APROXIMACIÓN POR MÍNIMOS CUADRADOS

"No manipules tus datos, pues ellos podrían estar correctos"
– Wilbur Wright

Introducción

En la investigación científica es esencial estudiar las relaciones entre distintos fenómenos con el objetivo de descubrir conexiones de causa-efecto; así como la relación entre diferentes parámetros o indicadores para determinar si se puede conocer el estado de un sistema según una variable relacionada al mismo. Esto conlleva una apreciable cantidad de trabajo a realizar: Mediciones y observaciones experimentales que raramente se comportan de manera exacta según una fórmula. Uno de los recursos del experimentador es lograr hallar la fórmula que mejor se aproxime al comportamiento mostrado por las mediciones.

Por ejemplo, en 2015 se publicó un estudio sobre el efecto tóxico que tiene la concentración de nitratos en los cultivos del camarón de cola blanca (*Litopenaeus Vannamei*), sobre el crecimiento de los mismos. Se determinó que la tasa de crecimiento semanal G (en gr/semana) decrece a medida que se incrementa la concentración de nitratos N (en mg/L), según la expresión:

$$G = 0,874 - 0,0007 \times N$$

La Figura 1 muestra uno de los resultados de ese estudio:

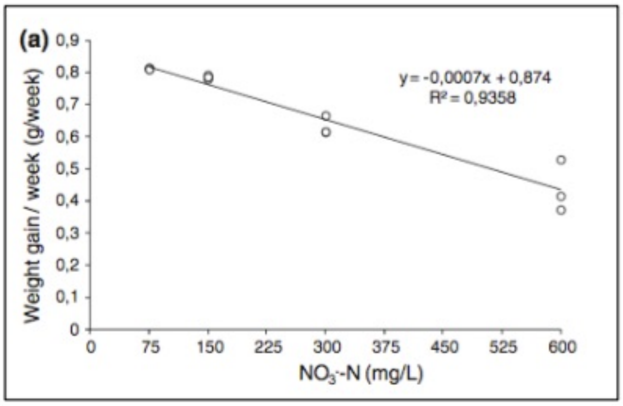

Figura 1: Decrecimiento de la ganancia de peso por semana, según la concentración de nitratos en las piscinas de cultivo de camarón de cola blanca, *L. Vannamei*.

En Figura 1, los pequeños círculos indican los datos reales medidos, mientras que la línea recta corresponde a la función lineal que mejor se aproxima a los datos medidos, no necesariamente que contenga dichos puntos.

REGRESIÓN LINEAL

Suponga que se dispone de dos secuencias de mediciones **x** e **y**:

$$\mathbf{x}=\begin{bmatrix} x_1\\ x_2\\ \vdots\\ x_n\end{bmatrix},\quad \mathbf{y}=\begin{bmatrix} y_1\\ y_2\\ \vdots\\ y_n\end{bmatrix}$$

con correspondencia entre mediciones (a la i-ésima medida x_i le corresponde la medición y_i). Suponga que hay evidencia experimental de una relación lineal entre los vectores dados, de modo que se desea encontrar una expresión de tipo y = *m*x + *b* (lineal), cuya gráfica represente mejor el conjunto de puntos (x_i, y_i) de las mediciones tomadas.

Si la relación fuera exacta, y las mediciones perfectas y sin errores, se cumpliría que:

$$\begin{aligned} y_1 &= b+mx_1\\ y_2 &= b+mx_2\\ &\vdots\\ y_n &= b+mx_n \end{aligned}$$

pero en la vida real esto nunca se cumple, por lo cual, al evaluar los x_i en la recta y = *m*x + *b* se producen como resultado $\hat{y}_i$ "teóricos" o "predichos", que no son iguales a los y_i "medidos":

$$\begin{bmatrix} \hat{y}_1\\ \hat{y}_2\\ \vdots\\ \hat{y}_n\end{bmatrix}=\begin{pmatrix} 1 & x_1\\ 1 & x_2\\ \vdots & \vdots\\ 1 & x_n\end{pmatrix}\begin{bmatrix} b\\ m\end{bmatrix}$$

O, $\hat{\mathbf{y}}=\mathbf{Au}$ en su forma matricial. ¿Cuáles son los valores de los parámetros *m* y *b* que producen la menor distancia entre los y_i y los $\hat{y}_i$? Es decir, que minimizan la distancia entre $\mathbf{y}$ y $\hat{\mathbf{y}}$, $\left\|\mathbf{y}-\hat{\mathbf{y}}\right\|$.

