

la ecuación de la recta de regresión de Y sobre X como:

$$y = \overline{y} + \frac{\sigma_{xy}}{\sigma_x}(x - \overline{x}) = \overline{y} + r(x - \overline{x}) = (\overline{y} - r\overline{x}) + rx$$

Para simplificar la ecuación usaremos:

$$b = (\overline{y} - r\overline{x}) \quad \text{y} \quad a = r$$

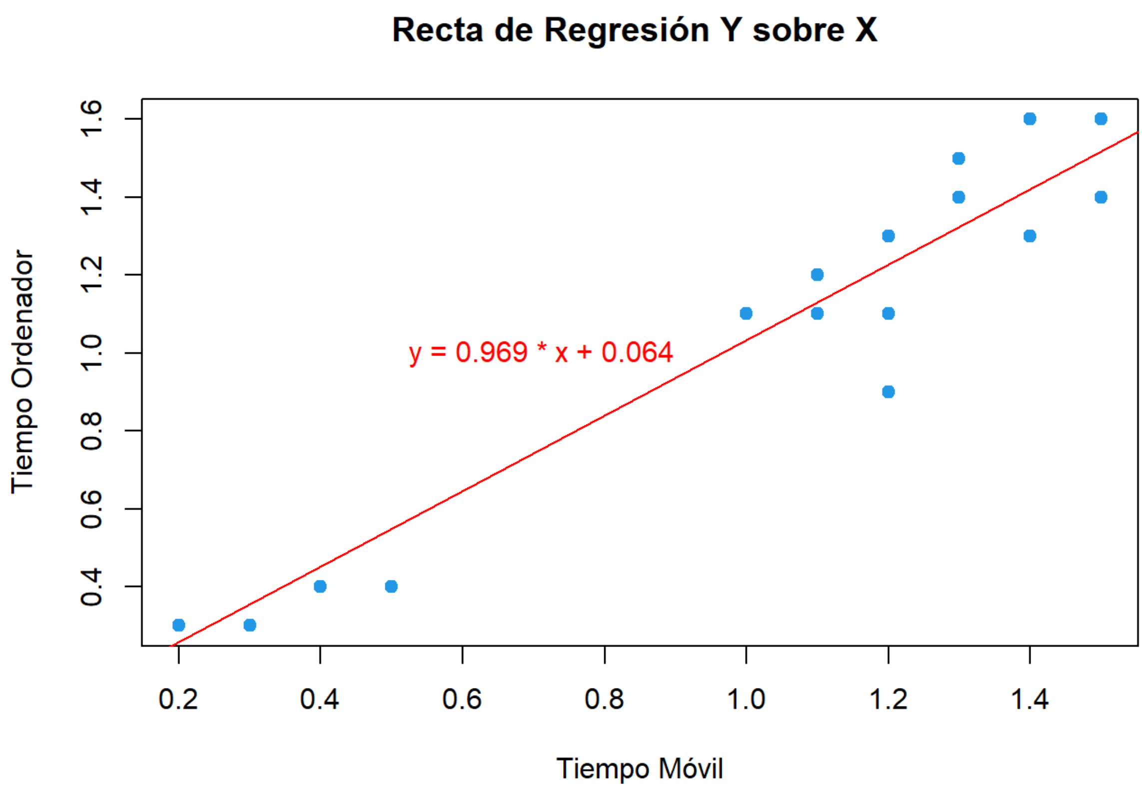
$$y = ax + b$$

```
#Recta de Regresión
a = Correlacion_XY
b = Media_Y -Correlacion_XY * Media_X

equation <- paste("y =", round(a,3), "* x +", round(b,3))
equation
```

```
## [1] "y = 0.969 * x + 0.064"
```

Entonces tenemos que la recta queda $y = 0.969x + 0.064$. Veamos su gráfica:

```
plot(X, Y,main = "Recta de Regresión Y sobre X ",
      xlab = "Tiempo Móvil",
      ylab = "Tiempo Ordenador",
      pch = 19, col = 4)
abline(b, a, col = "red")
text(0.5, 1, equation, pos = 4, col = "red")
```



Predicción y Causalidad

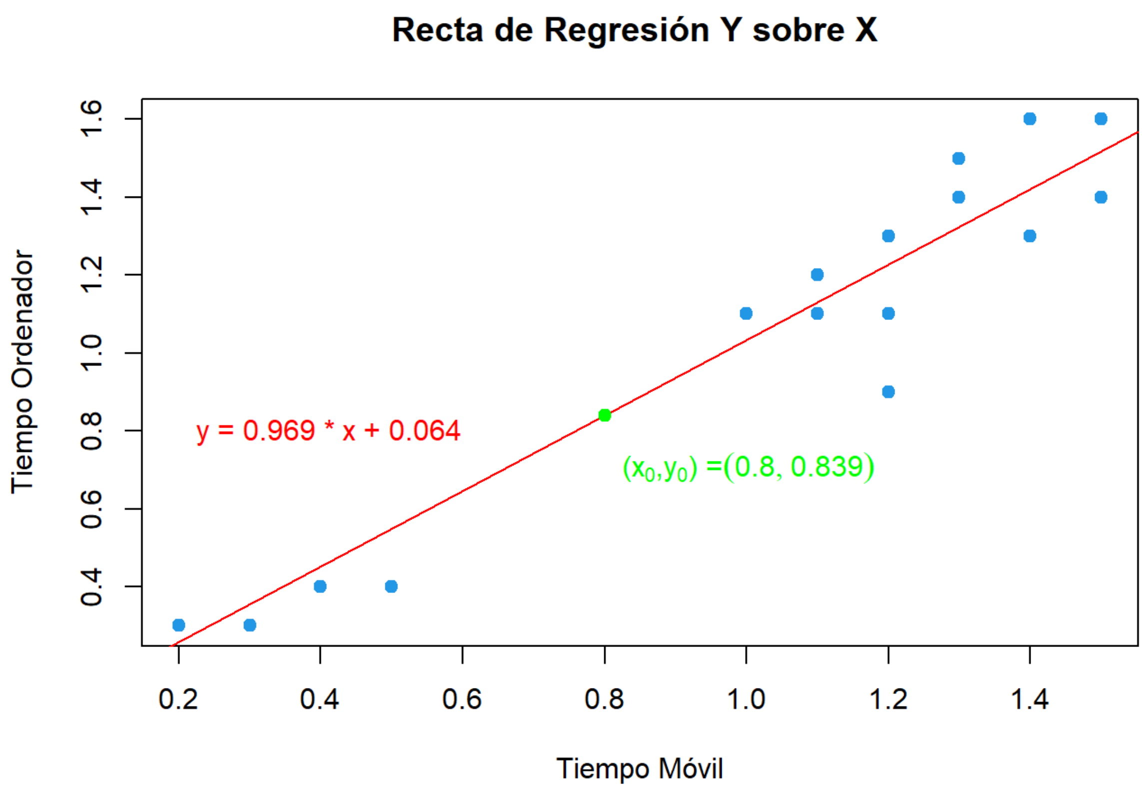
Queremos saber el tiempo de descarga en el ordenador, sabiendo que en móvil tardo 0.8 segundos. Es decir queremos hallar y_0 dado un $x = 0$ en la recta de regresión.

```
#Uso práctico de la recta de regresión
x_0 = 0.8
y_0 = a*x_0+b
cat( "(x_0, y_0) = (",x_0, ",", y_0,")")
```

```
## (x_0, y_0) = ( 0.8 , 0.839436 )
```

