Matemáticas y Estadística Farmacia

Alirio Cruz Irene García

2024-09-11

Tabla de contenidos

Presentación

Esto es una edición en línea de los apuntes de la asignatura de Matemáticas y Estadística.

El enfoque es teórico-práctico para el grado de Farmacia de la UIB.

1 Matemáticas básicas

En este primer tema, empezaremos por instalar el software necesario para el análisis estadístico y la visualización de datos, y luego revisaremos los conceptos básicos de funciones lineales, exponenciales, logarítmicas y polinómicas, que son fundamentales para modelar y analizar fenómenos farmacéuticos. Todo estos conocimientos son cruciales para realizar cálculos precisos de dosis, desarrollar y evaluar nuevos medicamentos, así como para llevar a cabo estudios de bioequivalencia y análisis de farmacoeconomía.

1.1 R y Jamovi

Es fundamental utilizar un software especializado que simplifique la resolución de problemas y optimice el análisis estadístico de datos. Estas herramientas no solo agilizan los cálculos y procesamientos complejos, sino que también mejoran la precisión y fiabilidad de los resultados obtenidos. Además, permiten una redacción más clara y estructurada de informes, facilitando la presentación de resultados de manera profesional y efectiva.

En este curso utilizaremos el software `jamovi` (interfaz gráfica que emplea como base R). `R` es un entorno de programación estadística de código abierto que ofrece una amplia gama de funciones y paquetes para el análisis de datos y la generación de gráficos. `Jamovi` es una herramienta más accesible que nos permitirá centrarnos en el análisis y la interpretación de datos, en lugar de en la sintaxis de programación.

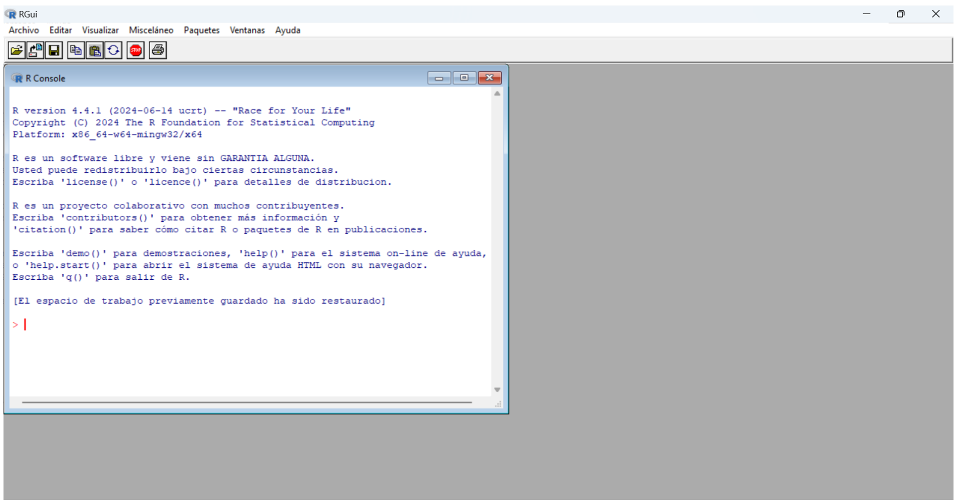
Instalar R es muy sencillo; pero es conveniente que dispongáis de su versión más reciente y que regularmente lo pongáis al día. Los pasos a realizar en Windows o Mac OS X para instalar su última versión son los siguientes:

- Si sois usuarios de Windows, acceded a la página web de la *CRAN* y pulsad sobre el enlace *Download R for Windows*. A continuación, entrad en el enlace *base*, descargad R y seguid las instrucciones de instalación del documento *Installation and other instructions* que encontraréis en esa misma página.
- Si sois usuarios de Mac OS X, acceded a la página web de la *CRAN* y pulsad sobre el enlace *Download R for Mac OS X*. A continuación, descargad el fichero `.pkg` correspondiente y, una vez descargado, abridlo y seguid las instrucciones del Asistente de Instalación.

- Si trabajáis con Ubuntu o Debian, para instalar la última versión de R basta que ejecutéis en una terminal, estando conectados a Internet, la siguiente instrucción:

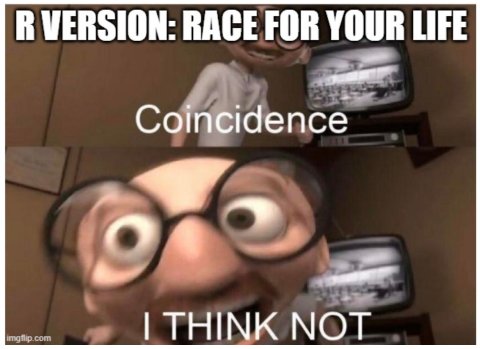
```
sudo aptitude install r-base
```

Cuando instaláis R para Windows o Mac OS X, con él también se os instala una interfaz gráfica que se abrirá al abrir la aplicación y en la que podréis trabajar.



La instalación para Linux no lleva una interfaz por defecto, así que sus usuarios tienen que trabajar con R en la terminal (ejecutando R para iniciar una sesión) o instalar aparte una interfaz. Independientemente de todas estas posibilidades, en este curso usaremos *RStudio* como interfaz gráfica de usuario de R para todos los sistemas operativos.