La expresión $\hat{\mathbf{y}}=\mathbf{Au}$ corresponde a un sistema de ecuaciones lineales, con $A_{n\times 2}$, y donde $\hat{\mathbf{y}}$ es un vector de $\mathbb{R}^n$ que pertenece a la imagen de A. Así también, Im(A) es un subespacio de dimensión menor o igual que 2, pues A solo tiene dos columnas. Según el teorema de la Aproximación de la Norma, la distancia $\left\|\mathbf{y}-\hat{\mathbf{y}}\right\|$ es mínima cuando $\mathbf{y}-\hat{\mathbf{y}}$ es perpendicular a Im(A).

Tarea 1: Enuncie y demuestre el Teorema de la Aproximación de la Norma.

Sea $\mathbf{u}=\bar{\mathbf{u}}$ el vector que minimiza $\left\|\mathbf{y}-\hat{\mathbf{y}}\right\|$, es decir, cuando se cumple que $A\mathbf{u}\perp\ \mathbf{y}-A\bar{\mathbf{u}}$, o:

$$\begin{gathered} \left\langle A\mathbf{u},\ \mathbf{y}-A\bar{\mathbf{u}}\right\rangle = 0\\ \left\langle A\mathbf{u},\ \mathbf{y}\right\rangle-\left\langle A\mathbf{u},\ A\bar{\mathbf{u}}\right\rangle = 0\\ \mathbf{u}^T(A^T\mathbf{y}-A^TA\bar{\mathbf{u}})=0 \end{gathered}$$

Lo que se cumple para cualquier $\mathbf{u}$ solo si $A^T\mathbf{y}=A^TA\bar{\mathbf{u}}$, en consecuencia:

$$\bar{\mathbf{u}}=(A^TA)^{-1}A^T\mathbf{y}$$

Tarea 2: Justificar por qué se afirma que necesariamente $A^T\mathbf{y}=A^TA\bar{\mathbf{u}}$.

Tarea 3: Deducir la formulación para la regresión cuadrática, cuando se quiere relacionar dos secuencias de mediciones $\mathbf{x}$ e $\mathbf{y}$ mediante la expresión $y_i = c + bx_i + ax_i^2$. ¿Qué forma tienen en este caso las matrices $\mathbf{A}$ y $\mathbf{u}$ en la ecuación $\hat{\mathbf{y}} = \mathbf{A}\mathbf{u}$? ¿Cuál es la expresión para el vector $\mathbf{u}$ que minimiza la distancia entre $\mathbf{y}$ e $\hat{\mathbf{y}}$?

CASO DE ESTUDIO: EL VUELO DEL PRIMER AEROPLANO

El final del Siglo XIX e inicio del Siglo XX vio una carrera contra el tiempo de varios grupos de investigación para lograr el primer vuelo controlado de un aeroplano. Los principales contendores estaban Samuel Pierpont Langley, Alexander Graham Bell, y los hermanos Orville y Wilbur Wright. Langley se dedicó al problema de la propulsión, desarrollo de motores y la potencia requerida para el vuelo. Graham Bell estudió la estructura del fuselaje, proponiendo la utilización de tetraedros para construir un fuselaje estable. Los hermanos Wright trabajaron mucho en la componente de balance de fuerzas, el peso del aeroplano y de la fuerza de sustentación requerida.

A pesar de la eminente competencia, los hermanos Wright fueron los primeros en realizar un vuelo controlado de un aeroplano; pero, ¿por qué? Se revisará uno de los muchos problemas técnicos que se tuvieron que resolver para conseguir el primer vuelo controlado.

Inicia cuando se abandona el ala plana (Fig. 2) y se empezó a experimentar con un ala curvada (Fig. 3), precursora de las actuales alas modernas codificada como NACA1408 (Fig. 4):

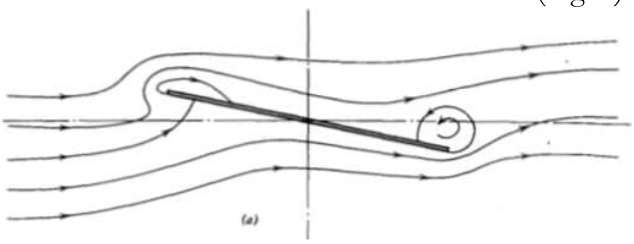

Figura 2: Ala plana [2].

