



Introducción a las relaciones empíricas



Jesús Antonio del Río Portilla
2022

Introducción a las relaciones empíricas

Universidad Nacional Autónoma de México
Instituto de Energías Renovables

© 2022 D.R. Universidad Nacional Autónoma de México

Todos los derechos reservados. Hecho en México.

Se autoriza la distribución bajo licencia creative commons:
atribución, no comercial y compartir igual .

 2022, Jesús Antonio del Río Portilla.

Diseño de Portada

Jesús Antonio del Río Portilla usando Canva <http://canva.com> e imagen de <https://storyset.com/>

ISBN 978-607-30-7275-5.

DOI

Publicado en formato digital en Temixco, Morelos, México, abril 2022.

Introducción a las relaciones empíricas

J. Antonio del Río

2022

Índice general

1. Representación de datos	3
1.1. Mediciones	4
1.1.1. Diferencias	5
1.2. Representación en tabla	6
1.3. Gráficas	8
2. Relaciones empíricas	11
2.1. Relaciones empíricas lineales	13
3. Relaciones exponenciales	17
4. Relaciones potenciales	23
5. Apéndice	31

Introducción

En actividades académicas, científicas, tecnológicas, de negocios donde se obtienen datos tanto experimentales como observacionales su representación por medio de fórmulas matemáticas ayuda a la construcción de modelos. Estos modelos con fundamento matemático facilitan el entendimiento de los fenómenos o situaciones de donde provienen los datos. También estas representaciones matemáticas son una herramienta en la predicción del comportamiento en el tiempo o bajo otras circunstancias. A estas representaciones matemáticas se les conoce como Ecuaciones Empíricas o Relaciones Empíricas o Fórmulas Empíricas. En todas estas formas de expresión se enfatiza en el carácter empírico de la construcción. Es decir, que la obtención de estas fórmulas se apega lo más posible a los datos desde donde se construyen.

En este texto introductorio a las relaciones empíricas se aborda de manera sencilla los conceptos de medición y diferencias en las mediciones. Estas diferencias en las mediciones en otros lugares se les llama errores en las mediciones. Continuamos con la descripción de la forma de presentar los datos que observamos o que obtuvimos de los experimentos para que sean claros y se comuniquen en forma adecuada para que otras personas puedan repetir nuestros procedimientos o con estos datos construyan nuevas formas de entender nuestros hallazgos. Así presentamos los requerimientos mínimos de las Tablas y de las Gráficas para que se puedan comunicar.

En la parte central de este texto abordamos ejemplos de datos observacionales para construir Relaciones Lineales, Exponenciales y Potenciales. La metodología se basa en utilizar las funciones de regresión lineal que aparecen en la mayoría de las hojas electrónicas como Excell, LibreCal, Hoja de Calculo de Google, etc., pero también se ofrece el código en Python de la regresión lineal en el apéndice donde se usaron los ejemplos abordados. En los dos últimos capítulos se ilustra la manera de realizar la transformación a los datos para obtener relaciones lineales en las variables transformadas y con ellas, fácilmente, obtener relaciones más complejas que las lineales.

El texto se considera que es útil para cualquier persona que desea expresar una colección de datos en una forma matemática. Se requiere una preparación de bachillerato con conocimientos elementales de geometría analítica, en particular, de la forma de la ecuación de una recta. Es ideal que se tenga un manejo intermedio de alguna hoja de cálculo o un nivel básico de programación en algún lenguaje como Python.

El objetivo de este libro es que las habilidades que adquirirá la persona que lea este texto y realice el seguimiento detallado de los desarrollos en paralelo a la lectura del texto tendrán aplicaciones en el ámbito de la ciencia, la ingeniería, los negocios y muchas otras actividades.

Iniciamos explicando la forma de presentar los datos de una manera comunicativa.

Capítulo 1

Representación de datos

En este curso se presentan diferentes estrategias sencillas para analizar datos experimentales y encontrar relaciones empíricas que puedan ser comparadas con modelos teóricos o con ellas proponer nuevos modelos matemáticos. Se invita a las personas que decidan incursionar en esta simple forma de analizar los datos a usar estas metodologías en sus clases de laboratorio o primeros escrutinios de sus datos en proyectos de investigación. Este texto fue inspirado en el trabajo pionero del profesor jubilado de la Facultad de Ciencias, UNAM, Juan Américo Gonzáles [1].

En física y otras ciencias uno de los aspectos fundamentales para avanzar en el conocimiento de los fenómenos son las mediciones de diferentes cantidades físicas. Una vez que se tienen las mediciones se puede comparar diferentes experimentos o contrastar las predicciones de diferentes teorías con los experimentos. En este sentido, la realización de buenos experimentos radica en la calidad de las mediciones que se realizan para poder discernir las similitudes o diferencias entre las teorías y los experimentos.

En este capítulo proporcionaremos una visión general del concepto de mediciones y la forma de presentar datos experimentales.

A partir de una serie de datos experimentales, el alumno obtendrá relaciones empíricas, gráficas y expresiones analíticas, entre las variables que describen al fenómeno, indicando la precisión de los datos y los límites de validez de los modelos matemáticos que describen las relaciones entre los datos.

Lo esencial entonces es medir, la forma de medir y cómo se reportan las mediciones es crucial para que puedan reproducir nuestros resultados.

1.1. Mediciones

Debe quedar claro el concepto de medición como una acción de comparar dos magnitudes del mismo tipo donde una de ellas ya ha sido comparada con una magnitud que se considera el patrón de esa magnitud. Generalmente, a este patrón de medición se le conoce como unidad de medición. Por ejemplo, una regla ya ha sido comparada con un patrón conocido “el metro patrón”; así en este ejemplo es la unidad metro.

Existen unidades que llamamos fundamentales como el metro, el segundo o la carga entre otras y llamamos unidades derivadas como el newton, watt, amper, etc. que son el resultado de medir cantidades que son determinadas mediante alguna fórmula.

Para que una medida sea aceptable debe ser reproducible, es decir debe dar “lo mismo” ante una repetición o la acción por diferentes experimentadores. Una medida reproducible por tanto requiere un proceso de medida o un ensayo no destructivo. Ejemplo: Si se mide cualquier número de veces un lado de un escritorio, siempre se debe obtener el mismo resultado. Las medidas reproducibles son procedimientos no destructivos que, además, no producen una alteración importante en el sistema físico sujeto a medición. En el caso que el proceso de medición sea destructivo, se deben preparar los objetos a medir idénticos y los resultados del proceso de comparación deben ser los mismos. Este concepto de medición opera para situaciones donde la aproximación clásica es válida. El concepto de medición se modifica cuando realizamos observaciones en el ámbito cuántico; pero no abordaremos estos temas en este libro.

Entonces bajo la premisa de que deseamos describir una situación clásica podemos afirmar que cuando se realiza una medición de la variable x n -veces podemos indicar el valor de la medición i como x_i para obtener una colección de mediciones conforme se repite la medición. Para poder comunicar a otras personas o tener nuestras notas, se puede construir una representación de los datos en un formato de tabla, como el indicado en la Tabla 1.1:

Primero, observemos algunos detalles de presentación en la Tabla 1.1. Es importante tener un encabezado que indique el nombre de la variable que se está midiendo, en la columna de los resultados debe estar clara la unidad en la que se presentan. Generalmente se indica entre paréntesis. En ciencia normalmente se usa el punto decimal, en lugar de la coma decimal. Es decir, la separación entre las unidades y los decimales es un punto. Es esencial hacer lo necesario para expresar adecuadamente las cantidades; esto ya que muchos paquetes informáticos al seleccionar español cambian este indicador de unidades y usan la coma para la separación de decimales. Segundo, en la

medición	resultado (unidades)
x_1	5.2
x_2	5.3
\vdots	\vdots
x_n	5.2

Tabla 1.1: Representación de x_i mediciones en formato tabular.

tabla se están representando n mediciones de la misma cantidad en las mismas circunstancias por lo tanto se espera obtener un valor similar; claramente los valores reportados pueden no ser iguales, pero esperaríamos que no fueran “muy” diferentes. Hagamos un paréntesis para contestar la pregunta que surge al leer estas líneas: ¿Cómo podemos entender estas posibles diferencias?

