

Unai Clavier
Miguel del Val
Martin Ybarra

2^{do} de Bachiller

Lycée Français de Bilbao

Tutora : Sandrine Romera

**¿ Qué herramientas estadísticas nos permiten
buscar un modelo
para el resalto hidráulico ?**

Resumen

El proyecto presentado en este dossier es el de entender y utilizar herramientas estadísticas adecuadas para el análisis de un fenómeno empírico. Gracias a un procedimiento experimental que acordamos, recopilamos una nube de puntos que quisimos analizar con la intención de buscar cuál sería la mejor relación matemática entre el radio del círculo formado por el resalto hidráulico y el caudal del agua saliendo del grifo de nuestros hogares.

Abstract

The project presented in this dossier is based on the comprehension and correct application of some statistical tools in an empirical phenomenon. Thanks to an agreed-upon experimental procedure, we had collected a scatter plot that we wished to analyze, with the intention of searching which would be the best mathematical relationship between the radius of the circle formed by hydraulic jump and the flow of water that comes out the tap of our homes.

Índice:

Introducción

- I. Procedimiento experimental**
- II. Recopilación de datos**
 - 1. Obtención de los radios**
 - 2. Gráfico del radio en función del caudal**
- III. Herramientas para el análisis de los datos**
 - 1. Aproximación de una nube de puntos por el método de los mínimos cuadrados**
 - 2. Varianza y varianza residual de una regresión**
 - 3. Coeficiente de determinación**
 - 4. Ejemplo**
- IV. Aplicación del análisis estadístico a nuestro problema**
- V. Posibles mejoras y extensiones**

Conclusión

Introducción

Leonardo da Vinci manifestó, en el siglo XVI, curiosidad por un fenómeno habitual y cotidiano, que la mayoría de las personas experimentamos a diario: el resalto hidráulico (*imagen n°1*). Este fenómeno que se observa en el fregadero cuando se abre el grifo, consiste en el súbito aumento del nivel del agua, formando un círculo bien visible. Hay que esperar hasta el año 1993 para que un grupo de científicos se interesara en buscar una relación entre el radio del círculo formado por el resalto y el caudal del agua que lo provoca. También se estudió por Rajesh Bhagat, estudiante de doctorado en física de la universidad de Cambridge, que trató de mejorar el modelo anterior en el 2018. Los científicos afirman hoy que el origen del resalto está basado en un cambio de energía a energía potencial, la cual genera la subida del nivel de agua. La teoría que describe la transferencia de energía suscita aún mucho debate sin llegar a poner a los diversos equipos de investigadores definitivamente de acuerdo.

Nuestro objetivo en este trabajo es, a partir de un procedimiento experimental, determinar con las herramientas de estadísticas apropiadas cuál es la relación más adecuada entre el radio del círculo formado por el resalto hidráulico y el caudal del agua. Así podremos, con material rudimentario y un protocolo casero, jugar a imitar a los grandes investigadores universitarios.



Imagen n°1 : ilustración del resalto atribuida a Leonardo da Vinci (cuadernos de la National Gallery London)

I. Procedimiento experimental

Para tener datos estadísticos, hemos decidido hacer nosotros mismos los experimentos, siguiendo un protocolo determinado y estricto, que cada uno de los tres miembros del grupo hemos seguido escrupulosamente.

El material necesario se compone, como se puede ver en la *imagen 1.1*, de:

- un cronómetro
- una báscula con precisión de 0,1g
- un recipiente
- una tabla
- una regla
- un grifo de agua



Imagen 1.1: Fotografía del montaje experimental.

Los pasos a seguir en el procedimiento experimental son:

1. Poner el recipiente sobre la báscula y tarar.
2. Abrir el grifo.

Poner el recipiente bajo el agua y simultáneamente encender el cronómetro.

3. Esperar un tiempo a que el recipiente se llene en parte, y retirar el recipiente a la vez que se para el cronómetro.
4. Sacar una fotografía en la que se vea claramente el círculo formado por el resalto hidráulico (*imagen 1.2*).
5. Apuntar los datos relativos al peso del agua recogida y al tiempo necesario para recogerla.
6. Repetir 20 veces, cambiando la apertura del grifo para modificar el caudal del agua.