Luego $(x_0, y_0 = 0.8, 0.839$. Veamos como queda este punto en la gráfica:

```
#Uso práctico de la recta de regresion
plot(X, Y,main = "Recta de Regresión Y sobre X ",
      xlab = "Tiempo Móvil",
      ylab = "Tiempo Ordenador",
      pch = 19, col = 4)
abline(b, a, col = "red")
points(x_0, y_0, col = "green", pch = 19)
text(0.2, 0.8, equation, pos = 4, col = "red")
text(0.8, 0.7, expression("(x"[0]*",y"[0]*") =" (0.8,0.839)) , pos = 4, col = "green")
```



Coeficiente de Determianción

Recordemos:

$$R^2 = r^2 \qquad \text{con} \quad 0 \leq R^2 \leq 1$$

```
#Coeficiente de determinación
R = Correlacion_XY^2
cat(R*100,"%")
```

```
## 93.9904 %
```

Observamos que el modelo explica el 93.99% de la variabilidad de los datos. Por lo que consideramos que es un muy buen modelo de regresión.

Recta X sobre Y

Para calcular la otra recta de regresión usaremos:

$$x = \overline{x} + \frac{\sigma_{xy}}{\sigma_y}(y - \overline{y}) \quad \Rightarrow \quad y = \overline{y} + \frac{\sigma_y}{\sigma_{xy}}(x - \overline{x}) \quad \Rightarrow \quad y = \frac{\sigma_y}{\sigma_{xy}} x + (\overline{y} - \frac{\sigma_y}{\sigma_{xy}} \overline{x})$$

Sea:

$$a_x = \frac{\sigma_y}{\sigma_{xy}} \qquad b_x = (\overline{y} - \frac{\sigma_y}{\sigma_{xy}} \overline{x})$$

Tenemos:

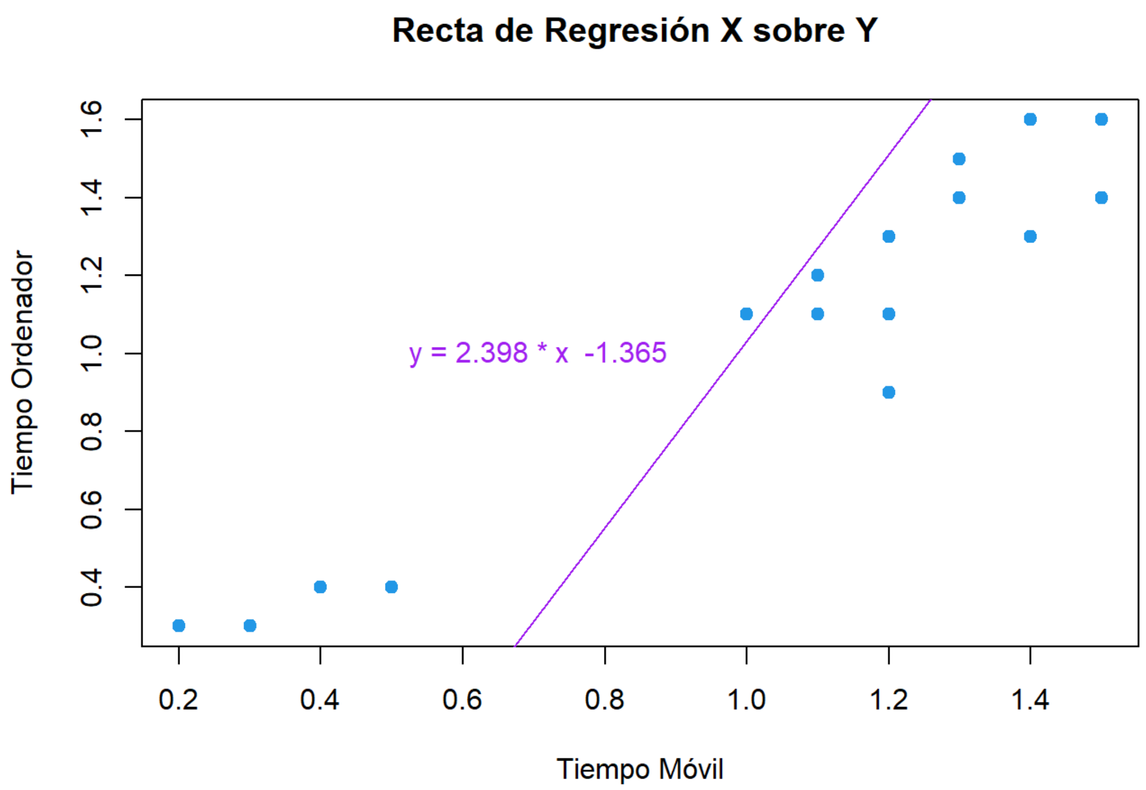
$$y = a_x X + b_x$$

```
#Recta regresión X sobre Y
a_x = sqrt(Varianza_Y)/Covarianza_XY
b_x = Media_Y - a_x * Media_X

equation_2 <- paste("y =", round(a_x,3), "* x ", round(b_x,3))
equation_2
```

```
## [1] "y = 2.398 * x  -1.365"
```