> **Nota**
>
> R version 4.4.1 (2024-06-14 ucrt) – “Race for Your Life” Copyright (C) 2024 The R Foundation for Statistical Computing Platform: x86_64-w64-mingw32/x64
> R es un software libre y viene sin GARANTIA ALGUNA. Usted puede redistribuirlo bajo ciertas circunstancias. Escriba ‘license()’ o ‘licence()’ para detalles de distribucion.
> R es un proyecto colaborativo con muchos contribuyentes. Escriba ‘contributors()’ para obtener más información y ‘citation()’ para saber cómo citar R o paquetes de R en publicaciones.
> Escriba ‘demo()’ para demostraciones, ‘help()’ para el sistema on-line de ayuda, o ‘help.start()’ para abrir el sistema de ayuda HTML con su navegador. Escriba ‘q()’ para salir de R.



Para que nuestra interfaz con R sea agradable, podemos usar varias aplicaciones disponibles, como `Rstudio`, `Visual Studio Code` o `Jamovi`. Por facilidad usaremos `Jamovi`

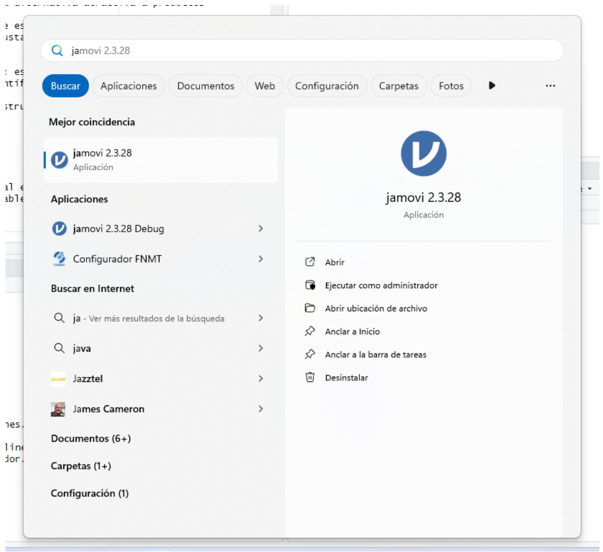
1.1.1 Instalación de Jamovi

Según la pagina de Jamovi, y en una traducción al castellano usando Google translate obtenemos:

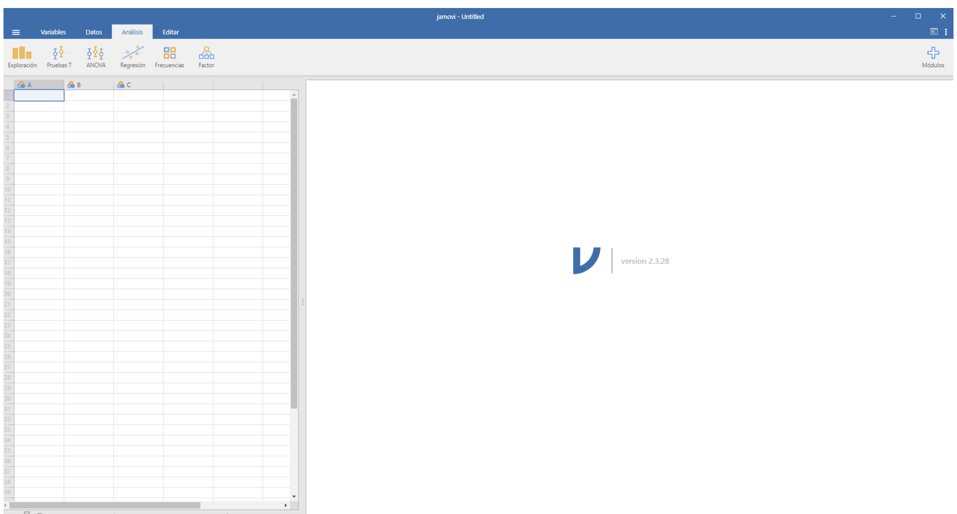
- Estadísticas simplificadas: Jamovi es una nueva hoja de cálculo estadística de "tercera generación". Diseñada desde cero para que sea fácil de usar, Jamovi es una alternativa atractiva a productos estadísticos costosos como SPSS y SAS.
- Integración con R: Jamovi está construido sobre el lenguaje estadístico R, lo que le brinda acceso a lo mejor que la comunidad estadística tiene para ofrecer. ¿Le gustaría el código R para sus análisis? Jamovi también puede proporcionárselo.
- Gratuito y abierto: Jamovi siempre será gratuito y abierto: ese es uno de nuestros valores fundamentales, porque Jamovi está hecho por la comunidad científica, para la comunidad científica.

Adicional a esto, instalaremos el software, siguiendo las instrucciones que aparecen en el botón **Download and install jamovi onto your computer**

Luego de la instalación debe aparecer en su busqueda de windows.



Y nos aparecerá esta interfaz para trabajar.



1.2 Matemáticas básicas

1.2.1 Funciones lineales

En geometría analítica y álgebra elemental, una función lineal es una función polinómica de primer grado, es decir, una función de una variable (normalmente esta variable se denota con

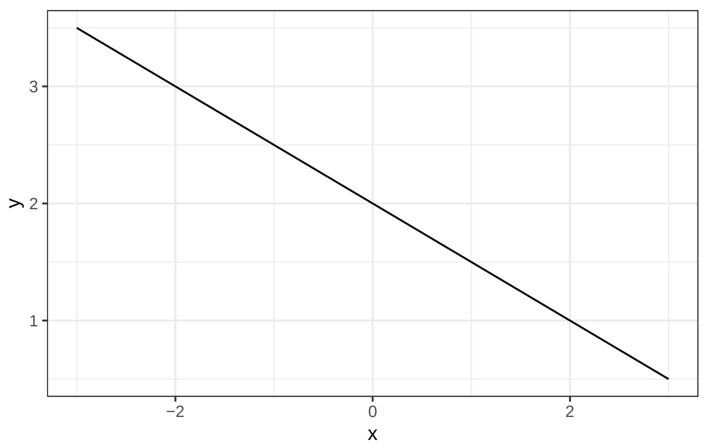
x, que puede ser escrita como la suma de términos de la forma

$$f(x) = mx + b$$

donde m determina la pendiente o inclinación de la recta, y la constante b determina el punto de corte de la recta con el eje vertical y.