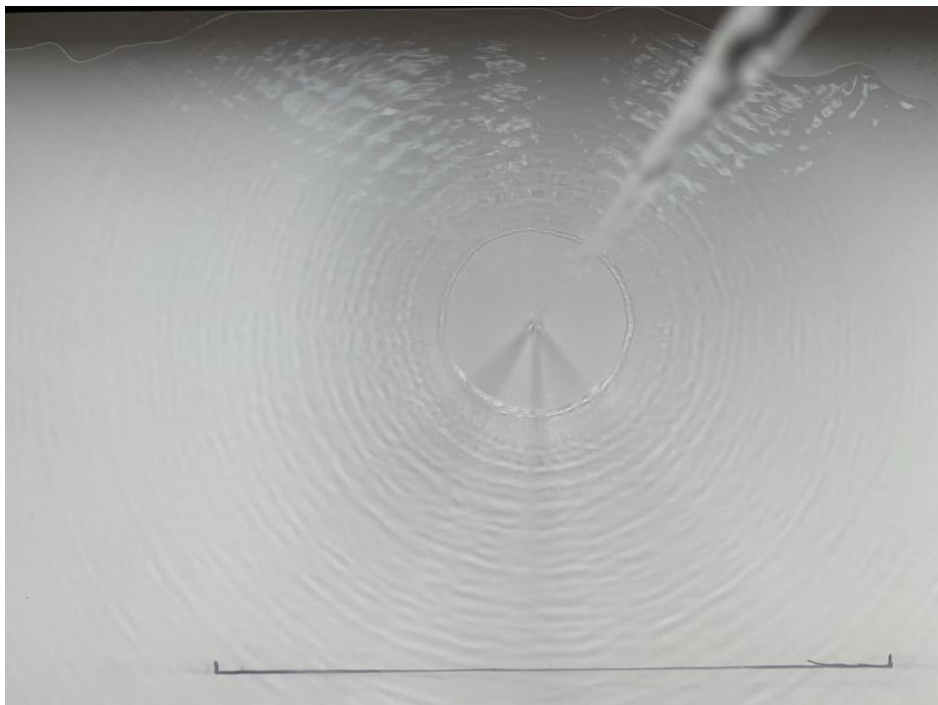


Imagen 1.2: Fotografía del salto hidráulico.

II. Recopilación de los datos

1. Obtención de los radios

A partir de las fotografías obtenidas, necesitamos medir el radio del círculo formado por el resalto. Hemos acordado usar, entre varios, el programa *freeware* Perfect Screen Ruler (*imagen II.1*).