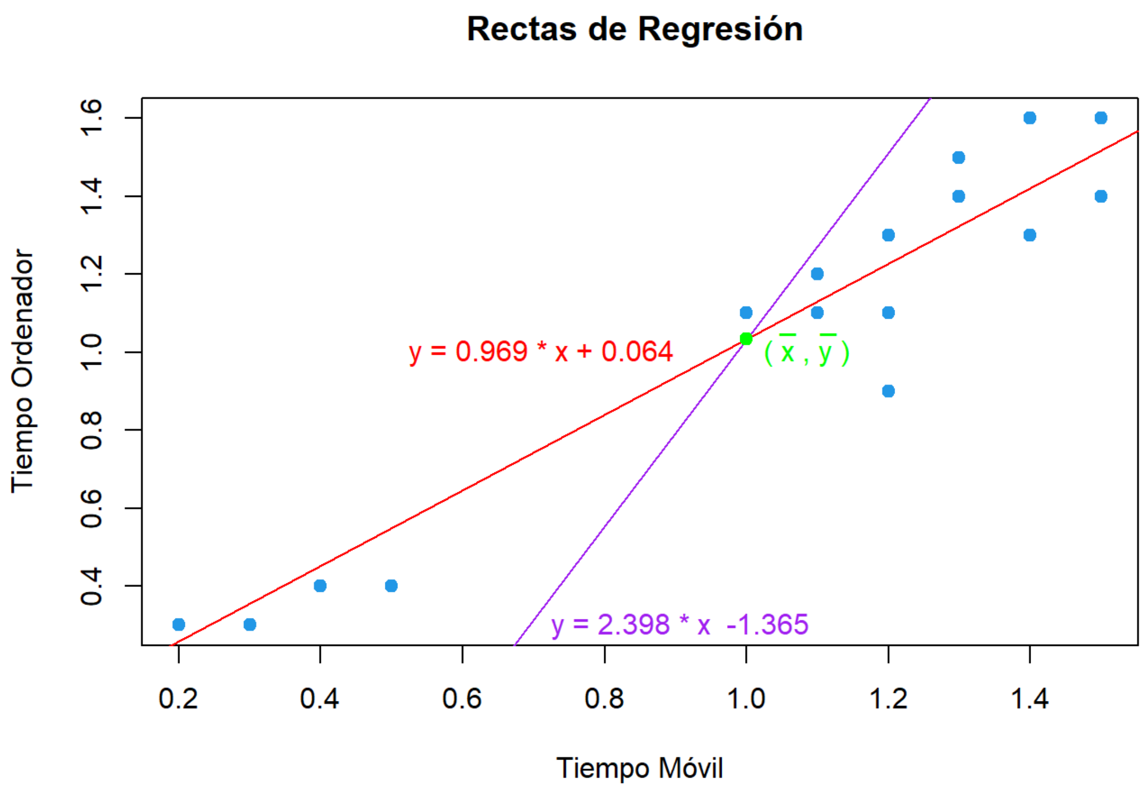
Entonces tenemos que la recta queda $y = 2.398x - 1.365$. Veamos su gráfica:

```
plot(X, Y,main = "Recta de Regresión X sobre Y ",
      xlab = "Tiempo Móvil",
      ylab = "Tiempo Ordenador",
      pch = 19, col = 4)
abline(b_x, a_x, col = "purple")
text(0.5, 1, equation_2, pos = 4, col = "purple")
```



Ahora veamos las dos rectas de regresión en la siguiente gráfica:

```
#Ambas rectas de regresión
plot(X, Y,main = "Rectas de Regresión ",
      xlab = "Tiempo Móvil",
      ylab = "Tiempo Ordenador",
      pch = 19, col = 4)
abline(b_x, a_x, col = "purple")
abline(b, a, col = "red")
points(Media_X, Media_Y, col = "green", pch = 19)
text(0.5, 1, equation, pos = 4, col = "red")
text(0.7, 0.3, equation_2, pos = 4, col = "purple")
text(1, 1, "( x , y )", pos = 4, col = "green")
text(1, 1.08, "  _   _ ", pos = 4, col = "green")
```



Como hemos visto, estas rectas se cortan por el centro de gravedad $(\overline{x}, \overline{y}) = (1, 1.03)$

Anexo IV. Comandos Básicos de Excel

Comandos básicos de Excel

- Sumas: para sumar valores de las celdas usaremos "SUMA". Soporta tanto celdas separadas como intervalos. Ejemplo: =**SUMA(A1:A50)**

- Restas: para restar los valores de dos celdas usaremos el símbolo de resta "−" entre ambas. Ejemplo: = **A2-A3**

- Multiplicaciones: para multiplicar los valores de dos celdas debemos intercalar entre ellas un asterisco ∗. Ejemplo: = **A1*A3* A5*A8**

- Divisiones: para dividir los valores de dos celdas debemos intercalar entre ellas el símbolo "/". Ejemplo: = **A2 / C2**

- Elevar al cuadrado: para elevar al cuadrado un valor se debemos incluir el símbolo "^2". Ejemplo: = **(A2)^2**

- Raíz Cuadrada: Para hacer la raíz cuadrada usaremos RAÍZ".
  Ejemplo: = **RAIZ(A2)**

Anexo V. Ejemplo Tablas de Doble Entrada

Ejemplos Excel Tablas de Doble Entrada

FRECUENCIAS ABSOLUTAS

X \ Y	Medicina A	Medicina B	Medicina C	Total (n_i)
Enfermedad A	55	76	89	220
Enfermedad B	76	64	1	141
Enfermedad C	42	3	40	85
Enfermedad D	28	75	41	144
Total (n_j)	201	218	171	590

FRECUENCIAS RELATIVAS

X \ Y	Medicina A	Medicina B	Medicina C	Total (n_i)
Enfermedad A	0,0932	0,1288	0,1508	0,3729
Enfermedad B	0,1288	0,1085	0,0017	0,2390
Enfermedad C	0,0712	0,0051	0,0678	0,1441
Enfermedad D	0,0475	0,1271	0,0695	0,2441
Total (n_j)	0,3407	0,3695	0,2898	1,0000

F{i}*F{j}

X \ Y	Medicina A	Medicina B	Medicina C
Enfermedad A	0,1270	0,1378	0,1081
Enfermedad B	0,0814	0,0883	0,0693
Enfermedad C	0,0491	0,0532	0,0418
Enfermedad D	0,0831	0,0902	0,0707