En Farmacia es útil observar la relación entre dosis y respuesta:

```
library(tidyverse)
x = seq(-3,3)
y = 2-0.5*x
datos = data.frame(x, y)
datos %>% ggplot(aes(x=x, y=y))+
  geom_line()+theme_bw()
```



1.2.2 Funciones exponenciales

Una función exponencial es una función matemática de la forma

$$f(x) = a \cdot b^x$$

,

donde:

- a es una constante que representa el valor inicial.
- b es la base de la función exponencial.
- x es la variable independiente.

Estas funciones son comunes en situaciones de crecimiento o decrecimiento rápido, como el crecimiento poblacional, la desintegración radiactiva, y el interés compuesto.

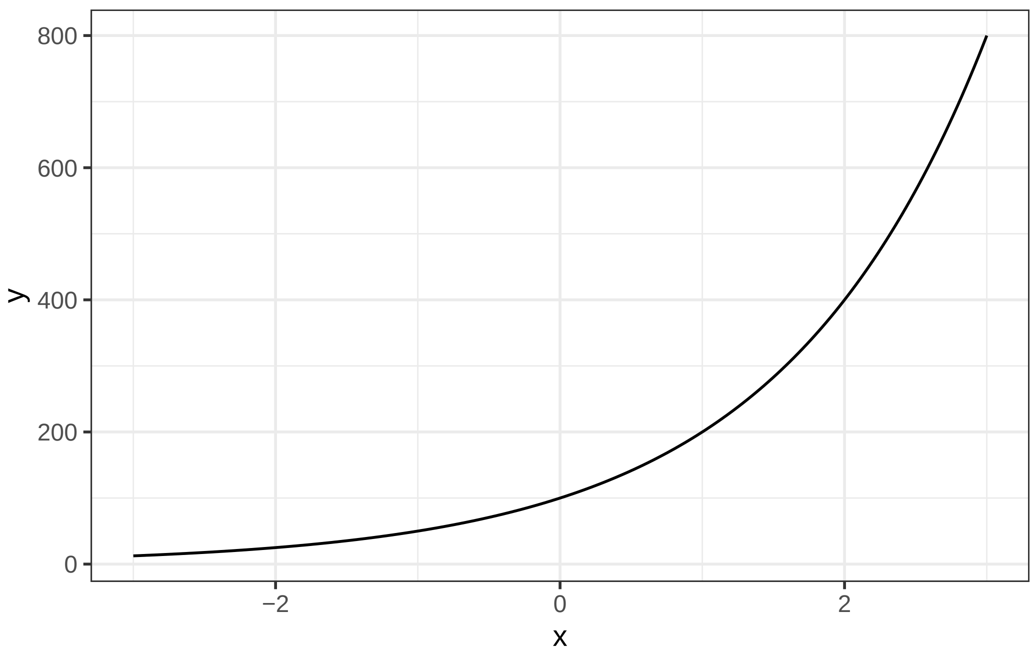
Consideremos una población de bacterias que se duplica cada hora. Si inicialmente hay 100 bacterias, podemos modelar el crecimiento de la población con la función exponencial:

$$P(t) = 100 \cdot 2^t$$

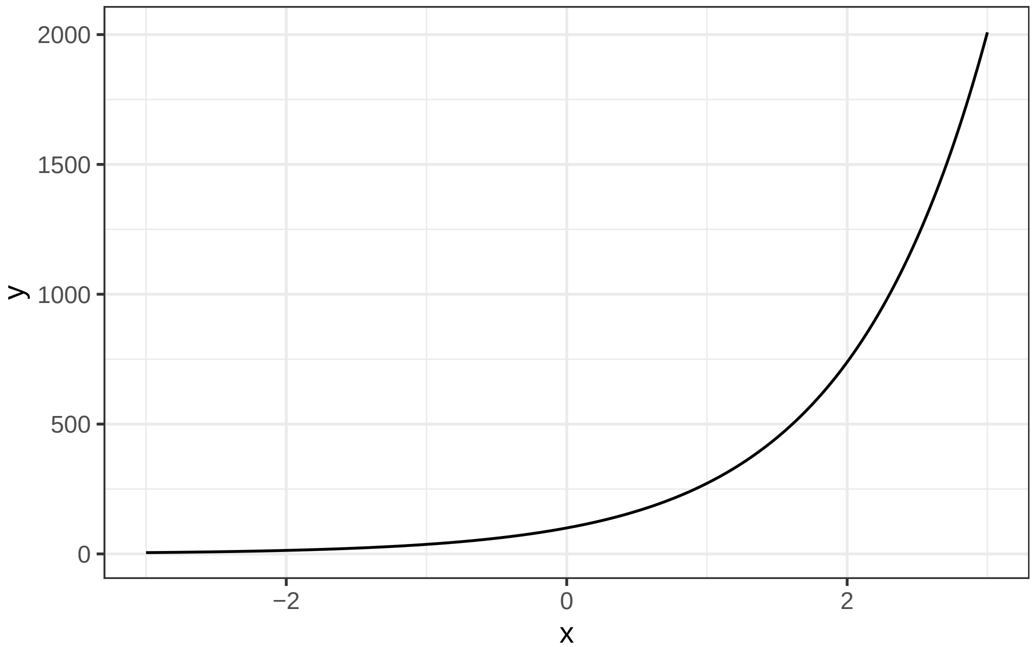
Aquí:

- $P(t)$ es la población de bacterias en el tiempo t (en horas).
- 100 es el valor inicial (a).
- 2 es la base (b), ya que la población se duplica cada hora.

```
x = seq(-3,3, 0.01)
y = 100*2^x
datos = data.frame(x, y)
datos %>% ggplot(aes(x=x, y=y))+
  geom_line()+theme_bw()
```



```
x = seq(-3,3, 0.01)
y = 100*exp(x)
datos = data.frame(x, y)
datos %>% ggplot(aes(x=x, y=y))+
  geom_line()+theme_bw()
```



Nota

Es muy importante reconocer el número $e $ 2.7182818 como la base más usada para tranajar con funciones exponenciales. Esto se debe a la propiedad siguiente:

$$y = ab^x = e^{ln(ab^x)} = e^{xlnb+a} = e^a e^{xlnb} = \tilde{a}e^{\tilde{b}x}$$

1.2.3 Funciones logarítmicas

Una función logarítmica es la inversa de una función exponencial. Si tenemos una función exponencial de la forma $y = b^x$, donde b es la base y x es el exponente, la función logarítmica correspondiente es $x = \log_b(y)$.

El logaritmo de un número y con base b es el exponente al cual hay que elevar la base b para obtener y. Matemáticamente, esto se representa como:

$$log_b(y) = x \Leftrightarrow b^x = y$$

En el contexto de la farmacia, las bases más comunes son e (el número de Euler, aproximadamente 2.718) y 10, dando lugar a los logaritmos naturales ($\ln(x)$) y logaritmos en base 10 ($\log_{10}(x)$), respectivamente.

Las propiedades básicas de los logaritmos son especialmente útiles para simplificar y resolver ecuaciones complejas:

a. $\log_b(1) = 0$ para cualquier base b.

b. Producto: $\log_b(xy) = \log_b(x) + \log_b(y)$.

c. Cociente: $\log_b(x/y) = \log_b(x) - \log_b(y)$.

d. Potencias: $\log_b(x^p) = p \cdot \log_b(x)$.

e. Cambio de base: $\log_b(x) = \frac{\log_a(x)}{\log_a(b)}$.

Las funciones logarítmicas son ampliamente utilizadas en diversos aspectos de la ciencia farmacéutica:

- **Farmacocinética**: Los logaritmos se utilizan para analizar cómo los medicamentos se distribuyen, metabolizan y eliminan en el cuerpo. Por ejemplo, la fórmula de eliminación de primer orden de un fármaco se representa mediante una función exponencial, y el tiempo de eliminación o la vida media se calculan usando logaritmos naturales.
- **pH y Química Farmacéutica**: En química, los logaritmos son fundamentales para calcular el pH, que se define como el logaritmo negativo de la concentración de iones de hidrógeno:

  $$pH = -\log_{10}([H^+])$$

  Este concepto es clave para entender la estabilidad y solubilidad de los fármacos, así como para desarrollar soluciones que mantengan un pH constante en formulaciones farmacéuticas.
- **Estudios de bioequivalencia**: Se comparan diferentes formulaciones de un mismo medicamento, el uso de logaritmos facilita el análisis de concentraciones plasmáticas a lo largo del tiempo, proporcionando una mejor comprensión de la absorción y distribución de los fármacos.
- **Escalas y Dosificación**: Los logaritmos se utilizan para interpretar datos que cubren un amplio rango de valores, como las curvas dosis-respuesta, donde la relación entre la concentración de un medicamento y su efecto biológico es no lineal y puede ser mejor representada en una escala logarítmica.

Consideremos un ejemplo práctico en el que se necesita determinar el tiempo necesario para reducir la concentración de un medicamento en sangre a la mitad (vida media). Supongamos que la eliminación del fármaco sigue una cinética de primer orden, es decir, la concentración

$$C(t) = C_0 \cdot e^{-kt}$$

Donde C_0 es la concentración inicial, k es la constante de eliminación y t es el tiempo.

La vida media ($t_{1/2}$) se define como el tiempo necesario para que la concentración del fármaco disminuya a la mitad de su valor inicial. Entonces, para un fármaco con cinética de primer orden, la vida media se calcula como:

$$t_{1/2} = \frac{\ln(2)}{k}$$

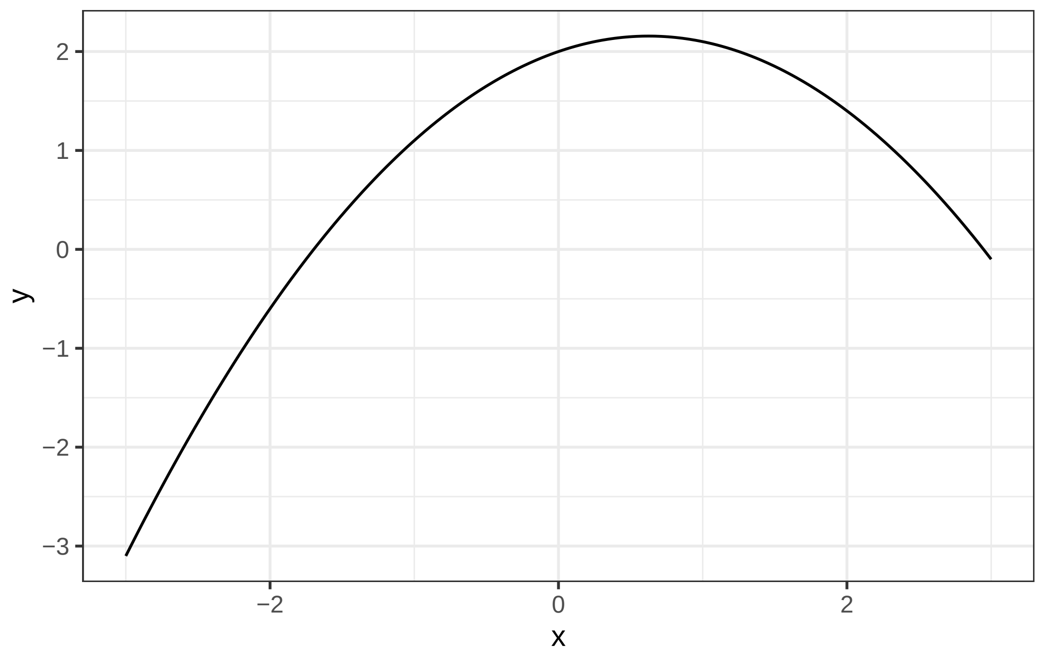
1.2.4 Funciones polinómicas

Una función polinomial de grado 2, también conocida como función cuadrática, tiene la forma $f(x) = ax^2 + bx + c$, donde:

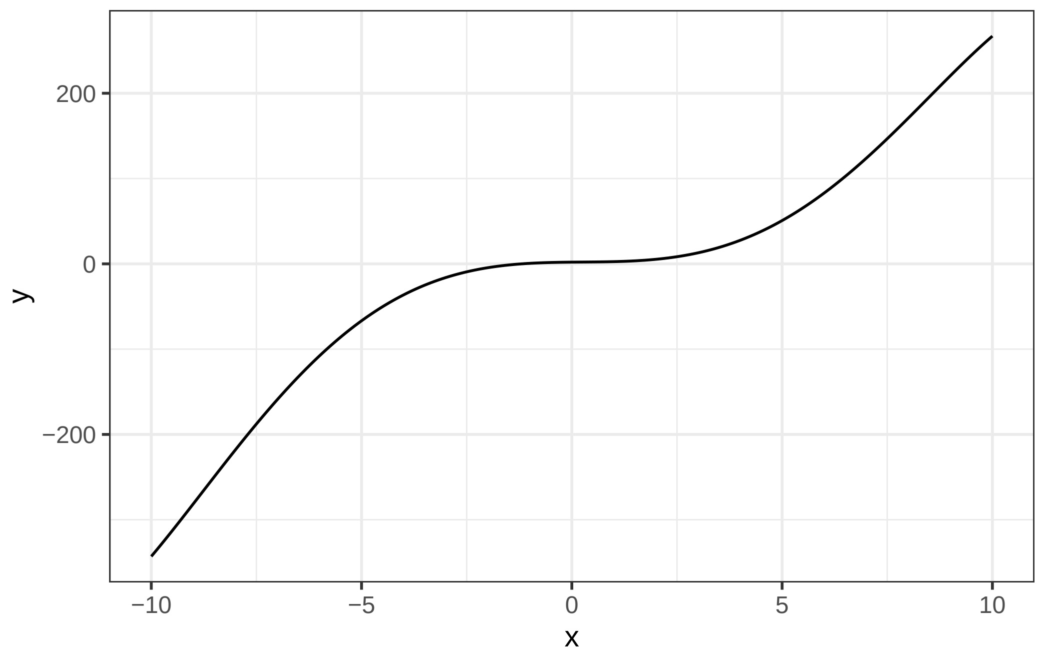
- a, b y c son constantes reales.
- a es distinto de 0.

La gráfica de una función cuadrática es una parábola que puede abrirse hacia arriba (si a es positivo) o hacia abajo (si a es negativo).

```
x = seq(-3,3, 0.01)
y = 2+0.5*x-0.4*x^2
datos = data.frame(x, y)
datos %>% ggplot(aes(x=x, y=y))+
  geom_line()+theme_bw()
```




```
x = seq(-10,10, 0.01)
y = 2+0.5*x-0.4*x^2 +0.5*x^3-0.002*x^5
datos = data.frame(x, y)
datos %>% ggplot(aes(x=x, y=y))+
  geom_line()+theme_bw()
```

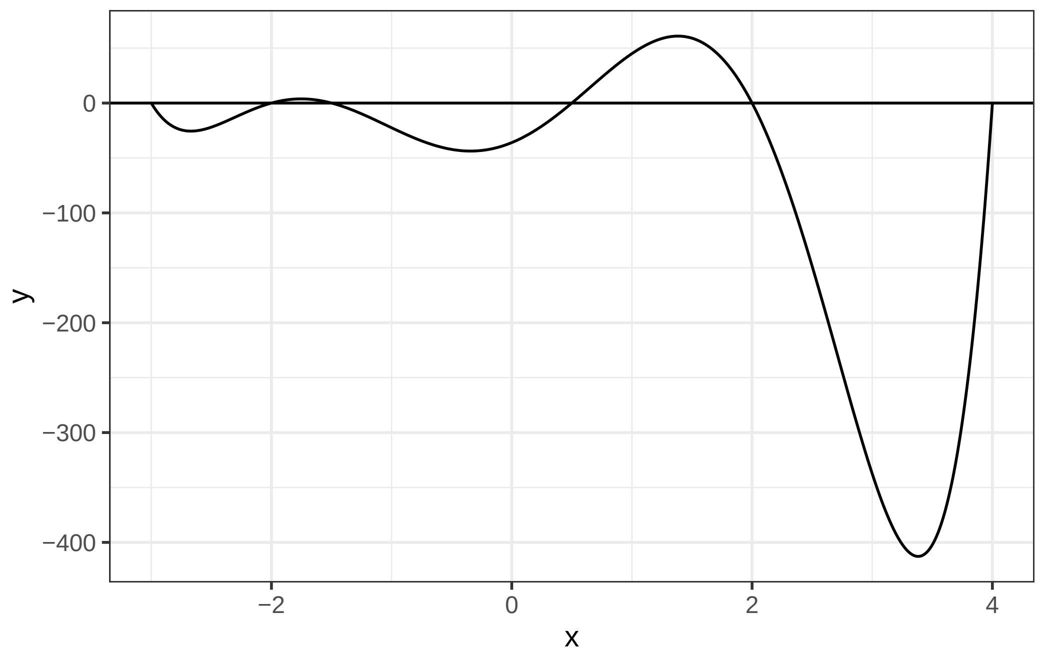


Nota

Para recordar factorización, ¿qué sucede si graficamos la siguiente función?