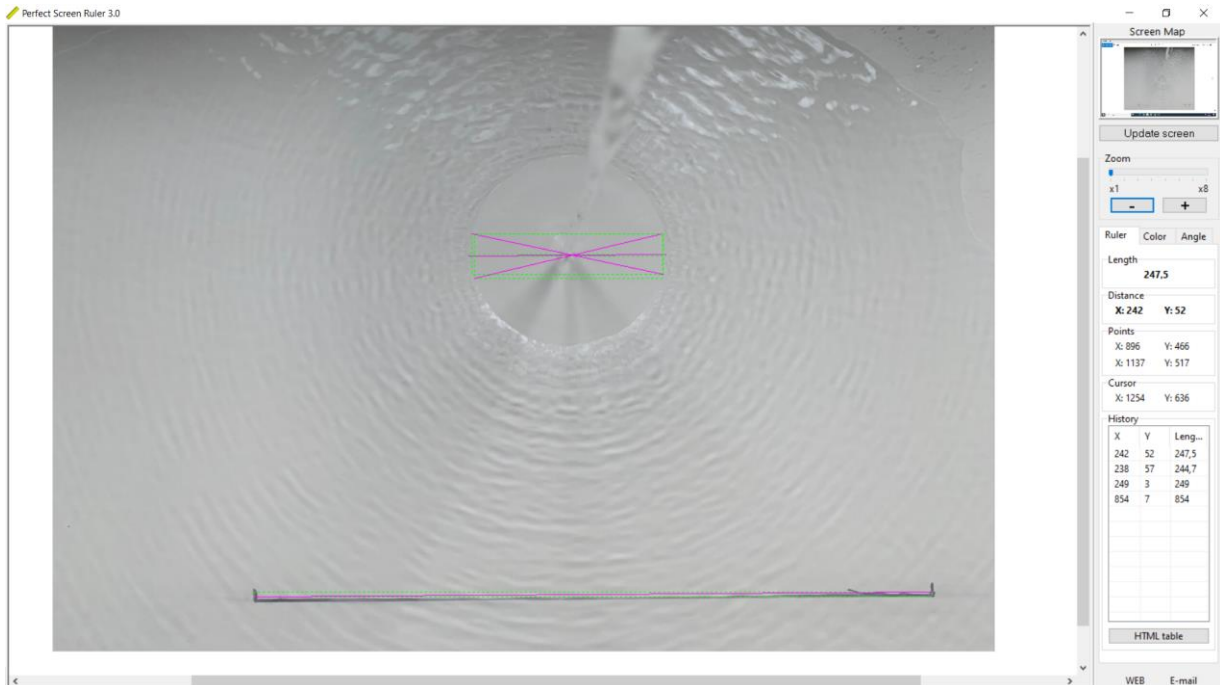


Imagen II.1: manipulación del programa Perfect Screen Ruler

Para poder medir los radios fotografiados, en primer lugar medimos el segmento dibujado en la plancha, y así tener el equivalente entre centímetros y píxeles del programa.

En segundo lugar, hacemos varias mediciones del diámetro del círculo creado por el resalto. Sacamos la media aritmética de esas mediciones para determinar un diámetro lo más correcto posible.

Finalmente, con una sencilla regla de tres, calculamos el diámetro en centímetros y obtenemos el radio dividiéndolo por dos.

Por ejemplo en el caso de la *imagen II.1*:

- El segmento para calibrar es de 30 cm correspondiendo a 854 píxeles.
- Las medidas del diámetro en píxeles son de 247.5, 244.7 y 249 cuya media es de 247.1.
- Estimación del radio : $\frac{30}{854} \times 247.1 \div 2 \approx 4.34$ cm

Así conseguimos medir unos 60 radios correspondientes a diferentes caudales de agua, de tal manera que disponíamos de datos experimentales para tratar de modelizar el fenómeno del resalto, para buscar una relación entre el caudal del agua y el radio del círculo formado.

2. Gráfico de los radios en función del caudal

Los radios obtenidos con el método explicado anteriormente son introducidos en una tabla Excel por cada participante (*imagen II.2*). De esta manera, podemos crear un gráfico en el que obtenemos el radio del círculo formado por el resalto en función del caudal de agua expresado en $\text{cm}^3 \text{s}^{-1}$.

Caudal en cm^3/s	Radio en cm	Caudal en cm^3/s	Radio en cm
11,05	1,7	66,93	6,6
11,74	1,5	67,05	6,7
14,05	2,0	68,91	7,1
17,86	2,7	69,22	6,6
20,46	3,5	71,4	6,7
23,9	2,6	78,3	6,8
24,76	2,9	79,49	7,2
24,78	3,9	85,3	7,8
25,27	4,0	94,16	6,2
29,76	4,1	104,45	7,0
29,9	4,4	106,9	7,0
30,04	4,3	110,1	6,8
33,46	4,4	118,9	7,3
38,25	4,9	124,6	7,7
40,02	4,5	135,7	7,8
46,44	4,5	142,23	7,8
50,81	4,8	164,93	8,3
54,61	5,6	177,1	8,5
57,93	6,6	191,48	8,6
59,3	6,6	203,79	8,7
63,7	6,7	218,14	9,0
64,05	7,4	226,45	9,1
64,4	6,7	232,32	9,4

Imagen II.2: Tabla Excel con los datos experimentales



Imagen II.3: Gráfico del radio en función del caudal de agua

Para interpretar la curva obtenida en la *imagen II.3*, encontramos en el programa Excel herramientas de base de la estadística que vamos a intentar profundizar en el párrafo siguiente.