1.1.1. Diferencias

Durante el proceso de medición es común que las mediciones no coincidan “exactamente”, como en el ejemplo de la tabla 1.1 y tengan pequeñas diferencias. Comúnmente a estas diferencias se les conoce como errores; sin embargo, en una interpretación moderna no podemos definir un resultado como exacto, ya que esa exactitud está limitada por la interacción entre el objeto que se usa para medir y el objeto que se está midiendo y existen condiciones que pueden variar ocasionando pequeñas diferencias que no permiten fácilmente decir cuál medición es “verdadera”. Por esta razón, en este texto quiero seguir la línea de pensamiento que considera la certeza en las mediciones y no enfatiza los errores que pueden conducir a malas interpretaciones del quehacer experimental. Uno de los aspectos importantes en el proceso de medición es tener claro el rango de validez de la medición, es decir, de la certeza que tenemos en ella. No es posible indicar las diferencias con adjetivos como muy, grande, pequeños, etc. En cuanto al reporte de las diferencias es esencial cuantificarlas para poder reportarlas. Es necesario expresar estas diferencias en forma cuantitativa. Para ello, podemos usar la suposición que las diferencias se deben a factores aleatorios y, entonces, suponer que las mediciones obedecen una distribución gaussiana o normal y describirlas por su promedio y desviación estándar (primer y segundo momento de la distribución), es decir, podemos calcular:

$$\bar{x} = \frac{\sum_i^n x_i}{n} \quad \text{y} \quad \sigma_x = \frac{\sum_i^n (x_i - \bar{x})^2}{n}, \quad (1.1)$$

donde \bar{x} y σ_x representan el promedio y la desviación estándar del conjunto de mediciones que se presentan en la tabla 1.1. Es importante enfatizar que con los valores de \bar{x} y σ_x estamos representando a todo el conjunto de mediciones reportadas en esa tabla, generalmente se representan en una forma muy sencilla:

$$\bar{x} \pm \sigma_x. \quad (1.2)$$

De esta manera simple, estamos indicando que hubo un proceso de medición y representando todo ese proceso de múltiples mediciones de la variable x con solamente dos números. Generalmente se sobreentiende que la probabilidad de obtener un valor en el intervalo entre $\bar{x} - \sigma_x$ y $\bar{x} + \sigma_x$ al realizar otra medición es aproximadamente 70 %. Por otro lado, también se asume que se realizaron más de tres mediciones, de otra forma se subestiman las posibles diferencias y la probabilidad de tener diferencias grandes en alguna otra repetición aumenta significativamente. Es conveniente realizar un número de mediciones de tal manera que se obtenga $\sigma \ll \bar{x}$, por ejemplo $\sigma \sim \frac{\bar{x}}{20}$. Esto indica que la precisión del aparato y de la metodología es adecuada para la medición.

En una forma resumida podemos decir que al expresar una medición como en la expresión 1.2 estamos indicando que al realizar otra medición tenemos la certeza que con un 70 % de probabilidad la nueva medición estará en el intervalo indicado. El intervalo de certeza envuelve a la mayoría de las mediciones y nos permite expresarlas en una forma sintética. De no envolver a la mayoría de las mediciones podría deberse a que la medición no es reproducible y tendríamos que tomar otras acciones. En este texto consideraremos que nuestras mediciones son reproducibles.

Regresemos a comentar la forma de presentar datos en el formato de tabla.

1.2. Representación en tabla

Frecuentemente estaremos buscando las relación entre dos variables, a una se le denomina independiente y a la otra dependiente y en este caso deberemos reportar las tablas en una forma general 1.2. La variable indepen-

diente es la que nosotros fijamos en el experimento y la o las dependientes son el resultado de nuestra medición.

Por ejemplo consideremos un experimento donde se calienta un cuerpo y medimos la evolución de la temperatura. Expresamos estos datos de la Tabla 1.2 donde se muestra una colección de datos de tiempo contra temperatura:

Tiempo $\pm 10^4 \%$ (s)	Temperatura $\pm 0.1 \%$ (K)
1	347
2	427
3	516
4	664
5	783
6	929
7	1.20×10^3
8	1.48×10^3
9	1.69×10^3
10	2.06×10^3

Tabla 1.2: Ejemplo de tabla con los requerimientos mínimos.

Observemos que dado que en el encabezado se estableció que el intervalo de confianza en la temperatura es de un 0.1 % implica que hasta un milésimo de la magnitud es confiable. Para ejemplificar si la magnitud es 3,895 una certeza del 0.1 % se expresaría explícitamente como $3,895 \pm 4$, en otras palabras no debemos usar más de cuatro cifras para expresar esta cantidad. Al número de dígitos con los que podemos escribir una cantidad sin tener conflictos con el rango de certeza se le llama cifras significativas. Dado que las posibles mediciones que obtuviéramos estarían entre 3,899 y 3,891, por esta razón, los últimos datos no tienen cifras decimales. Ya que con una confianza del 0.1 % la cuarta cifra no es de confiar. En cambio en el tiempo se ahorraron las cifras decimales, pero se sobreentiende que están presentes, es decir los números en el tiempo para todo cálculo deben ser considerados 1.000, 2.000, etc.

En el caso en que en las mediciones de temperatura la certeza estuviera en el 0.01 %, entonces los valores con los que representaríamos a las mediciones deberían representarse en forma diferente para dejar clara esta certeza. En la

tabla 1.3 presentamos como se presentarían los datos con una mayor certeza en la temperatura.

Tiempo $\pm 10^4 \%$ (s)	Temperatura $\pm 0.1 \%$ (K)
1	347.2
2	427.4
3	516.7
4	664.3
5	783.1
6	929.2
7	1206
8	1486
9	1698
10	2067

Tabla 1.3: Otro ejemplo de presentación de datos mediante una tabla con cifras significativas diferentes, comparar con la Tabla 1.2.

Notemos que a diferencia de la tabla 1.2 en la tabla 1.3 no se usó la notación científica, dado que las cifras significativas son 4, aunque también se podrían haber expresado como 2.067×10^3 . Siempre se debe revisar que la forma de expresar las mediciones tenga el número correcto de cifras significativas. No es adecuado reportarlas tal y como se leen en las hojas de cálculo o como se expresan en las salidas de los programas de cómputo.

No siempre la representación de los datos en una tabla es la más adecuada para el análisis y en ocasiones para entender un conjunto de mediciones se prefiere la representación en un gráfica. En la siguiente sección discutiremos algunas de las características para la representación gráfica de colecciones de datos experimentales u observacionales que dependen de una variable.

1.3. Gráficas

Cuando se tiene una colección de datos en función de alguna variable, sea espacial, temporal o de alguna otra cantidad física su representación en forma

de gráfica puede ser más adecuada que en formato de tabla. La visualización de los datos en esta manera gráfica puede aportar información global sobre el comportamiento de la variable dependiente medida en función de la variable con la que se asoció su dependencia.