$$y = (x-2)(x+1.5)(x-0.5)(x+3)(x+2)(x-4)$$

```
x = seq(-3,4, 0.01)
y = (x-2)*(x+1.5)*(x-0.5)*(x+3)*(x+2)*(x-4)
datos = data.frame(x, y)
datos %>% ggplot(aes(x=x, y=y))+
  geom_line()+theme_bw()+
  geom_hline(yintercept = 0)
```

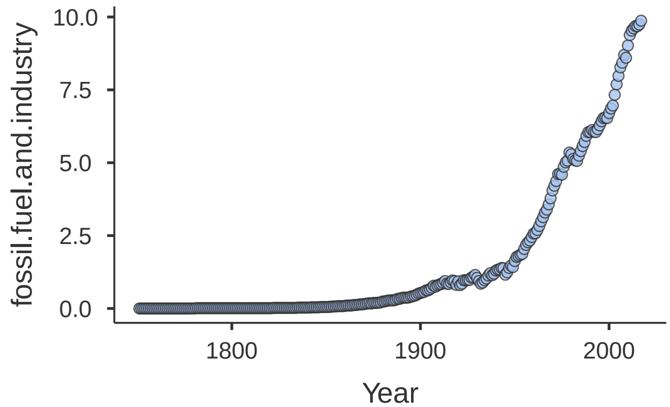


1.2.5 *Práctica 1*:

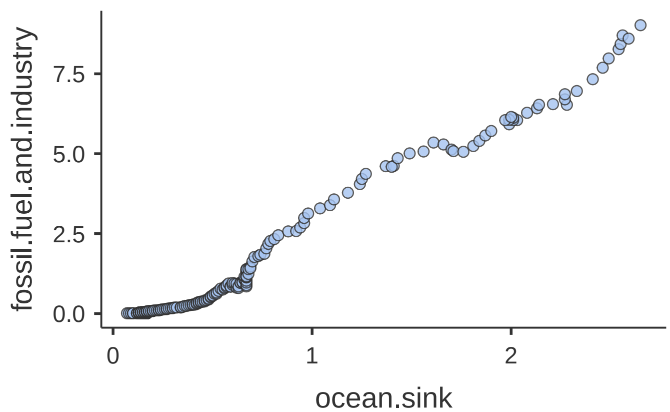
- Formad grupos de 1, 2 o 3 integrantes.
- Trabajaréis con los datos que aparecen en el enlace. Abrimos el enlace y con click derecho –> Guardar como y seguimos las instrucciones.
- Estos datos son tomados de la investigación sobre calentamiento global del Global Carbon Project. Allí se encontrará más información por si os interesa. Vamos a introducir los gráficos de dispersión y la visualización de posibles relaciones entre variables cuantitativas.

Seguiremos los siguientes pasos.

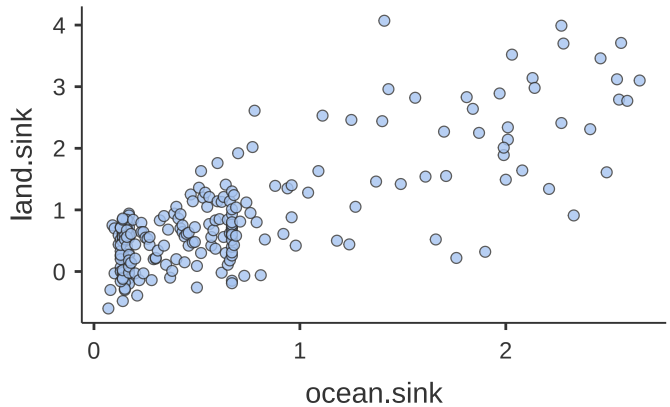
- Con la opción abrir, importamos la base de datos en JAMOVI.
- Verificamos que todas las variables estén cargadas
- Reproduciremos esta gráfica de dispersión. ¿Cómo se realiza la gráfica? ¿De los tipos de funciones, cuál se ajustaria mejor?



- Reproduciremos esta gráfica de dispersión. ¿Cómo se realiza la gráfica? ¿De los tipos de funciones, cuál se ajustaria mejor?



- Reproduciremos esta gráfica de dispersión. ¿Cómo se realiza la gráfica? ¿De los tipos de funciones, cuál se ajustaria mejor?



- Debéis exportar las gráficas en formato PDF.
- Entregad el reporte en la tarea de Aula Digital disponible. Revisad la fecha en que cierra la tarea.

2 Representación gráfica y análisis de Datos

2.1 La estadística y el método científico

- La ciencia avanza definiendo teorías que intentan explicar el mundo.
- La comunidad científica elabora teorías/hipótesis que intentan explicar hechos que ocurren. Una hipótesis es científica si existe alguna manera de comprobar su veracidad.
- Podemos diseñar experimentos para comprobar si se cumplen las afirmaciones de la teoría.
- Como la naturaleza tiene un comportamiento con "incertidumbre", es decir, que si repetimos el experimento se obtienen resultados similares pero no idénticos, la estadística permite analizar estos resultados y ver si las desviaciones de la teoría son razonables o no.
- Se ha definido estadística de muchas maneras. La que más nos gusta, y que está relaciona con la situación que acabamos de explicar, es que:

  La **estadística** es la ciencia que permite adquirir conocimiento generalizable a partir de datos.