III. Herramientas para el análisis de los datos

1. Aproximación de una nube de puntos por el método de los mínimos cuadrados

El método de los mínimos cuadrados es un procedimiento de análisis numérico en el que, dados un conjunto de datos, se intenta determinar la función continua que mejor se aproxime a los datos. Busca minimizar la suma de cuadrados de las diferencias ordenadas entre los puntos generados por la función (valores teóricos) y los correspondientes datos (valores experimentales).

Por ejemplo, si se quiere determinar la recta que se ajuste lo mejor posible a unos datos, como se ve en la *imagen III.1*, se busca su ecuación: $y = mx + b$

Los números m y b se obtienen buscando cuáles de los valores minimizan la suma de los cuadrados. Las fórmulas son:

$$m = \frac{\Sigma(xy) - \frac{(\Sigma x)(\Sigma y)}{N}}{\Sigma(x^2) - \frac{(\Sigma x)^2}{N}} \quad \text{y} \quad b = \bar{y} - m\bar{x},$$

siendo \bar{y} y \bar{x} las medias de las variables y y x respectivamente, y siendo N el número de observaciones.

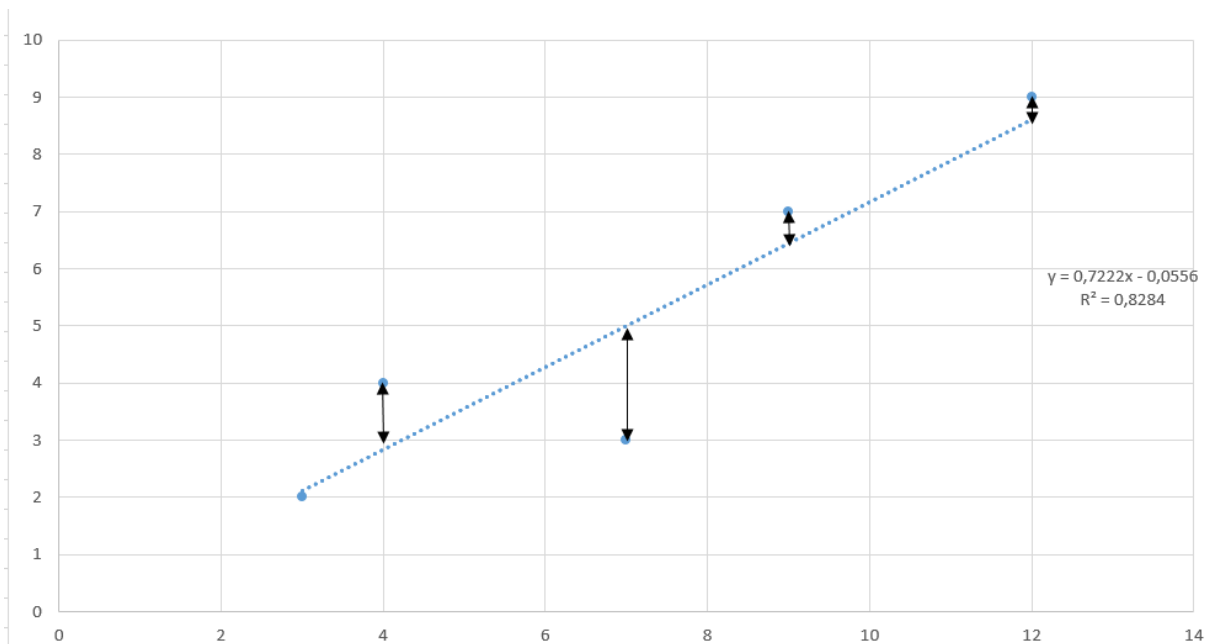


Imagen III.1: Ejemplo de aproximación lineal de una nube de puntos con el método de los mínimos cuadrados

Tras proponer una metodología para aproximar una nube de puntos, la estadística también se preocupa de medir las dispersiones alrededor del modelo.

2. Varianza y varianza residual de una regresión

La dispersión de una variable X está relacionada con su **varianza**, que es una medida de la distancia de cada uno de los datos al valor medio de X . Una varianza cero supone que todos los datos coinciden con la media y por lo tanto todos los datos son iguales.

Se define la varianza como la media de los cuadrados de las diferencias entre cada dato y la media de la variable: $\sigma_X^2 = \frac{\sum(x_i - \bar{X})^2}{N}$, siendo x_i el valor número i de la variable x , \bar{X} la media de la variable X y N el número de observaciones.