	A	B	C	D	E	F	G	H	I	J	K	L
1	Frecuencias Aboslutas											
2		X\Y	0	1	2	3	4	5				
3		0	12	18	19	16	10	6				
4		1	8	41	32	34	37	27				
5		2	15	26	62	37	36	20				
6		3	4	22	53	45	64	69				
7		4	3	19	48	59	30	28				
8												
9	Frecuencias Aboslutas											
10		X\Y	0	1	2	3	4	5	fi.			
11		0	12	18	19	16	10	6	81			
12		1	8	41	32	34	37	27	179			
13		2	15	26	62	37	36	20	196			
14		3	4	22	53	45	64	69	257			
15		4	3	19	48	59	30	28	187			
16		f.j	42	126	214	191	177	150	900			
17												
18												
19	Frecuencias Marginal X											
20												
21		X	fi.	fi. * X	X^2	fi. *X2						
22		0	81	0	0	0		Media X	2,32222222			
23		1	179	179	1	179		Var X	1,5717284			
24		2	196	392	4	784		Desv.Tip. X	1,25368592			
25		3	257	771	9	2313						
26		4	187	748	16	2992						
27		Sumatorios	900	2090		6268						
28												
29												
30	Frecuencias Marginal Y											
31												
32		Y	f.j	f.j *Y	Y^2	f.j * Y^2						
33		0	42	0	0	0		Media Y	2,87222222			
34		1	126	126	1	126		Var Y	2,06478395			
35		2	214	428	4	856		Desv.Tip. Y	1,43693561			
36		3	191	573	9	1719						
37		4	177	708	16	2832						
38		5	150	750	25	3750						
39		Sumatorios	900	2585		9283						
40												
41												
42	X*Y											
43		X\Y	0	1	2	3	4	5		Suma X*Y* fij	6300	
44		0	0	0	0	0	0	0		Covarianza_XY	0,330061728	
45		1	0	1	2	3	4	5		Correlacion_XY	0,183218411	
46		2	0	2	4	6	8	10				
47		3	0	3	6	9	12	15				
48		4	0	4	8	12	16	20				
49										Recta de regresion		
50	X*Y*fij									y=ax+b	a =coef. Cor.	0,18321841
51		X\Y	0	1	2	3	4	5			b= y_media-a*media_x	2,44674836
52		0	0	0	0	0	0	0				
53		1	0	41	64	102	148	135		Recta de regresion:	y = 0,183x+2,447	
54		2	0	52	248	222	288	200				
55		3	0	66	318	405	768	1035				
56		4	0	76	384	708	480	560				

Ejemplos Excel Tablas de Doble Entrada con Formulas

FRECUENCIAS ABSOLUTAS

X \ Y	Medicina A	Medicina B	Medicina C	Total (n_i)
Enfermedad A	55	76	89	=SUMA(C3:E3)
Enfermedad B	76	64	1	=SUMA(C4:E4)
Enfermedad C	42	3	40	=SUMA(C5:E5)
Enfermedad D	28	75	41	=SUMA(C6:E6)
Total (n_j)	=SUMA(C3:C6)	=SUMA(D3:D6)	=SUMA(E3:E6)	=SUMA(C7:E7)

FRECUENCIAS RELATIVAS

X \ Y	Medicina A	Medicina B	Medicina C	Total (n_i)
Enfermedad A	=C3/590	=D3/590	=E3/590	=SUMA(C11:E11)
Enfermedad B	=C4/590	=D4/590	=E4/590	=SUMA(C12:E12)
Enfermedad C	=C5/590	=D5/590	=E5/590	=SUMA(C13:E13)
Enfermedad D	=C6/590	=D6/590	=E6/590	=SUMA(C14:E14)
Total (n_j)	=SUMA(C11:C14)	=SUMA(D11:D14)	=SUMA(E11:E14)	=SUMA(C15:E15)

F{i}*F{j}

X \ Y	Medicina A	Medicina B	Medicina C
Enfermedad A	=$F11*C$15	=$F11*D15	=$F11*E15
Enfermedad B	=$F12*C$15	=$F12*D$15	=$F12*E$15
Enfermedad C	=$F13*C$15	=$F13*D$15	=$F13*E$15
Enfermedad D	=$F14*C$15	=$F14*D$15	=$F14*E$15

	A	B	C	D	E	F	G	H	I	J	K	L
1	bla de Frecuencias Aboslut											
2		X\Y	0	1	2	3	4	5				
3	0		12	18	19	16	10	6				
4	1		8	41	32	34	37	27				
5	2		15	26	62	37	36	20				
6	3		4	22	53	45	64	69				
7	4		3	19	48	59	30	28				
8												
9	bla de Frecuencias Aboslut											
10		X\Y	0	1	2	3	4	5	fi.			
11	0		=C3	=D3	=E3	=F3	=G3	=H3	=SUMA(C11:H11)			
12	1		=C4	=D4	=E4	=F4	=G4	=H4	=SUMA(C12:H12)			
13	2		=C5	=D5	=E5	=F5	=G5	=H5	=SUMA(C13:H13)			
14	3		=C6	=D6	=E6	=F6	=G6	=H6	=SUMA(C14:H14)			
15	4		=C7	=D7	=E7	=F7	=G7	=H7	=SUMA(C15:H15)			
16		f.j	=SUMA(C11:C15)	=SUMA(D11:D15)	=SUMA(E11:E15)	=SUMA(F11:F15)	=SUMA(G11:G15)	=SUMA(H11:H15)	=SUMA(C16:H16)			
17												
18												
19	bla de Frecuencias Margina											
20												
21		X	fi.	fi. * X	X^2	fi. *X2						
22	0		=I11	=B22*C22	=B22^2	=C22*E22		Media X	=D27/C27			
23	1		=I12	=B23*C23	=B23^2	=C23*E23		Var X	=(F27/C27)-I22^2			
24	2		=I13	=B24*C24	=B24^2	=C24*E24		Desv.Tip. X	=RAIZ(I23)			
25	3		=I14	=B25*C25	=B25^2	=C25*E25						
26	4		=I15	=B26*C26	=B26^2	=C26*E26						
27		Sumatorios	=SUMA(C22:C26)	=SUMA(D22:D26)		=SUMA(F22:F26)						
28												
29												
30	bla de Frecuencias Margina											
31												
32		Y	f.j	f.j *Y	Y^2	f.j * Y^2						
33	0		=C16	=B33*C33	=B33^2	=C33*E33		Media Y	=D39/C39			
34	1		=D16	=B34*C34	=B34^2	=C34*E34		Var Y	=(F39/C39)-I33^2			
35	2		=E16	=B35*C35	=B35^2	=C35*E35		Desv.Tip. Y	=RAIZ(I34)			
36	3		=F16	=B36*C36	=B36^2	=C36*E36						
37	4		177	=B37*C37	=B37^2	=C37*E37						
38	5		150	=B38*C38	=B38^2	=C38*E38						
39		Sumatorios	=SUMA(C33:C38)	=SUMA(D33:D38)		=SUMA(F33:F38)						
40												
41												
42	X*Y											
43		X\Y	0	1	2	3	4	5		Suma X*Y* fij	=SUMA(C52:H56)	
44	0		=$B44*C$43	=$B44*D$43	=$B44*E$43	=$B44*F$43	=$B44*G$43	=$B44*H$43		Covarianza_XY	=(K43/I16)-(I22*I33)	
45	1		=$B45*C$43	=$B45*D$43	=$B45*E$43	=$B45*F$43	=$B45*G$43	=$B45*H$43		Correlacion_XY	=K44/(I35*I24)	
46	2		=$B46*C$43	=$B46*D$43	=$B46*E$43	=$B46*F$43	=$B46*G$43	=$B46*H$43				
47	3		=$B47*C$43	=$B47*D$43	=$B47*E$43	=$B47*F$43	=$B47*G$43	=$B47*H$43				
48	4		=$B48*C$43	=$B48*D$43	=$B48*E$43	=$B48*F$43	=$B48*G$43	=$B48*H$43				
49										Recta de regresion		
50	X*Y*fij									y=ax+b	a =coef. Cor.	=K45
51		X\Y	0	1	2	3	4	5			b= y_media-a*media_x	=I33-I22*L50
52	0		=C44*C3	=D44*D3	=E44*E3	=F44*F3	=G44*G3	=H44*H3				
53	1		=C45*C4	=D45*D4	=E45*E4	=F45*F4	=G45*G4	=H45*H4		Recta de regresion:	y = 0,183x+2,447	
54	2		=C46*C5	=D46*D5	=E46*E5	=F46*F5	=G46*G5	=H46*H5				
55	3		=C47*C6	=D47*D6	=E47*E6	=F47*F6	=G47*G6	=H47*H6				
56	4		=C48*C7	=D48*D7	=E48*E7	=F48*F7	=G48*G7	=H48*H7				