- La estadística ayuda en todas las fases del método científico:
  - *Planteamiento del problema*: Diseño de experimentos y encuestas, determinación del tamaño de la muestra y métodos de muestreo adecuados para garantizar que los datos recopilados sean representativos de la población objetivo.
  - *Recopilación de datos*: Proporciona herramientas para recopilar y organizar datos relevantes sobre el problema.
  - *Análisis de datos*: Aplicación de técnicas descriptivas (Análisis explorartorio de datos), así como técnicas inferenciales (contrastes de hipótesis, ajustes de modelos,etc) para sacar conclusiones sobre la población en función de la muestra recopilada.
  - *Interpretación de resultados*: Ayuda a los científicos a determinar si los resultados son estadísticamente significativos y si las conclusiones se pueden generalizar a la población más amplia.

  - *Comunicación de hallazgos*: La estadística se usa para comunicar los resultados de manera efectiva a través de gráficos, tablas y tests estadísticos. Esto es esencial para que otros investigadores puedan comprender y evaluar los resultados.
  - *Reproducibilidad*: Proporciona métodos estadísticos claros y transparentes, se permite que otros repitan los experimentos y análisis para verificar la validez de los hallazgos.
  - *Toma de decisiones*: En muchos campos científicos, los resultados estadísticos se utilizan para tomar decisiones importantes. Por ejemplo, en la medicina, la estadística se usa para evaluar la eficacia de tratamientos y tomar decisiones sobre su uso en la práctica clínica.

En resumen, la estadística es una herramienta esencial que ayuda a garantizar que la investigación científica sea rigurosa, confiable y basada en evidencia sólida.

2.2 Conceptos básicos

2.2.1 Estudio clínico

Un estudio clínico es un proceso cuyo objetivo es obtener evidencia empírica sobre alguna cuestión. En el caso de los estudios que nos ocupan en este curso, esta cuestión es, naturalmente, sobre algún aspecto como la efectividad de un medicamento o algún tema de salud pública.

2.2.2 Unidad de observación

En un estudio estadístico, la unidad de observación es, para entendernos, el tipo de qué o de quiénes que son objeto de medición durante la investigación. En los estudios médicos normalmente serán personas, pero no siempre. Por ejemplo, pueden ser sucesos que les pasen a personas, de manera que una misma persona pueda ser observada varias veces: embarazos, operaciones quirúrgicas. Por ejemplo, podemos medir en diferentes centros de educación primaria de una ciudad el gasto medio diario en sus máquinas expendedoras de alimentos procesados y la proporción de miopes entre sus alumnos, para estimar si hay alguna relación entre el consumo de alimentos procesados y la miopía. Aquí, la unidad de observación son los centros de educación primaria, no los alumnos.

2.2.3 Población y muestra

- **Población**: Es el conjunto de todas las unidades de observación sobre los que queremos conocer alguna información. Esta población puede estar perfectamente definida en un lugar y tiempo: por ejemplo, los empadronados en Mallorca a día de hoy. Pero normalmente su definición será difusa. Si, por ejemplo, queremos estimar algo sobre "los españoles diabéticos mayores de 65 años", ¿de quiénes estamos hablando exactamente? ¿De los que están vivos justo ahora? ¿De todos los que ha habido en España desde su fundación? ¿Incluimos los que aún no han nacido? ¿Qué hacemos con los que son diabéticos pero no han sido diagnosticados, ni lo serán nunca?

- **Muestra**: Es un subconjunto de la población que se ha seleccionado para ser observado. La idea es que la muestra sea representativa de la población, de manera que los resultados obtenidos de la muestra puedan generalizarse a la población. En el ejemplo anterior, podríamos seleccionar una muestra de españoles diabéticos mayores de 65 años, y a partir de esta muestra intentar estimar alguna característica de la población de españoles diabéticos mayores de 65 años.