Cuando se quiere medir la dispersión de unos puntos ($X_i ; Y_i$) respecto a una línea de regresión definida por ($X_i ; \hat{Y}_i$) se usa la **varianza residual** definida por: $V_r(Y) = \frac{\sum_i (Y_i - \hat{Y}_i)^2}{N}$ con las notaciones definidas anteriormente.

La diferencia ($Y_i - \hat{Y}_i$) mide el error, también llamado residuo, que cometemos al sustituir el valor observado Y_i por el valor estimado mediante la regresión \hat{Y}_i .

Un valor elevado de la varianza residual significa que el error residual es importante y por lo tanto que la ecuación propuesta en la regresión es poco coherente.

El concepto de varianza residual permite afinar la herramienta de medida de dispersión con el coeficiente de determinación.

3. Coeficiente de determinación

El **coeficiente de determinación** de una regresión se calcula comparando a 1 el cociente de la varianza residual por la varianza de los valores experimentales.

Se calcula mediante la fórmula: $R^2 = 1 - \frac{\sum(Y_i - \hat{Y}_i)^2}{\sum(Y_i - \bar{Y})^2}$

R^2 es el indicador que permite juzgar la calidad de una regresión, medir la adecuación del modelo a las observaciones. Varía entre 0 y 1, y cuanto más cerca de 1 se sitúe su valor, mayor será el ajuste del modelo a la variable cuyo comportamiento estamos intentando describir.

4. Ejemplo

Para entender mejor el método de los mínimos cuadrados, y el cálculo del coeficiente de determinación, hemos hecho cada paso “a mano” (pero con la hoja de cálculo) y hemos comprobado que el valor que obtenemos para R^2 es el mismo que el valor obtenido directamente desde el programa Excel.

En la *imagen III.1*, podemos observar la regresión lineal propuesta por Excel con una muestra ficticia de puntos, y $R^2 = 0,8284$.

En la *imagen III.2*, vemos el detalle del cálculo de R^2 :

- en las columnas B y C están los valores de X_i e Y_i , con sus medias respectivas \bar{X} e \bar{Y} en la línea 9.
- en la columna D se calculan los valores Y_t obtenidos a través de la ecuación de la recta.
- en la columna F está la diferencia entre D y C al cuadrado, y la suma de estas diferencias en la línea 9.
- en la columna G están las etapas del cálculo de la varianza de Y .
- por último, para calcular el R^2 solo queda aplicar $1 - \frac{F_9}{G_9}$.

	B	C	D	E	F	G
1						
2	Valor Xi	Valor Yi	Valor Yt en la recta		(Yi-Yt)^2	(Yi - Y)^2
3	3	2	2,1110		0,0123	9
4	4	4	2,8332		1,3614	1
5	7	3	4,9998		3,9992	4
6	9	7	6,4442		0,3089	4
7	12	9	8,6108		0,1515	16
8						
9	X = 7	Y = 5			5,83	34,00
10						
11					Coef R^2	0,8284

Imagen III.2: Tabla del cálculo de los mínimos cuadrados

Así hemos comprobado que nuestro valor de R^2 coincide con el de Excel, y hemos entendido mejor lo que representaba.

IV. Aplicación del análisis estadístico a nuestro problema

En la *imagen IV.1* podemos observar la propuesta del programa Excel de una aproximación mediante una función lineal. El R^2 correspondiente es de **0,9134**.



Imagen IV.1: aproximación de la nube de puntos por una función lineal.

En la *imagen IV.2* tenemos una aproximación por función polinómica de segundo grado. El R^2 es de **0,9396** y al ser más cercano a 1 que el anterior podemos concluir que es un mejor modelo.



Imagen IV.2: aproximación de la nube de puntos por una función cuadrado.

En el tercer caso que se ilustra en la *imagen IV.3*, se aproxima la nube de puntos por una función de tercer grado que como podemos observar se acopla aún mejor, por lo que el R^2 correspondiente de **0,9541** se acerca aún más del 1 que en los otros casos.



Imagen IV.3: aproximación de la nube de puntos por una función de tercer grado.