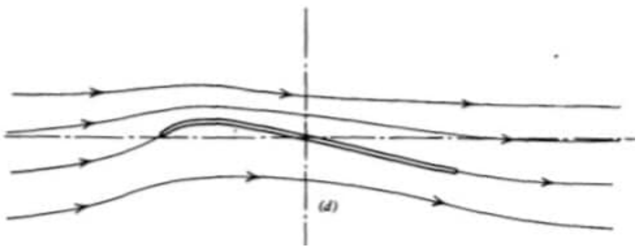

Figura 3: Ala curvada [2].

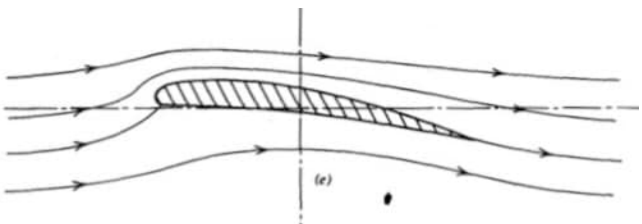

Figura 4: Ala NACA1408 actual [2].

Las características de estas alas eran estudiadas mediante sus curvas de coeficiente de sustentación **Ct**, el coeficiente de momento **Cm**, y el coeficiente de arrastre, todos medidos en función del ángulo de ataque α, es decir, la inclinación del ala respecto a la horizontal.

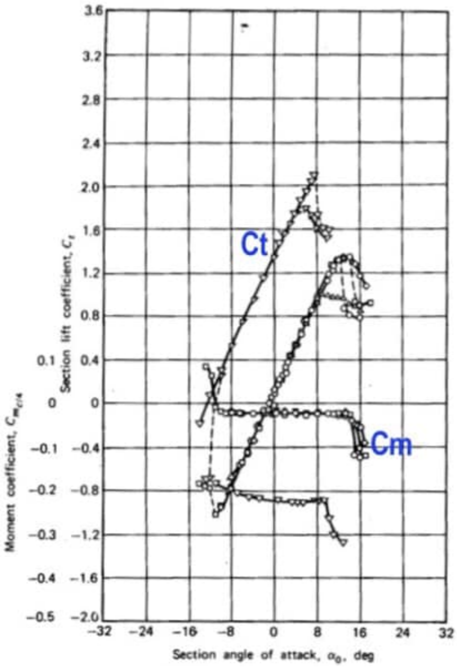

Figura 5: Características del ala curvada, NACA1408 [2]

Como se observa en Fig. 5, el coeficiente de sustentación **Ct**, se comporta de manera lineal durante la mayor parte de los ángulos de ataque medidos. La forma curvada del ala hace que, aun a 0°, el coeficiente de sustentación sea positivo, mientras que es necesario un ángulo negativo para lograr una sustentación de cero. Aun era desconocido para la época que, al alcanzar un cierto ángulo de ataque, el flujo en la parte posterior del ala empieza a separarse de la misma y a formar turbulencias (Fig. 6). Esta turbulencia incrementa el coeficiente de arrastre, disminuyendo gradualmente la sustentación.

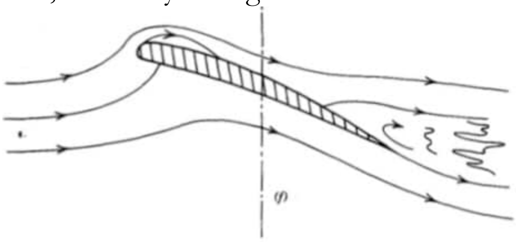

Figura 6: Turbulencia al final del ala; depende del ángulo de ataque y el número de Reynolds

Luego de alcanzar el punto máximo, la sustentación gradualmente cae a cero cuando el ángulo de ataque llega a los 90°. El grupo de trabajo de S. P. Langley notó esta pequeña disminución de **Ct** luego del máximo, pero decidió elegir una curva muy alisada ("suave") a través de sus datos. Para los hermanos Wright, esta curva tenía un máximo y, por lo tanto, un ángulo en el cual el ala produce una máxima sustentación. Según se evidencia por una conferencia dictada en 1901 [2], los hermanos Wright ya sabían que el ala curvada era mejor que el ala plana, y que los ángulos de ataque mejores estaban en un intervalo entre 5 y 12 grados.