Anexo VI. Tabla de Competencias Especificas, Criterios de Evaluación y Descriptores Operativos

Estadística Bidimensional de Matemáticas Aplicadas a las Ciencias Sociales I de 1º Bachillerato con el programa R: Competencias específicas, criterios de evaluación y descriptores operativos, según Anexo II del Real Decreto 243/2022, del 5 de abril, del Boletín Oficial del Estado (2022)

Competencias específicas	**Criterios de evaluación**	**Descriptores operativos**
1. Modelizar y resolver problemas de la vida cotidiana y de las ciencias sociales aplicando diferentes estrategias y formas de razonamiento para obtener posibles soluciones.	1.1. Emplear algunas estrategias y herramientas, incluidas las digitales, en la resolución de problemas de la vida cotidiana y de las ciencias sociales, valorando su eficiencia en cada caso. 1.2. Obtener todas las posibles soluciones matemáticas de problemas de la vida cotidiana y de las ciencias sociales, describiendo el procedimiento realizado.	STEM1, STEM2, STEM3, CD2, CD5, CPSAA4, CPSAA5, CE3.
2. Verificar la validez de las posibles soluciones de un problema empleando el razonamiento y la argumentación para contrastar su idoneidad.	2.1. Comprobar la validez matemática de las posibles soluciones de un problema, utilizando el razonamiento y la argumentación.	STEM1, STEM2, CD3, CPSAA4, CC3, CE3.

	2.2. Seleccionar la solución más adecuada de un problema en función del contexto (de sostenibilidad, de consumo responsable, equidad...), usando el razonamiento y la argumentación.	
3. Formular o investigar conjeturas o problemas, utilizando el razonamiento, la argumentación, la creatividad y el uso de herramientas tecnológicas, para generar nuevo conocimiento matemático.	3.1. Adquirir nuevo conocimiento matemático mediante la formulación de conjeturas y problemas de forma guiada. 3.2. Emplear herramientas tecnológicas adecuadas en la formulación o investigación de conjeturas o problemas.	CCL1, STEM1, STEM2, CD1, CD2, CD3, CD5, CE3.
4. Utilizar el pensamiento computacional de forma eficaz, modificando, creando y generalizando algoritmos que resuelvan problemas mediante el uso de las matemáticas, para modelizar y resolver situaciones de la vida cotidiana y del ámbito de las ciencias sociales.	4.1. Interpretar, modelizar y resolver situaciones problematizadas de la vida cotidiana y de las ciencias sociales, utilizando el pensamiento computacional, modificando y creando algoritmos.	STEM1, STEM2, STEM3, CD2, CD3, CD5, CE3.

5. Establecer, investigar y utilizar conexiones entre las diferentes ideas matemáticas estableciendo vínculos entre conceptos, procedimientos, argumentos y modelos para dar significado y estructurar el aprendizaje matemático.	5.1. Manifestar una visión matemática integrada, investigando y conectando las diferentes ideas matemáticas. 5.2. Resolver problemas, estableciendo y aplicando conexiones entre las diferentes ideas matemáticas.	STEM1, STEM3, CD2, CD3, CCEC1.
6. Descubrir los vínculos de las matemáticas con otras áreas de conocimiento y profundizar en sus conexiones, interrelacionando conceptos y procedimientos, para modelizar, resolver problemas y desarrollar la capacidad crítica, creativa e innovadora en situaciones diversas.	6.1. Resolver problemas en situaciones diversas, utilizando procesos matemáticos, estableciendo y aplicando conexiones entre el mundo real, otras áreas de conocimiento y las matemáticas.	STEM1, STEM2, CD2, CPSAA5, CC4, CE2, CE3, CCEC1.

	6.2. Analizar la aportación de las matemáticas al progreso de la humanidad reflexionando sobre su contribución en la propuesta de soluciones a situaciones complejas y a los retos en las ciencias sociales que se planteen.	
7.Representar conceptos, procedimientos e información matemáticos seleccionando diferentes tecnologías, para visualizar ideas y estructurar razonamientos matemáticos.	7.1. Representar ideas matemáticas, estructurando diferentes razonamientos matemáticos y seleccionando las tecnologías más adecuadas. 7.2. Seleccionar y utilizar diversas formas de representación, valorando su utilidad para compartir información.	STEM3, CD1, CD2, CD5, CE3, CCEC4.1, CCEC4.2.
8. Comunicar las ideas matemáticas, de forma individual y colectiva, empleando el soporte, la terminología y el rigor apropiados, para organizar y consolidar el pensamiento matemático.	8.1. Mostrar organización al comunicar las ideas matemáticas, empleando el soporte, la terminología y el rigor apropiados.	CCL1, CCL3, CP1, STEM2, STEM4, CD2, CD3, CCEC3.2.