En la Tabla 1.2 vemos una serie de datos de temperatura en función del tiempo. De su simple inspección no podemos inferir en forma simple su comportamiento, en cambio al escoger su representación gráfica en la Fig. 1.1 podemos reconocer con mayor facilidad su comportamiento.

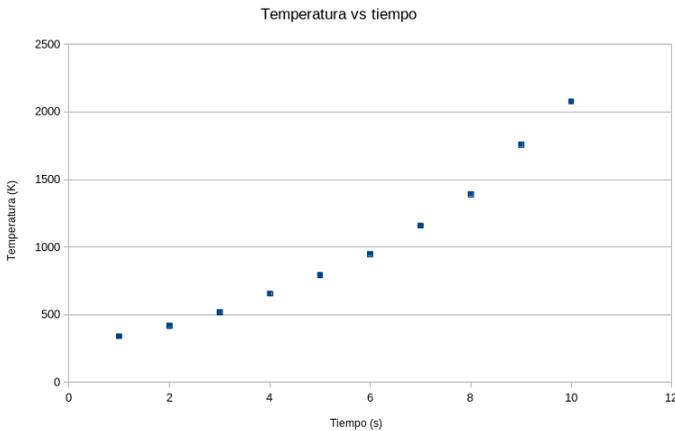


Figura 1.1: Gráfica representando los datos de temperatura y tiempo de la Tabla 1.2.

Antes de interpretar los resultados es importante observar algunos detalles en las gráficas, como vemos en la fig. 1.1. Es conveniente que la gráfica tenga un título, puede ser tan simple como escribir “temperatura vs tiempo” o decir el nombre del experimento. Para una lectura adecuada de las gráficas, es importante que se escriban los nombres de los ejes, deben escribirse explícitamente las unidades en las que fueron tomadas las mediciones o se realizaron los cálculos. También es conveniente escribir un pie de figura, donde se expliquen brevemente los datos que se están presentando.

Las gráficas pueden tener retícula para facilitar la lectura de los datos, pero debe ser espaciada para que no interfiera con los mismos datos o líneas que se tracen con los datos. Una gráfica ilustrativa hace que los datos ocupen la mayor parte del área. Así que la forma de escoger los límites en los ejes es

importante. La mayoría de los programas para manejar datos como Excell u OpenCalc tienen las facilidades para hacer la selección adecuada de la escala. En ocasiones cuando se grafican varias series de datos, es conveniente usar leyendas para distinguirlas.

En cuanto al comportamiento, claramente, esta gráfica nos indica que la temperatura se incrementa mientras el tiempo crece; pero no en la misma proporción, es decir la relación no es lineal. Si intentamos trazar una línea sobre estos datos veremos que no puede ser una de un solo trazo sino que tendría ondulaciones. En esta parte del texto no deseamos obtener la expresión matemática que pueda representar o describir estos datos. En cambio hemos ilustrado claramente los elementos mínimos de una gráfica comunicativa. La figura 1.1 ilustra una gráfica que tiene la información para que otra persona la entienda ¹.

En las próximas secciones describiremos como a partir de una serie de datos experimentales, se pueden obtener relaciones empíricas y así poder comparar con expresiones analíticas, teóricas u otros experimentos entre las variables que describen al fenómeno, indicando la precisión de los datos y los límites de validez de los modelos matemáticos que describen las relaciones entre los datos.

¹Para una discusión más detallada de la forma de graficar datos se sugiere consultar el libro de Enrique Galindo, [2].

Capítulo 2

Relaciones empíricas

En esta sección describiremos como podemos obtener relaciones empíricas de datos experimentales o de situaciones reales. Empezaremos con una colección de datos pequeña. En muchos casos, cuando la colección de datos considera un intervalo pequeño de la variable independiente, se puede encontrar que la ecuación de una recta es adecuada para describir el comportamiento de la colección de datos. Es así, como se puede decir que la aproximación lineal describe el fenómeno o el experimento. En esta sección daremos un ejemplo de esta última situación.

Consideremos los datos de la pandemia de COVID-19 para el estado de Morelos. Contamos con la información disponible de los Datos Abiertos de la Secretaría de Salud[3]. En la tabla 2.1 se presentan los datos del 8 de abril al 15 de abril del 2020, un intervalo en las primeras semanas.

Para tener una idea de la posible relación que describiría estos datos es conveniente graficarlos. Presentamos estos datos en forma gráfica en la figura 2.1, claramente se ve un incremento sostenido y al ver esta gráfica, nuestros ojos pueden unir los puntos (datos) con una línea imaginaria, es decir, parece un crecimiento lineal. Nuestro reto es poder expresar matemáticamente la ecuación de una recta que pase muy cerca de todos los puntos. En otras palabras, podríamos obtener una ecuación para esta línea al hacer un ajuste de regresión lineal.

Hace algunas décadas para obtener la expresión matemática de la recta que pasa más cerca de los datos experimentales se usaban métodos gráficos y se trazaba a mano esa recta y se obtenía por inspección de la gráfica la pendiente, m , y la ordenada al origen, b , de esa recta para, así tener la ecuación de la recta en la forma:

fecha AAAA-MM-DD	Casos confirmados COVID-19 (número de personas)
2020-04-08	30
2020-04-09	36
2020-04-10	40
2020-04-11	43
2020-04-12	48
2020-04-13	54
2020-04-14	58
2020-04-15	62

Tabla 2.1: Datos de personas confirmadas por el Modelo Centinela de enfermedades respiratorias como COVID-19

$$y = mx + b. \tag{2.1}$$

Hoy en día, con el uso masivo de las computadoras, tenemos muy diversas herramientas para calcular la m y la b de la recta que más se aproxima a nuestros datos experimentales. En general el método que usan estas herramientas computacionales es el método de mínimos cuadrados. Esta idea consiste en encontrar la pendiente y la ordenada al origen que minimice la distancia en la componente y entre los datos y la recta que deseamos. Para conocer los detalles matemáticos les sugiero consultar el libro de Baird [4] o directamente en Wikipedia [5]. También pueden analizar los programas que se han desarrollado en Python y que acompañan a este texto; también están disponibles en GitHub [9].

En la siguiente sección ilustraremos cómo obtener esta expresión matemática.

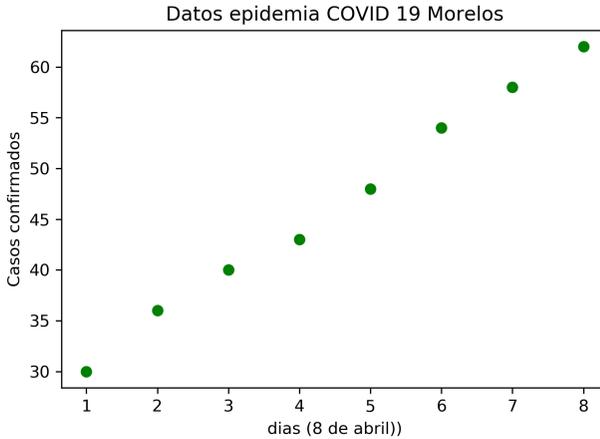


Figura 2.1: Gráfica representando los casos confirmados de COVID-19 de la tabla 2.1

2.1. Relaciones empíricas lineales

Antes de introducir los datos en alguna herramienta hay que prepararlos. Aquí solamente diremos que esta forma puede ser realizada colocando los datos en una hoja de cálculo como: Excell, LibreCalc o hacer uso de un lenguaje de programación como Python (ver apéndice).