Con una función de cuarto grado (*imagen IV.4*) obtenemos un R^2 de **0,9603**. Así pues, podemos comprobar que una función del cuarto grado se aproxima mejor a nuestra nube de puntos que las demás.



Imagen IV.4: aproximación de la nube de puntos por una función de cuarto grado.

V. Posibles mejoras y extensiones

Después de haber entendido y profundizado las herramientas disponibles en Excel, nos hemos preguntado si nuestro modelo era pertinente, ya que cada vez que probábamos modelizar con un polinomio de grado superior aumentaba el R^2 . Buscando en internet, encontramos el estudio de un equipo de Cambridge donde el modelo expresa el radio del círculo como proporcional al caudal a la potencia $3/4$. Este dato nos llevó a poner en cuestión la única medida por el R^2 de la pertinencia de un modelo. Encontramos entonces que había más herramientas estadísticas, como el coeficiente de determinación ajustado, el índice AIC o el índice BIC, pero no pudimos profundizar ésta parte por falta de tiempo, y por la dificultad que presentan estas otras herramientas.

También nos preguntamos cómo encontrar una función potencia decimal, y hemos intentado escribir un programa que determine la potencia decimal más adecuada, para ver si nos acercábamos un tanto a los resultados del equipo de Cambridge. Desafortunadamente, las bases que tenemos de programación con Python no nos permitieron llevar a cabo éste programa. En la *imagen V.1* aparece una propuesta inacabada para buscar el mejor valor de a en una aproximación evaluada por los mínimos cuadrados del tipo $f(x) = x^a$ siendo a un número real entre 0 y 1, que evaluamos con un paso de 0,01. Los nombres de las variables están en adecuación con los de la *imagen III.2*.

```
#Primero, insertamos los datos en el programa#
from google.colab import files
uploaded=files.upload()
#A partir de los datos, se deben de crear los valores teoricos#
mediayi=stat.mean(yi)
yt=[]
F=[]
G=[]
R2=[]
R2min=0
amin=0
for a in range (0,1,0.01):
    yt.append(x[i]**a)
    #Ahora, realizamos la técnica de los mínimos cuadrados#
    F=[(yi[i]-yt[i])**2]
    sumaF=sum(F)
    G=[(yi[i]-mediayi)**2]
    sumaG=sum(G)
    R2=[1-(sumaF/sumaG)]
    if R2>R2min:
        R2min=R2
        amin=a
#Mostramos los valores de a y de R2 más adaptados#
print(R2min)
print(amin)
```

Imagen V.1: propuesta de programa en Python

Conclusión

Con este trabajo, hemos descubierto unas herramientas estadísticas que permiten buscar una descripción matemática de un fenómeno empírico: el resalto hidráulico. Hemos profundizado el conocimiento que teníamos de la varianza de una variable, y hemos descubierto y entendido que el método de los mínimos cuadrados permite encontrar las ecuaciones que modelizan una nube de puntos. También hemos estudiado la noción de coeficiente de determinación, herramienta clave para medir cual de varias aproximaciones es la mejor. Aplicando estos conceptos y métodos a nuestro experimento, nos hemos acercado a cómo los científicos llegan a revelar una ley de la física, aproximando siempre con mejor precisión el modelo que describe la observación de la experiencia. Pero también llegamos a criticar las herramientas básicas que propone Excel, ya que nuestras conclusiones eran poco coherentes a pesar de tener un coeficiente de determinación muy elevado. Por falta de tiempo no pudimos estudiar otras herramientas que tal vez convenían mejor.

A pesar de cierta frustración por no llegar a un modelo en adecuación con la propuesta de Cambridge, y por la falta de tiempo para mejorar nuestro estudio, este proyecto fue muy satisfactorio porque, a partir de un protocolo sencillo y totalmente casero, hemos estudiado un fenómeno que interesó a grupos de investigadores de universidades prestigiosas como la de Cambridge o el MIT. La sensación grata de jugar en el patio de los grandes nos motivó mucho, no solo para llevar a cabo este trabajo, sino también para proyectarnos en la vida universitaria que emprenderemos los tres en la facultad de tecnologías y ciencias a partir del próximo septiembre.