	8.2. Reconocer y emplear el lenguaje matemático en diferentes contextos, comunicando la información con precisión y rigor.	
9. Utilizar destrezas personales y sociales, identificando y gestionando las propias emociones, respetando las de los demás y organizando activamente el trabajo en equipos heterogéneos, aprendiendo del error como parte del proceso de aprendizaje y afrontando situaciones de incertidumbre, para perseverar en la consecución de objetivos en el aprendizaje de las matemáticas.	9.1. Afrontar las situaciones de incertidumbre, identificando y gestionando emociones y aceptando y aprendiendo del error como parte del proceso de aprendizaje de las matemáticas. 9.2. Mostrar una actitud positiva y perseverante, aceptando y aprendiendo de la crítica razonada al hacer frente a las diferentes situaciones de aprendizaje de las matemáticas.	CP3, STEM5, CPSAA1.1, CPSAA1.2, CPSAA3.1, CPSAA3.2, CC2, CC3, CE2.

Tabla 19

Tabla relaciona las competencia específicas, criterios de evaluación y descriptores operativos

Anexo VII. Saberes Básicos

Saberes básicos.

Estadística Bidimensional de Matemáticas Aplicadas a las Ciencias Sociales I de 1º Bachillerato con el programa R: Saberes básicos según Anexo II del Real Decreto 243/2022, del 5 de abril, del Boletín Oficial del Estado (2022).

A. Sentido numérico.

1. Conteo.

1.1.Estrategias y técnicas de recuento sistemático (diagramas de árbol, técnicas de combinatoria...).

2. Cantidad.

2.1. Números reales (racionales e irracionales): comparación, ordenación, clasificación y contraste de sus propiedades.

3. Sentido de las operaciones.

3.1. Potencias, raíces y logaritmos: comprensión y utilización de sus relaciones para simplificar y resolver problemas.

4. Educación financiera.

4.1. Resolución de problemas relacionados con la educación financiera (cuotas, tasas, intereses, préstamos. . .) con herramientas tecnológicas.

B. Sentido de la medida.

1. Medición.

1.1.La probabilidad como medida de la incertidumbre asociada a fenómenos aleatorios.

2. Cambio.

2.1. Límites: estimación y cálculo a partir de una tabla, un gráfico o una expresión algebraica.

2.2. Continuidad de funciones: aplicación de límites en el estudio de la continuidad.

2.3. Derivada de una función: definición a partir del estudio del cambio en contextos de las ciencias sociales.

C. Sentido algebraico.

1. Patrones.

1.1. Generalización de patrones en situaciones sencillas.

2. Modelo matemático.

2.1. Relaciones cuantitativas esenciales en situaciones sencillas: estrategias de identificación y determinación de la clase o clases de funciones que pueden modelizarlas.

2.2. Ecuaciones, inecuaciones y sistemas: modelización de situaciones de las ciencias sociales y de la vida real.

3. Igualdad y desigualdad.

3.1.Resolución de ecuaciones, inecuaciones y sistemas de ecuaciones e inecuaciones no lineales en diferentes contextos.

4. Relaciones y funciones.

4.1. Representación gráfica de funciones utilizando la expresión más adecuada.

4.2. Propiedades de las distintas clases de funciones, incluyendo, polinómica, exponencial, racional sencilla, irracional, logarítmica, periódica y a trozos: comprensión y comparación.

4.3 Álgebra simbólica en la representación y explicación de relaciones matemáticas de las ciencias sociales.

5. Pensamiento computacional.

5.1. Formulación, resolución y análisis de problemas de la vida cotidiana y de las ciencias sociales utilizando programas y herramientas adecuados.

5.2. Comparación de algoritmos alternativos para el mismo problema mediante el razonamiento lógico.

D. Sentido estocástico.

1. Organización y análisis de datos.

1.1. Organización de los datos procedentes de variables bidimensionales: distribución conjunta y distribuciones marginales y condicionadas. Análisis de la dependencia estadística.

1.2. Estudio de la relación entre dos variables mediante la regresión lineal y cuadrática: valoración gráfica de la pertinencia del ajuste. Diferencia entre correlación y causalidad.

1.3. Coeficientes de correlación lineal y de determinación: cuantificación de la relación lineal, predicción y valoración de su fiabilidad en contextos de las ciencias sociales.

1.4. Calculadora, hoja de cálculo o software específico en el análisis de datos estadísticos.

2. Incertidumbre.

2.1. Estimación de la probabilidad a partir del concepto de frecuencia relativa.

2.2. Cálculo de probabilidades en experimentos simples: la regla de Laplace en situaciones de equiprobabilidad y en combinación con diferentes técnicas de recuento.

3. Distribuciones de probabilidad.

3.1. Variables aleatorias discretas y continuas. Parámetros de la distribución.

3.2. Modelización de fenómenos estocásticos mediante las distribuciones de probabilidad binomial y normal. Cálculo de probabilidades asociadas mediante herramientas tecnológicas.

3.3. Estimación de probabilidades mediante la aproximación de la binomial por la normal.

4. Inferencia.

4.1. Diseño de estudios estadísticos relacionados con las ciencias sociales utilizando herramientas digitales. Técnicas de muestreo sencillas.

4.2. Análisis de muestras unidimensionales y bidimensionales con herramientas tecnológicas con el fin de emitir juicios y tomar decisiones: estimación puntual.

E. Sentido socioafectivo.

1. Creencias, actitudes y emociones.

1.1. Destrezas de autoconciencia encaminadas a reconocer emociones propias, afrontando eventuales situaciones de estrés y ansiedad en el aprendizaje de las matemáticas.

1.2. Tratamiento del error, individual y colectivo como elemento movilizador de saberes previos adquiridos y generador de oportunidades de aprendizaje en el aula de matemáticas.

2. Trabajo en equipo y toma de decisiones.

2.1. Reconocimiento y aceptación de diversos planteamientos en la resolución de problemas y tareas matemáticas, transformando los enfoques de los demás en nuevas y mejoradas estrategias propias, mostrando empatía y respeto en el proceso.