Antes de poner los datos en esas herramientas conviene transformar las fechas en días a partir del 8 de abril del 2020. La idea es que se puedan hacer los cálculos en forma decimal en lugar de usar fechas, que no siempre pueden ser usadas en cálculos numéricos de manera simple. De esta manera, los datos se pueden expresar en la tabla 2.2.

Notemos el cambio en el nombre de la primera columna de la Tabla 2.2. Al introducir los datos en alguna herramienta de cómputo podemos obtener la línea recta que más se aproxima.

En las hojas de cálculo normalmente se colocan los datos de la variable independiente (abscisa, x) en una columna y a su derecha se colocan los correspondientes datos de la variable dependiente (ordenadas, y) y en un rango de celdas se usa la función “regresión lineal”. En el programa de cómputo se introducen los datos en él o se mandan llamar de un archivo texto. En el ejemplo del apéndice, se introducen los datos en formato de “array”. La he-

días (8 de abril)	Casos confirmados COVID-19 (número de personas)
1	30
2	36
3	40
4	43
5	48
6	54
7	58
8	62

Tabla 2.2: Datos de personas confirmadas por el Modelo Centinela de enfermedades respiratorias como COVID-19.

ramienta nos dará como resultado tanto $m = 4.53$ (casos/día) como $b = 26$ (casos) y algunos otros parámetros del cálculo. Por esta razón, es importante leer los manuales de las hojas de cálculo o, en su caso, entender el programa que utilizaremos. Con estos valores podemos escribir la siguiente fórmula:

$$\text{Casos} = 4.53\text{día} + 26 = 479, \quad (2.2)$$

donde 4.53 tiene unidades de casos *confirmados/días* y 26 es el *número de casos* confirmados al día cero (7 de abril). Con este resultado, graficamos la ecuación 2.2 y los datos en una sola gráfica en la figura 2.2 para comparar el resultado del modelo lineal con los datos.

Observamos que la recta se acerca bastante a los datos y podemos concluir que cualitativamente la ecuación 2.2 representa a los datos. De esta manera, hemos obtenido la primera ecuación empírica que aproxima a los primeros datos de la epidemia de COVID-19 en el estado de Morelos. Existen parámetros para establecer que tan confiable es una relación empírica, pero eso lo dejaremos para otra ocasión.

Es esencial comentar que si bien se nos menciona que una epidemia crece exponencialmente con los datos de estos 8 días no podemos rechazar la descripción más sencilla. Esta descripción sencilla es una línea recta. De hecho el ajuste es muy bueno, la inspección visual y los parámetros de bondad de ajuste son más que adecuados. De esta manera, una conclusión con un número pequeño de observaciones (datos experimentales) puede llevarnos a conclu-

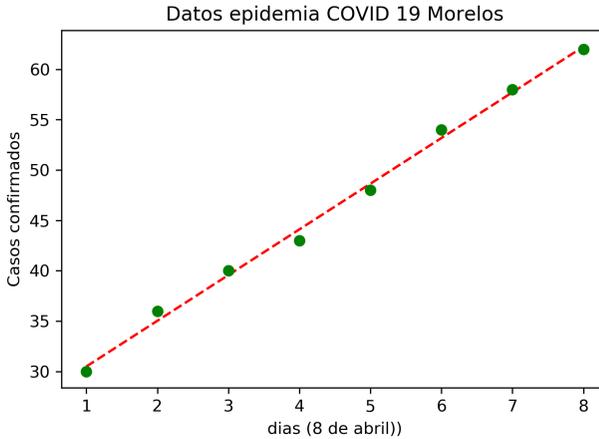


Figura 2.2: Datos de personas confirmadas por el Modelo Centinela de enfermedades respiratorias como COVID-19, con datos de la Tabla 2.2.

siones apresuradamente. Llevemos este ejercicio a explorar conclusiones más allá de las limitaciones que tiene. Como una aplicación de este modelo podemos calcular el número de posibles confirmados al día 100 después del día 8 de abril usando la ecuación 2.2, es decir, podemos hacer el siguiente caso:

$$4.53(100) + 26 = 479.$$

Este resultado de 479 confirmados como se verá posteriormente no es adecuado, pero los datos se ajustan. La validez de las ecuaciones empíricas debe ser analizada con detalle; sin embargo, el trabajo científico puede ser valioso si se explicitan las condiciones que se tienen para obtener estas ecuaciones. Del análisis de estas condiciones o limitaciones puede obtenerse información para modificar las hipótesis de las relaciones empíricas obtenidas. En este ejemplo, podemos pensar que son muy pocos los días analizados y, entonces, para cambiar de hipótesis a una más compleja necesitamos más datos. Por esta razón, ahora analizaremos más de 30 días en la siguiente sección e ilustraremos cómo obtener una relación exponencial.

Capítulo 3

Relaciones exponenciales

En esta sección describiremos cómo obtener ecuaciones empíricas de forma exponencial. La esencia de este capítulo es que utilizaremos las mismas herramientas de cálculo (hojas electrónicas o programas); pero antes de usarlas transformaremos los datos.

De la misma base de datos, pero ahora considerando una serie de 32 días construimos la tabla 3.1, donde mostramos los datos a partir del mismo día 8 de abril. En esta ocasión, la inspección visual de los datos es muy complicada y no parece que podamos obtener alguna información rápidamente. Sin embargo observamos que el número de casos crece lentamente en el inicio de la pandemia y posteriormente crece rápidamente.

Para intentar describir los datos de la Tabla 3.1 los mostramos en la figura 3.1. Ahí observamos que los primeros ocho días que presentamos en la figura 1 del módulo anterior, solo eran el comienzo de un crecimiento mucho más rápido.

Al ver esta curva parece que se asemeja a una curva exponencial. Es decir una curva que matemáticamente se representa por:

$$y = Ae^{mx}, \quad (3.1)$$

donde A indica una amplitud o el valor inicial (cuando $x = 0$ y m es la tasa de crecimiento). Por simplicidad en este cálculo hemos usado la base e para el comportamiento exponencial; pero el análisis es similar si se usa la base 10. Para saber si esta ecuación (3.1) describe a los datos podemos hacer otra gráfica. En una hoja de cálculo como Excell o LibreCalc basta con cambiar a escala logarítmica en el eje “Y” de la gráfica para observar el cambio en ella;

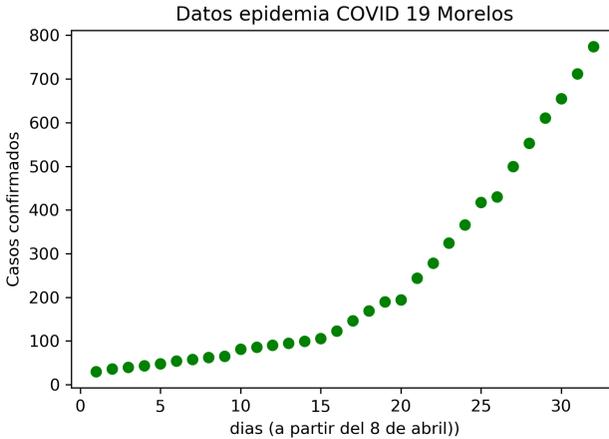


Figura 3.1: Datos de personas confirmadas por el Modelo Centinela de enfermedades respiratorias como COVID-19. Días contados a partir del 8 de abril con datos de la Tabla 3.1.

sin embargo, en este texto tomaremos los logaritmos naturales de la segunda columna de la Tabla 3.1. Es decir, si hacemos la siguiente transformación:

$$z = \ln(y), \quad (3.2)$$

$$= \ln(Ae^{mx}), \quad (3.3)$$

$$= \ln(A) + mx. \quad (3.4)$$

Observemos como la expresión para z se comporta como una recta con pendiente m y ordenada al origen $\ln(A)$.