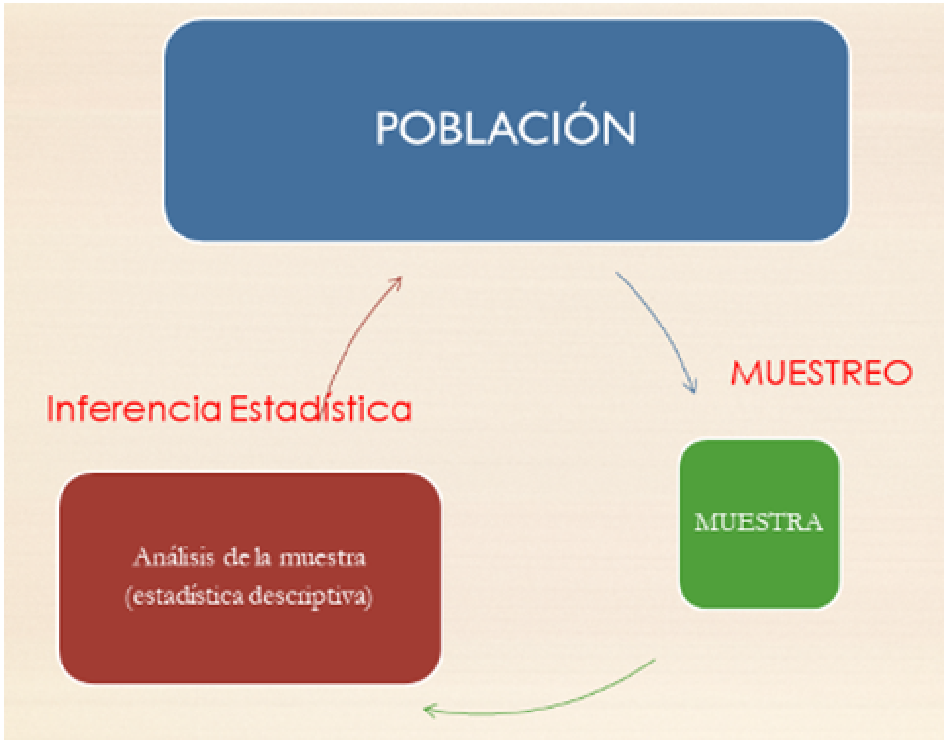
2.2.4 Tipos de estadística

En estadística, siempre se empieza obteniendo unos **datos** sobre una muestra de una población. Bueno, en realidad, no se empieza obteniendo los datos, sino planificando cuidadosamente cómo se van a obtener.

Se **generaliza la información** que se ha obtenido sobre este grupo de personas al total de la población. Y no se trata de trucos de magia adivinatoria, sino de una **ciencia** cuya metodología ha sido validada por medio de demostraciones matemáticas o, en el peor de los casos, mediante simulaciones numéricas (el equivalente en matemáticas de los experimentos en las otras ciencias).

Así pues, la situación de partida a la hora de aplicar técnicas estadísticas es que disponemos de un conjunto de datos que describen algunas características de un grupo de individuos. El análisis estadístico de estos datos puede ser entonces de dos tipos básicos:

- **Análisis exploratorio de datos**, cuando nuestro objetivo sea simplemente resumir, representar y explicar los datos concretos de los que disponemos. La **estadística descriptiva** es el conjunto de técnicas que se usan con este fin.
- **Análisis inferencial**, si nuestro objetivo es deducir (**inferir**), a partir de estos datos, información significativa sobre el total de la población de interés. Las técnicas que se usan en este caso forman la **estadística inferencial**.



Ambos tipos de análisis están relacionados. Por un lado, porque es conveniente (obligatorio, en nuestra opinión) empezar cualquier análisis inferencial dando un vistazo a los datos que se usarán.

Por otro, porque muchas técnicas descriptivas permiten estimar propiedades de la población de la que se ha extraído la muestra. Por citar un ejemplo, la media aritmética de las alturas de un grupo de individuos nos da un valor más o menos representativo de sus alturas, pero también sirve para *estimar* la altura media de los individuos de la población total.

La estadística inferencial entra en juego cuando se quiere obtener información sobre una población y no se puede acceder a todos sus integrantes. Si por ejemplo queremos conocer la altura media de los estudiantes matriculados en esta asignatura de la UIB en este curso, en principio no necesitamos para nada la estadística inferencial. Sois pocos, os mediríamos a todos y calcularíamos la media. En todo caso, usaríamos técnicas de estadística descriptiva para arropar este valor representando la distribución de vuestras alturas de manera adecuada.

Pero si quisiéramos conocer la altura media de los mallorquines entre 18 y 25 años, sería muy complicado medirlos a todos. Entonces, lo que haríamos sería tomar una muestra representativa de esta población, medirlos y a partir de sus alturas estimar dicha altura media. Naturalmente, lo más seguro es que de esta manera no obtuviéramos el valor exacto de la altura media de los mallorquines de 18 años, nos tendríamos que conformar con obtener una aproximación dentro de un cierto margen de error y determinar la probabilidad de acertar con nuestra estimación y este margen de error. La estadística inferencial es la que nos permite acotar el error que podamos haber cometido y calcular la probabilidad de cometerlo, incluyendo la metodología que tendríamos que haber usado para tomar la muestra en primer lugar.

2.2.5 Tipos de estudios

Podemos clasificar los estudios clínicos de diferentes maneras:

- Según su **intención**:
  - **Descriptivos**: Se limitan a describir las características de los individuos de la muestra.
  - **Analíticos**: Intentan inferir conclusiones sobre el total de la población.
- Según el **papel jugado por el investigador**:
  - **Observacionales**: El investigador se limita a recoger datos, sin ejercer ninguna influencia planificada sobre los acontecimientos que generan dichos datos.
  - **Intervencionista**: El investigador lleva a cabo una intervención en los participantes (por ejemplo, les administra tratamientos farmacológicos o recomienda cambios de comportamiento) de manera planificada y con el objetivo de generar los datos que permitan evaluar el efecto de dicha intervención.
- Según el **lapso de tiempo** sobre el que se recoge la información:
  - **Transversales**: Se recoge información sobre un solo momento.
  - **Longitudinales**: Se recoge información sobre varios momentos de tiempo y se estudian los cambios producidos entre los mismos.

    A su vez, estos últimos suelen dividirse en:
    - **Prospectivos**: Se recoge información en un momento concreto (normalmente, al inicio del estudio) y en momentos posteriores.
    - **Retrospectivos**: Se recoge información en un momento concreto (de nuevo, normalmente, al inicio del estudio) y sobre momentos anteriores.

Combinando los tipos de estudio según el papel jugado por el investigador y según el lapso de tiempo sobre el que se recoge la información, tenemos estudios:

- Observacionales transversales
- Observacionales prospectivos
- Observacionales retrospectivos
- Intervencionistas transversales

- Intervencionistas prospectivos
- Intervencionistas retrospectivos

Acabamos de ver que hay estudios intervencionistas prospectivos. ¿Los hay de las otras cinco clases de estudios de esta lista, o hay algún par de características que es imposible que se den simultáneamente?