2.2. Técnicas y estrategias de trabajo en equipo para la resolución de problemas y tareas matemáticas, en grupos heterogéneos.

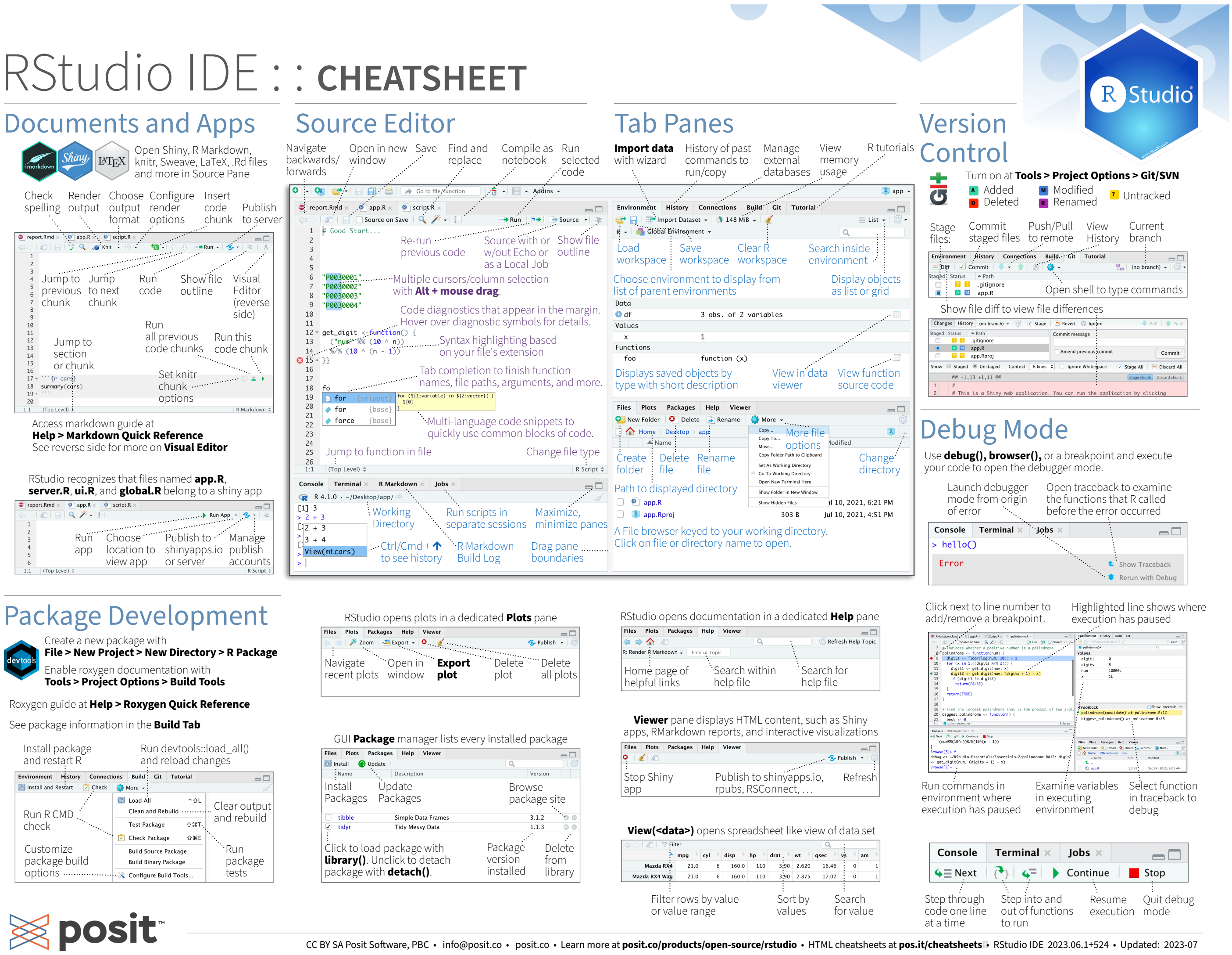
3. Inclusión, respeto y diversidad.

3.1. Destrezas para desarrollar una comunicación efectiva: la escucha activa, la formulación de preguntas o solicitud y prestación de ayuda cuando sea necesario.

3.2. Valoración de la contribución de las matemáticas y el papel de matemáticos y matemáticas a lo largo de la historia en el avance de las ciencias sociales.

Anexo VIII. Guía Rápida de los Paneles de RStudio.

Una cheatsheet que nos ofrece información sobre los paneles de RStudio creada por CC BY SA Posit Software, PBC (2023).



RStudio IDE : : CHEATSHEET

Documents and Apps

Open Shiny, R Markdown, knitr, Sweave, LaTeX, .Rd files and more in Source Pane

Check spelling | Render output | Choose output format | Configure render options | Insert code chunk | Publish to server

Jump to previous chunk | Jump to next chunk | Run code | Show file outline | Visual Editor (reverse side)

Jump to section or chunk | Run all previous code chunks | Run this code chunk | Set knitr chunk options

Access markdown guide at **Help > Markdown Quick Reference**

See reverse side for more on **Visual Editor**

RStudio recognizes that files named **app.R**, **server.R**, **ui.R**, and **global.R** belong to a shiny app

Run app | Choose location to view app | Publish to shinyapps.io or server | Manage publish accounts

Source Editor

Navigate backwards/forwards | Open in new window | Save | Find and replace | Compile as notebook | Run selected code

Re-run previous code | Source with or w/out Echo or as a Local Job | Show file outline

Multiple cursors/column selection with **Alt + mouse drag**.

Code diagnostics that appear in the margin. Hover over diagnostic symbols for details.

Syntax highlighting based on your file's extension

Tab completion to finish function names, file paths, arguments, and more.

Multi-language code snippets to quickly use common blocks of code.

Jump to function in file | Change file type

Working Directory | Run scripts in separate sessions | Maximize, minimize panes

Ctrl/Cmd + ↑ to see history | R Markdown Build Log | Drag pane boundaries

Tab Panes

Import data with wizard | History of past commands to run/copy | Manage external databases | View memory usage | R tutorials

Load workspace | Save workspace | Clear R workspace | Search inside environment

Choose environment to display from list of parent environments | Display objects as list or grid

Displays saved objects by type with short description | View in data viewer | View function source code

Create folder | Delete file | Rename file | More file options | Change directory

Path to displayed directory

A File browser keyed to your working directory. Click on file or directory name to open.