Veamos el resultado de esta transformación de variable graficando en la figura 3.2, pero ahora con los logaritmos de los casos confirmados.

Ahora, claramente, los datos se alinean y nuestros ojos pueden imaginar una recta pasando por la mayoría de los datos. Esta línea imaginaria reafirma que nuestra hipótesis de tener una relación exponencial parece adecuada.

Entonces podemos hacer nuevamente la regresión lineal, pero ahora con $z = \ln(y)$ en lugar del número de casos. Es decir en la columna o en el arreglo (array) se colocan los valores del logaritmo natural de los casos. Al usar las herramientas para hacer el cálculo de regresión lineal nos arrojarán

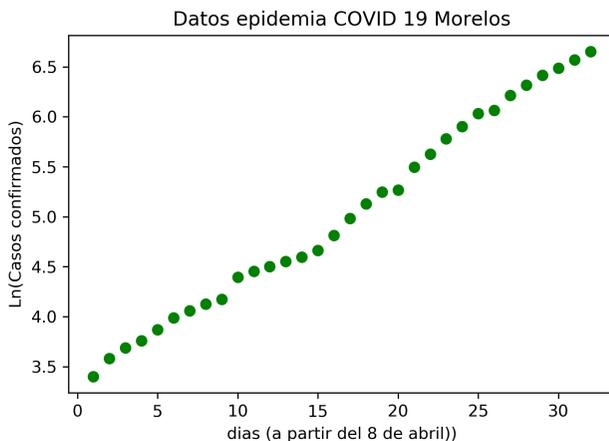


Figura 3.2: Datos de personas confirmadas por el Modelo Centinela de enfermedades respiratorias como COVID-19. Días contados a partir del 8 de abril 3.1.

una $m = 0.1059$ y una $b = 3.2767$. Entonces la expresión matemática del resultado de hacer esta regresión lineal en forma similar a la desarrollada para la sección anterior es

$$z = 0.1059x + 3.2767. \quad (3.5)$$

Podemos comparar la recta que obtenemos y los logaritmos de los datos para observar si hay coincidencias. Vemos el resultado en la figura 3.3, donde claramente observamos que la línea coincide con la mayoría de los datos experimentales. Por lo tanto, podemos pensar que la relación empírica que proviene de los datos es adecuada para describirlos y con ella podríamos hacer predicciones.

Así, con este resultado parece que hemos construido una relación empírica que describe el crecimiento exponencial que los casos confirmados de la epidemia mostró en las primeras semanas. Hemos construido una relación empírica que tiene bases teóricas sustentadas por el crecimiento exponencial de los infectados en una epidemia en los primeros días.

Es importante regresar los datos a la expresión original, sacando los antilogaritmos o elevando e a los exponentes correspondientes, es decir:

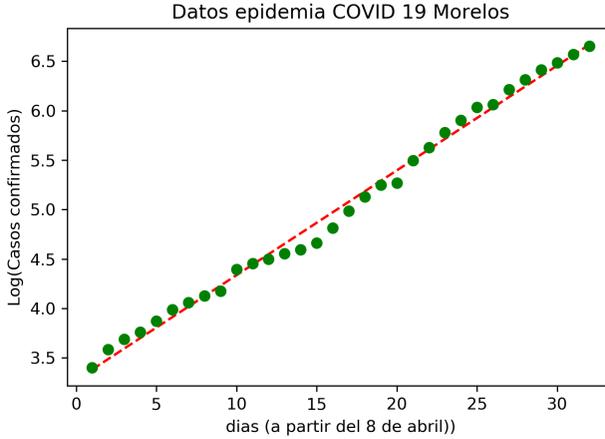


Figura 3.3: Datos de personas confirmadas por el Modelo Centinela de enfermedades respiratorias como COVID-19. Días contados a partir del 8 de abril de la Tabla 3.1.

$$y = e^z, \quad (3.6)$$

$$= e^{0.1059x+3.2767}, \quad (3.7)$$

$$= e^{3.2767} e^{0.1059x}, \quad (3.8)$$

$$= 26.49 \times e^{0.1059x}. \quad (3.9)$$

Así la relación empírica final es

$$y = 26.49 \times e^{0.1059x}. \quad (3.10)$$

Notemos que esta expresión es sencilla y solo necesita los dos parámetros que hemos calculado. Finalmente, en la figura 3.4 podemos ver la comparación del modelo ec. (3.10) con los datos confirmados, es decir con los datos sin transformación. Observamos que prácticamente todos los puntos están muy cerca de la línea roja que representa el modelo. Así, la tasa de crecimiento es $m = 0.1059$ y la constante o amplitud es $A = 26.49$. Observamos que la constante o valor inicial es muy similar al valor de la ordenada al origen de la aproximación lineal de la sección anterior (ver ec. (2.2)), el dato inicial parece claro, pero no así el resto del comportamiento.

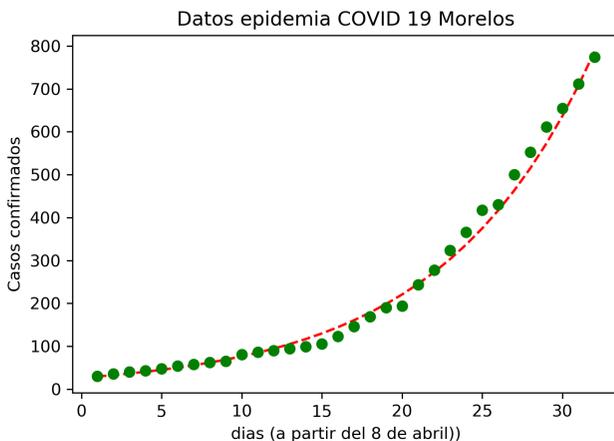


Figura 3.4: Comparación de los datos de personas confirmadas por el Modelo Centinela de enfermedades respiratorias como COVID-19 con un modelo exponencial. Días contados a partir del 8 de abril Tabla 3.1.

Regresemos a la conjetura final de la sección anterior y calculemos el número de infectados al día 100. Al usar la expresión (3.10) obtenemos 1,060,917 infectados a los 100 días, es decir la mitad de la población del estado. Recordemos, el modelo lineal predice solamente unos cuantos cientos, pero el exponencial predice un millón. Este último resultado muestra también las limitaciones del modelo exponencial para los casos epidemiológicos. De hecho, los modelos epidemiológicos muestran un comportamiento complejo donde una primera fase es de crecimiento lineal seguido de otro exponencial, pero luego las limitaciones de los contagios y lo finito de la población frenan este crecimiento exponencial. Nuevamente, tenemos que precisar cuáles son las condiciones en las que hemos basado el cálculo para obtener las ecuaciones empíricas y explicarlas con detalle en nuestras posibles conclusiones. Sin embargo, la descripción de estos modelos más complejos sale fuera de los objetivos de este texto introductorio.

En el siguiente capítulo también utilizaremos la regresión lineal para mediante un cambio de variable obtener una relación de potencia o relación potencial $y = Ax^m$.

Días (8 de abril)	Casos confirmados COVID-19 (número de personas)
1	30
2	36
3	40
4	43
5	48
6	54
7	58
8	62
9	65
10	81
11	86
12	90
13	95
14	99
15	106
16	123
17	146
18	169
19	190
20	194
21	244
22	278
23	324
24	366
25	417
26	430
27	500
28	553
29	611
30	655
31	712
32	774

Tabla 3.1: Personas confirmadas con COVID-19.