Cuando los hermanos Wright vieron el trabajo de Langley, notaron que sus datos también mostraban la región de cambio de función creciente a decreciente, lo cual indicó que Langley no ajustó bien la curva a sus datos. Como escribiría Wilbur Wright al ingeniero Octave Chanute, el 1 de diciembre de 1901:

> "Si Langley hubiera atendido a sus observaciones, su curva habría estado probablemente más cerca de la verdadera. A mí también se me hace muchas veces difícil dejar que las líneas sigan su curso, en lugar de hacerlas pasar por donde yo creo que deben pasar. Mi conclusión es que es mucho más seguro seguir las observaciones exactamente como son, y que sean otros los que añadan sus propias correcciones si desean. [3]"

El 17 de diciembre de 1903, los hermanos Wright lograron con éxito el primer vuelo controlado de la historia. Lejos de ser un golpe de suerte, ellos habían construido un túnel de viento, y probado muchos diseños de alas, simulado vuelos mediante cometas; midiendo de manera metódica y obteniendo conclusiones con estricto apego al método científico. Los vuelos que Langley programó para octubre y diciembre de 1903 se precipitaron a tierra "como un saco de cemento" [4].

Aplique los métodos de aproximación por mínimos cuadrados, para obtener una expresión que permita calcular el coeficiente **Ct** según el ángulo de ataque, para el ala NACA1408 según los datos descritos en el archivo Excel adjunto, obtenido de [5] y [6]. Pruebe con polinomios de grado 1, 2, 3 , 4 y 5; y responda a las cuestiones planteadas en los entregables. En caso de ser de su interés, datos con otros números de Reynolds están disponibles en [7]; mientras que un calculador del Reynolds es asequible en [8].

Nota: Los datos se proveen en el archivo adjunto:
Proyecto_Aeronautica_Datos.xlsx.

REFERENCIAS:

[1] Grossman, S: Aplicaciones del Álgebra Lineal, Grupo Editorial Iberoamérica, México DF, pp. 124-132. México.

[2] McCormick, B. W: Aerodynamics, Aeronautics and Flight Mechanics, Wiley & Sons, New York, pp. 63-79.

[3] MacFarland, M. W: The papers of Wilbur and Orville Wright, including the Chanute-Wright letters. MacGraw-Hill, New York, 1953.

[4] Hansen, Liane: The race for flight. The National Public Radio, 2003. Recuperado en https://www.npr.org/templates/story/story.php?storyId=1469463 .

[5] Abbott, I. H., Von Doenhoff, A., and Stivers Jr., L. S: Summary of Airfoil Data. Report No. 824. National Advisory Committee for Aeronautics (NACA), 1945.

[6] Airfoil Tools: NACA 1408 Xfoil prediction polar with Reynold number 500000. Recuperado http://airfoiltools.com/polar/details?polar=xf-naca1408-il-500000

[7] Airfoil Tools: NACA 1408 http://airfoiltools.com/site/search?cx=partner-pub-2601928470036827%3A5136476429&cof=FORID%3A10&ie=UTF-8&q=NACA+1408&sa=Search&siteurl=airfoiltools.com%2Fairfoil%2Fdetails%3Fairfoil%3Dn0012-il&ref=&ss=0j0j1 .

[8] Airfoil Tools: Reynolds Number Calculator. Recuperado de: http://airfoiltools.com/calculator/reynoldsnumber

Entregables

Tenga presente lo siguiente:

- Se debe proveer un reporte con dos secciones. La primera parte se constituye de un marco teórico adecuado, junto con las soluciones de las tareas propuestas en el documento.
- En la segunda sección debe constar exclusivamente el análisis de las conexiones que existe entre

el coeficiente **Ct** y el ángulo de ataque α, proveyendo respuestas a las siguientes cuestiones:

1. Elabore un gráfico de los datos originales provistos (**Ct** versus α).
2. Decida en base al gráfico, cuál es la mejor relación polinomial (cuadrática, cúbica, etc) entre estas variables.
3. A partir de los cálculos hechos, provea una tabla con los coeficientes del polinomio que mejor se ajusta a los datos dados.
4. Grafique el polinomio hallado
5. Uno de los principales problemas filosóficos relacionados es la cuestión de la correlación y la causa-efecto. La frase "correlación no implica causación" acompaña a los científicos durante toda su vida como admonición. Explique en sus propias palabras y mediante ejemplos esta frase, y distinga entre los dos casos: la correlación y la causa-efecto.

NOTA: Es lícito apoyarse en la tecnología, pero si necesita utilizar un software o calculadora (Matlab®, Python, etc), o algún sitio de resolución de matrices online (Online Matrix Calculator de blue-bit), debe ser indicado en el documento a entregar, planteando la fórmula teórica, e indicando si se utilizó para resolver esa ecuación; y repitiendo la indicación para cada una de las ecuaciones así resueltas. Si consultó un libro o artículo, se debe incluir en una sección Bibliografía o Referencias del documento.