Version Control

Turn on at **Tools > Project Options > Git/SVN**

A Added | M Modified | ? Untracked | D Deleted | R Renamed

Stage files: | Commit staged files | Push/Pull to remote | View History | Current branch

Open shell to type commands

Show file diff to view file differences

Debug Mode

Use **debug()**, **browser()**, or a breakpoint and execute your code to open the debugger mode.

Launch debugger mode from origin of error | Open traceback to examine the functions that R called before the error occurred

Click next to line number to add/remove a breakpoint. | Highlighted line shows where execution has paused

Run commands in environment where execution has paused | Examine variables in executing environment | Select function in traceback to debug

Step through code one line at a time | Step into and out of functions to run | Resume execution | Quit debug mode

Package Development

Create a new package with **File > New Project > New Directory > R Package**

Enable roxygen documentation with **Tools > Project Options > Build Tools**

Roxygen guide at **Help > Roxygen Quick Reference**

See package information in the **Build Tab**

Install package and restart R | Run devtools::load_all() and reload changes

Run R CMD check | Clear output and rebuild

Customize package build options | Run package tests

RStudio opens plots in a dedicated **Plots** pane

Navigate recent plots | Open in window | **Export plot** | Delete plot | Delete all plots

GUI **Package** manager lists every installed package

Install Packages | Update Packages | Browse package site

Click to load package with **library()**. Unclick to detach package with **detach()**. | Package version installed | Delete from library

RStudio opens documentation in a dedicated **Help** pane

Home page of helpful links | Search within help file | Search for help file

Viewer pane displays HTML content, such as Shiny apps, RMarkdown reports, and interactive visualizations

Stop Shiny app | Publish to shinyapps.io, rpubs, RSConnect, … | Refresh

View(<data>) opens spreadsheet like view of data set

Filter rows by value or value range | Sort by values | Search for value

posit

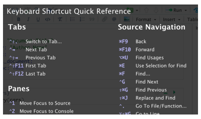
Keyboard Shortcuts

RUN CODE	Windows/Linux	Mac
Search command history	Ctrl+↑	Cmd+↑
Interrupt current command	Esc	Esc
Clear console	Ctrl+L	Ctrl+L
NAVIGATE CODE		
Go to File/Function	Ctrl+.	Ctrl+.
WRITE CODE		
Attempt completion	Tab or Ctrl+Space	Tab or Ctrl+Space
Insert <− (assignment operator)	Alt+-	Option+-
Insert |> or %>% (pipe operator)	Ctrl+Shift+M	Cmd+Shift+M
(Un)Comment selection	Ctrl+Shift+C	Cmd+Shift+C

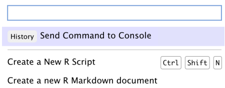
MAKE PACKAGES	Windows/Linux	Mac
Load All (devtools)	Ctrl+Shift+L	Cmd+Shift+L
Test Package (Desktop)	Ctrl+Shift+T	Cmd+Shift+T
Document Package	Ctrl+Shift+D	Cmd+Shift+D

DOCUMENTS AND APPS		
Knit Document (knitr)	Ctrl+Shift+K	Cmd+Shift+K
Insert chunk (Sweave & Knitr)	Ctrl+Alt+I	Cmd+Option+I
Run from start to current line	Ctrl+Alt+B	Cmd+Option+B
MORE KEYBOARD SHORTCUTS		
Keyboard Shortcuts Help	Alt+Shift+K	Option+Shift+K
Show Command Palette	Ctrl+Shift+P	Cmd+Shift+P

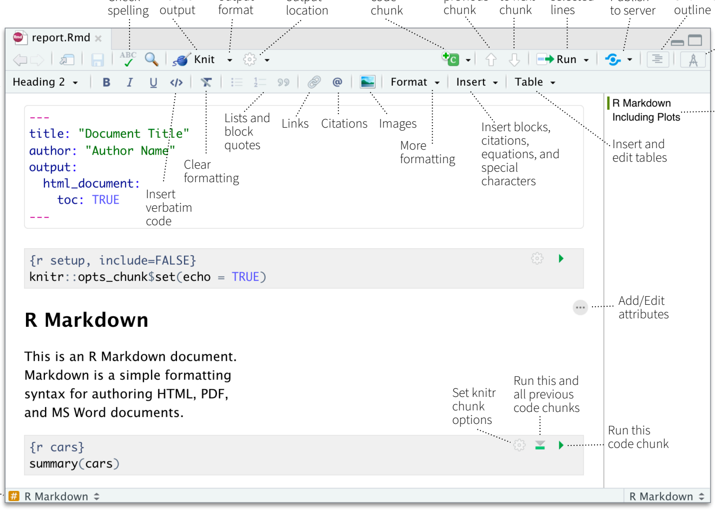
View the Keyboard Shortcut Quick Reference with **Tools > Keyboard Shortcuts** or **Alt/Option + Shift + K**



Search for keyboard shortcuts with **Tools > Show Command Palette** or **Ctrl/Cmd + Shift + P.**



Visual Editor



RStudio Workbench

WHY RSTUDIO WORKBENCH?

Extend the open source server with a commercial license, support, and more:

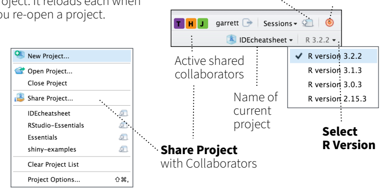
- open and run multiple R sessions at once
- tune your resources to improve performance
- administrative tools for managing user sessions
- collaborate real-time with others in shared projects
- switch easily from one version of R to a different version
- integrate with your authentication, authorization, and audit practices
- work in the RStudio IDE, JupyterLab, Jupyter Notebooks, or VS Code

Download a free 45 day evaluation at **www.rstudio.com/products/workbench/evaluation/**

Share Projects

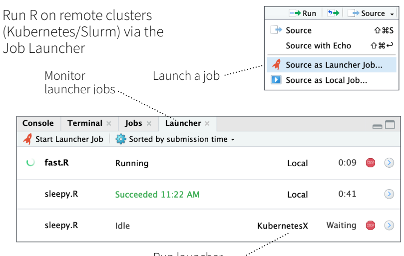
File > New Project

RStudio saves the call history, workspace, and working directory associated with a project. It reloads each when you re-open a project.



Run Remote Jobs

Run R on remote clusters (Kubernetes/Slurm) via the Job Launcher



posit