Capítulo 4

Relaciones potenciales

En este capítulo analizaremos otro caso de transformación de variable para obtener una ecuación empírica a partir de datos observacionales o experimentales.

Ahora analizaremos un conjunto de datos que provienen de un comercio. En la siguiente Tabla 4.1 se presentan los precios de venta de láminas cuadradas de acrílico. Aquí surge la pregunta ¿los precios están calculados de alguna forma?, ¿cuál es esa forma?

Longitud (m)	Precio (\$)
10	40
20	160
30	360
40	640
50	1000

Tabla 4.1: Lista de precios de las láminas cuadradas de acrílico en términos de su longitud.

De una inspección de la tabla no se observa un crecimiento lineal, pero ya sabemos qué hacer para alimentar nuestra intuición: grafiquemos para poder analizar.

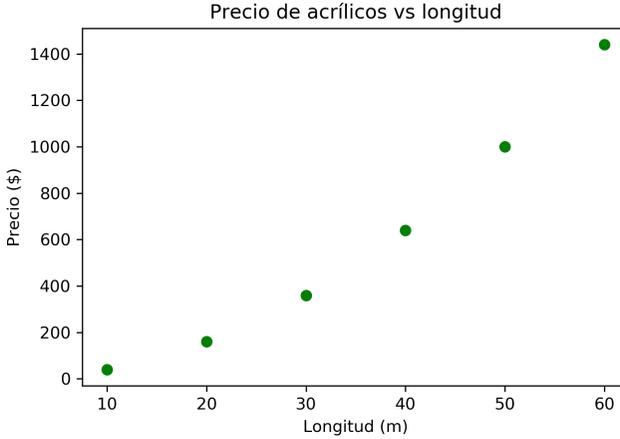


Figura 4.1: Precios en función del largo del acrílico, ver Tabla 4.1.

Efectivamente, en la figura 4.1 se puede apreciar que no hay una relación lineal, al tratar de trazar una línea recta que pase cerca de los puntos nuestros ojos no pueden hacerlo sin que se tuerza. Así que usemos la transformación de la sección anterior y transformemos los datos sacando los logaritmos de los precios, es decir $z = \ln(\text{precio})$ y graficamos z vs longitud (l). El resultado lo vemos en la fig. 4.2.

A diferencia del caso de la sección anterior, en esta ocasión no funcionó el cambio de variable. Nuestros ojos no pueden trazar una línea sobre los puntos transformados, pero transformemos el eje “ X ” también, es decir, sea $\xi = \ln(l)$ y grafiquemos z vs ξ . Los resultados se ven en la fig. 4.3.

Claramente, ahora los datos si pueden ser unidos por una línea recta imaginaria que nuestros ojos pueden construir pasando por los puntos. Entonces podemos ajustar una línea con los logaritmos de los datos (tanto para “ Y ” como para “ X ”). Para obtener la expresión lineal en los logaritmos utilizamos las mismas herramientas de las secciones pasadas. En esta ocasión tanto en la columna de las y como de las x usamos los valores de z y de x_i para hacer el cálculo de regresión lineal y obtener m y b dado que la expresión sería:

$$z = m\xi + b, \quad (4.1)$$

donde después de hacer la regresión lineal sabemos que $m = 2$ y $b = -0.916$. En esta ocasión tomamos el logaritmo natural, por lo tanto para regresar a

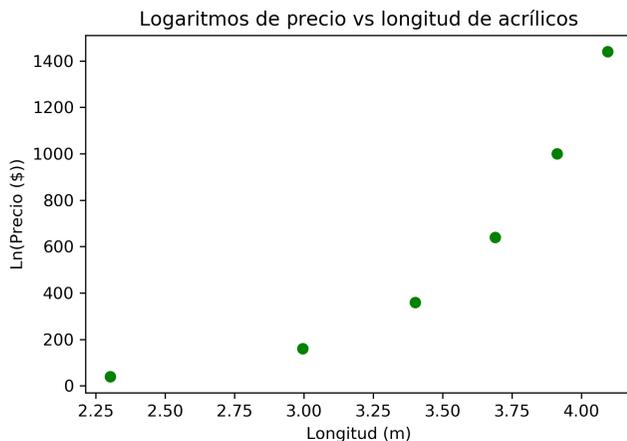


Figura 4.2: Logaritmos de los precios en función del largo del acrílico ver Tabla 4.1.

los datos originales tenemos que usar la función exponencial, es decir:

$$e^z = e^{m\xi+b}, \quad (4.2)$$

$$precio = e^b l^m, \quad (4.3)$$

$$precio = A l^m, \quad (4.4)$$

$$precio = (0.40)l^2. \quad (4.5)$$

Con esta expresión si podemos interpretar los resultados. La expresión l^2 se refiere al área de la lámina, ya que es una lámina cuadrada, y entonces $A = 0.4\$/m^2$ es el precio por metro cuadrado de acrílico. La gráfica final donde se comparan los datos originales con la descripción empírica (4.5) que encontramos se presenta en la figura 4.4 y se observa una buena correspondencia.

Esta última parte, donde se da significado a los parámetros es muy importante, ya que después de haber obtenido la expresión empírica podemos interpretarla desde los contextos físicos, de ingeniería, epidemiología, comercial, etc. Una ecuación empírica nunca estará completa sin la interpretación de los parámetros involucrados.

Aquí es esencial notar que la expresión (4.1) involucra una gran diversi-

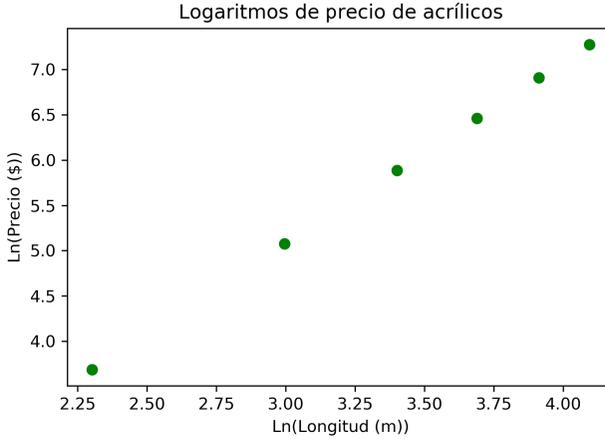


Figura 4.3: Logaritmos de los precios en función del logaritmo de los largos de los acrílicos ver Tabla 4.1.

dad de comportamientos. Por ejemplo, si $m > 0$ implica que la relación entre las variables es potencial y puede ser cúbica ($m = 3$) o cuártica ($m = 4$). Además en el caso de ser negativa ($m < 0$) se tiene comportamiento inverso, es decir que cuando una variable crece la otra decrece.

Con estos sencillos ejemplos hemos concluido este texto introductorio para el tratamiento de datos experimentales y obtención de ecuaciones empíricas.

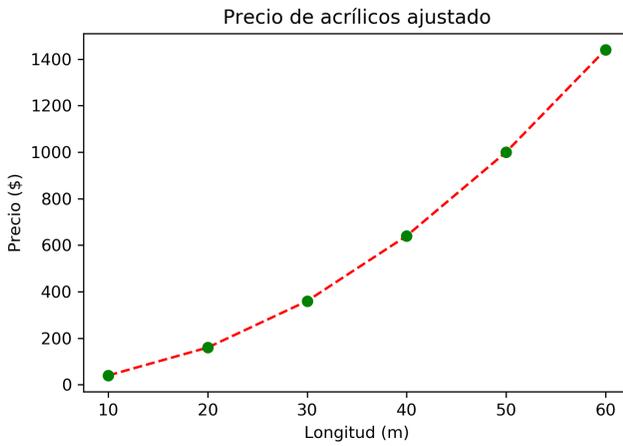


Figura 4.4: Comparación del modelo potencial y los precios observados ver Tabla 4.1.

Bibliografía

- [1] J.A. González “Gráficas y Ecuaciones Empíricas: Texto Programado” Centro de Enseñanza de la Física, Facultad de Ciencias (UNAM, 1977).
- [2] E. Galindo Fentanes “El quehacer de la ciencia experimental” (Siglo XXI, 2013).
- [3] Los datos fueron obtenidos de www.gob.mx/salud/documentos-abiertos-152127.
- [4] D.C. Baird “Experimentation : an introduction to measurement theory and experiment design” (Prentice Hall, 1995).
- [5] https://es.wikipedia.org/wiki/Regresi%C3%B3n_lineal
- [6] <https://github.com/jadelrio/empirical-relationships/blob/master/RelacionesLineales2.ipynb>
- [7] <https://github.com/jadelrio/empirical-relationships/blob/master/RelacioneExponenciales2.ipynb>
- [8] <https://github.com/jadelrio/empirical-relationships/blob/master/RelacionesPotenciales2.ipynb>
- [9] <https://github.com/jadelrio/empirical-relationships>

Capítulo 5

Apéndice

En este apéndice se presentan los códigos en Python de los programas que se usaron para la obtención de las ecuaciones empíricas del texto. También están disponibles en GitHub [9]

Programa Regresión lineal

En esta sección se presenta el programa en Python de regresión lineal disponible en GitHub [6].

```
"""J.A. del Rio 2022.
```

```
En esta programa se ilustra cómo usando una lista de datos se  
puede obtener una relación lineal. Primero se grafican para luego  
obtener la ecuación empírica a partir de ajustes de regresión  
lineal. El encabezado del código en Python, empieza con las  
siguientes instrucciones
```

```
"""
```

```
import matplotlib  
import matplotlib.pyplot as plt  
import numpy as np  
#matplotlib inline  
import math  
import cmath
```

```

"""Se defienen los arreglos. Utilizaremos los datos de casos
confirmados de COVID-19 del día 8 al 16 de Abril 2020 en Morelos.
Notemos que los tenemos que escribir en formato array"""

d= np.array([1, 2, 3, 4, 5, 6, 7, 8])
Casos=np.array([30., 36., 40., 43., 48., 54., 58., 62.])

"""Primero graficaremos los datos para intuir si hay una relación
lineal"""

plt.figure()
plt.plot(d,Casos,'go',markersize=6)
plt.xlabel('dias (8 de abril)')
plt.ylabel('Casos confirmados')
plt.title('Datos epidemia COVID 19 Morelos')
plt.savefig('epilindatos.png',dpi=300)
# se salva la figura en eps con 300 dpi

"""La inspección visual sugiere una relación lineal. Definiremos
una función regrelineal para obtener el ajuste de regresión lineal
de los datos """

# definimos la funcion regrelineal
# en ella hemos codificado las formulas de regresion lineal por
minimos cuadrados

def regrelineal(x, y):
    # numero de parejas de datos
    n = np.size(x)
    ny= np.size(y)
    if n != ny:
        return('no son parejas de datos')

    # promedio de los arreglos x y y
    p_x = np.mean(x)
    p_y = np.mean(y)

    # calculando la desviacion cruzada y desviacion en x
    SS_xy = np.sum(y*x) - n*p_y*p_x
    SS_xx = np.sum(x*x) - n*p_x*p_x

    # calculo de m y b

```

```
p1 = SS_xy / SS_xx
p0 = p_y - p1*p_x

# calculo de R cuadrada
error = y - (p1*x + p0)
se = np.sum(error**2)
SSt = np.sum((y - p_y)**2)
p2 = 1- (se/SSt)
# el resultado p0 es la ordenada al origen, p1 es la pendiente
y p2 es R cuadrada
return (p1, p0, R2)

# ajustamos la recta a los datos
p1, p0, R2 = regrelineal(d, Casos)
#valores de y con el ajuste
y = p1 * d + p0

plt.figure()
plt.plot(d,y,'r--',linewidth=2, markersize=6)
plt.xlabel('dias (8 de abril)')
plt.ylabel('Casos confirmados')
plt.title('Datos epidemia COVID 19 Morelos')

"""Observamos los puntos alineados de acuerdo a la ecuación para
y Obtenemos los valores de la ordenada al origen y la pendiente
de la línea roja"""

print(p0,p1)

"""Graficamos lo Casos ( verde) y el ajuste (rojo) """

plt.figure()
plt.plot(d,y,'r--', d,Casos,'go',markersize=6)
plt.xlabel('dias (8 de abril)')
plt.ylabel('Casos confirmados')
plt.title('Datos epidemia COVID 19 Morelos')
plt.savefig('epilin.png',dpi=300)
# se salva la figura en eps con 300 dpi

"""Con este ajuste podríamos predecir el número de casos a los
100 días"""
```

```
# en el dia 100 habria
print('en el dia 100 habria ',int(p1 * 100 + p0), 'casos')
#se manda a la pantalla combinacion de texto con lista

"""Como se podrá ver en el próximo ejercicio esta extrapolación
no es correcta. Ya que son muy pocos datos para observar un
comportamiento exponencial. La calidad del ajuste se observa
analizando el valor de R cuadrada. """

print('En este caso es',R2,'. Más cercano a 1 (-1) es más adecuado')
```

Programa regresión exponencial

En esta sección, se presenta el programa en Python de regresión exponencial usando las fórmulas para una regresión lineal. También están disponibles en GitHub [7].

```
# -*- coding: utf-8 -*-
"""J.A. del Rio 2022.
En esta programa se ilustra cómo usando una lista de datos se
puede obtener una relación exponencial. Primero se grafican y se
realiza una inspección en el formato gráfico para luego obtener
la ecuación empírica a partir de ajustes de regresión lineal.
"""

import matplotlib
import matplotlib.pyplot as plt
import numpy as np
#matplotlib inline
import math
import cmath

"""Se definen los arreglos. Utilizaremos los datos de casos
confirmados de COVID-19 del día 8 al 16 de Abril 2020 en Morelos.
Notemos que los tenemos que escribir en formato array
"""

d1=np.array([1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15,
16, 17, 18, 19,20,21,22,23,24,25,26,27,28,29,30,31,32])
Casos1=np.array([30, 36, 40, 43, 48, 54, 58, 62, 65, 81,
86, 90, 95, 99, 106, 123, 146, 169, 190, 194, 244, 278, 324,
366, 417, 430, 500, 553, 611, 655, 712, 774])

"""Ahora se ejemplificará un ajuste exponencial, pero primero
haremos una inspección visual de los datos"""

plt.figure()
plt.plot(d1,Casos1,'go',markersize=6)
plt.xlabel('dias (8 de abril)')
plt.ylabel('Casos confirmados')
plt.title('Datos epidemia COVID 19 Morelos');
plt.savefig('epiexpdatos.png',dpi=300)
```

```
"""Hacemos un cambio de variable con logaritmos naturales, se puede
hacer con logaritmos en cualquier base."""
```

```
ye=np.log(Casos1) # obtener el logaritmo de los datos
```

```
"""Graficamos los logaritmos de los casos contra los dias"""
```

```
plt.figure()
plt.plot(d1,ye,'go',markersize=6)
plt.xlabel('dias (8 de abril)')
plt.ylabel('Log(Casos confirmados)')
plt.title('Datos epidemia COVID 19 Morelos');
plt.savefig('epiexplogdatos.png',dpi=300)
```

```
"""Observamos que visualmente podemos unir a casi todos los puntos
con una recta imaginaria. Calculamos el ajuste de regresión lineal."""
```

```
# definimos la funcion regrelineal
# en ella hemos codificado las formulas de regresion lineal por
minimos cuadrados
```

```
def regrelineal(x, y):
    # numero de parejas de datos
    n = np.size(x)
    ny= np.size(y)
    if n != ny:
        return('no son parejas de datos')
```

```
    # promedio de los arreglos x y y
    p_x = np.mean(x)
    p_y = np.mean(y)
```

```
    # calculando la desviacion cruzada y desviacion en x
    SS_xy = np.sum(y*x) - n*p_y*p_x
    SS_xx = np.sum(x*x) - n*p_x*p_x
```

```
    # calculo de m y b
    p1 = SS_xy / SS_xx
    p0 = p_y - p1*p_x
```

```
    # calculo de R cuadrada
    error = y - (p1*x + p0)
```

```
se = np.sum(error**2)
SSt = np.sum((y - p_y)**2)
p2 = 1- (se/SSt)
# el resultado p0 es la ordenada al origen p1 es la pendiente y
p2 es R cuadrada
return (p1, p0, p2)

# ajustamos la recta a los logaritmos de los datos
q1, q0, q2 = regrelineal(d1, ye)
#valores de y con el ajuste
yye = q1 * d1 + q0

plt.figure()
plt.plot(d1,yye,'r--', d1,ye,'go',markersize=6)
plt.xlabel('dias (8 de abril)')
plt.ylabel('Log(Casos confirmados)')
plt.title('Datos epidemia COVID 19 Morelos');
plt.savefig('epiexp.png',dpi=300)

plt.figure()
plt.plot(d1,np.exp(yye),'r--', d1,Casos1,'go',markersize=6)
plt.xlabel('dias (8 de abril)')
plt.ylabel('Casos confirmados')
plt.title('Datos epidemia COVID 19 Morelos');
plt.savefig('epicompara.png',dpi=300)

"""observamos que el ajuste en los datos sin transformar tambien
parece adecuado. Los valores de la pendiente y la ordenada al
origen son"""

print(q1,q0, q2)

"""Lo valores de la tasa de crecimiento y de la condicion inicial
son:"""

print('La tasa de crecimiento es', q1 , 'la amplitud o el valor
inicial es',np.exp(q0))

print('La calidad del ajuste se observa con un valor',q2,'de R
cuadrada, mientras más cercano a 1 (-1) es mejor')
```

Programa regresión potencial

En esta sección se presenta el programa en Python de regresión potencial usando las fórmulas para una regresión lineal. También están disponibles en GitHub [8].

```
# -*- coding: utf-8 -*-
"""J.A. del Rio 2022.
En esta programa se ilustra cómo usando una lista de datos se
puede obtener una relación de potencia a partir de datos
experimentales. Primero se grafican para luego obtener la ecuación
empírica a partir de ajustes de regresión lineal. Para ello se usa
la transformación de datos mediante la función logaritmo.
"""

import matplotlib
import matplotlib.pyplot as plt
import numpy as np
#matplotlib inline
import math
import cmath

"""Definición de los datos"""

lon=[10, 20, 30, 40, 50, 60]
precio=[40, 160, 360, 640, 1000, 1440]

"""grafica de los datos"""

plt.figure()
plt.plot(lon,precio,'go',markersize=6)
plt.xlabel('Longitud (m)')
plt.ylabel('Precio ($)')
plt.title('Precio de acrílicos vs longitud')
plt.savefig('acrili.png',dpi=300)
# se salva la figura en eps con 300 dpi

"""transformación de los datos en el eje Y"""

precioln=np.log(precio)

plt.figure()
```

```
plt.plot(lon,precioln,'go',markersize=6)
plt.xlabel('Longitud (m)')
plt.ylabel('Log(Precio ($))')
plt.title('Logaritmos de Precio de acrílicos')
plt.savefig('acrililn.png',dpi=300)
# se salva la figura en eps con 300 dpi

"""transformación de los datos en el eje X"""

lonln=np.log(lon)

plt.figure()
plt.plot(lonln,precio,'go',markersize=6)
plt.xlabel('Log(Longitud (m))')
plt.ylabel('Precio ($)')
plt.title('Logaritmos de Precio vs longitud de acrílicos')
plt.savefig('acrililn.png',dpi=300)
# se salva la figura en png con 300 dpi

plt.figure()
plt.plot(lonln,precioln,'go',markersize=6)
plt.xlabel('Log(Longitud (m))')
plt.ylabel('Log(Precio ($))')
plt.title('Logaritmos de Precio de acrílicos')
plt.savefig('acrililnln.png',dpi=300)
# se salva la figura en png con 300 dpi

"""La inspección visual sugiere una relación lineal. Definiremos
una función regrelineal para obtener el ajuste de regresión lineal
de los datos transformados en X y Y"""

# definimos la funcion regrelineal
# en ella hemos codificado las formulas de regresion lineal por
minimos cuadrados
def regrelineal(x, y):
    # numero de parejas de datos
    n = np.size(x)
    ny= np.size(y)
    if n != ny:
        return('no son parejas de datos')

    # promedio de los arreglos x y y
```

```

p_x = np.mean(x)
p_y = np.mean(y)

# calculando la desviacion cruzada y desviacion en x
SS_xy = np.sum(y*x) - n*p_y*p_x
SS_xx = np.sum(x*x) - n*p_x*p_x

# calculo de m y b
p1 = SS_xy / SS_xx
p0 = p_y - p1*p_x

# calculo de R cuadrada
error = y - (p1*x + p0)
se = np.sum(error**2)
SSt = np.sum((y - p_y)**2)
p2 = 1- (se/SSt)
# el resultado p0 es la ordenada al origen p1 es la pendiente y
p2 es R cuadrada
return (p1, p0, p2)

# ajuste lineal a los datos transformados
m1, b0, R2 = regrelineal(lonln, precioln)
#valores de y con el ajuste
logP= m1 * lonln + b0

plt.figure()
plt.plot(lonln,logP,'r--', lonln, precioln,'go',markersize=6)
plt.xlabel('Log(Longitud (m))')
plt.ylabel('Log(Precio ($))')
plt.title('Logaritmos de Precio de acrílicos')
plt.savefig('acrililnajus.png',dpi=300)
# se salva la figura en png con 300 dpi

precioajuste=np.exp(logP)

""""Resultado final""""

plt.figure()
plt.plot(lon,precioajuste,'r--', lon, precio,'go',markersize=6)
plt.xlabel('Longitud (m)')
plt.ylabel('Precio ($)')
plt.title('Precio de acrílicos ajustado')

```

```
plt.savefig('acriliaajuste.png',dpi=300)
# se salva la figura en eps con 300 dpi

"""Obtención de los parámetros de ajuste"""

print('La potencia es',m1,'la constante multiplicativa o valor
inicial es ',np.exp(b0))

print('La bondad del ajuste se observa en el valor',R2,' mientras
más cercano a 1 (-1) es más adecuado el ajuste.')

print('La ordenada al origen es ',b0)
```