# LA FUNCIÓN LINEAL Y EL MODELO DE REGRESIÓN LINEAL

**Agnelli, Héctor; Konic, Patricia; Peparelli, Susana; Zón, Nora**
***Universidad Nacional de Río Cuarto – Argentina***
**e-mail:** speparelli@exa.unrc.edu.ar

## 1. INTRODUCCION

En los cursos de matemática se utiliza la función lineal como de uno de los ejemplos básicos del concepto de función. Dada la expresión de la función, $f_{(x)} = m\,x + b$, en general se pasa a asociar esta función con su representación gráfica: la recta. Nominados ahora b y m como la ordenada al origen y la pendiente, se pone de manifiesto que las rectas y por lo tanto las funciones lineales quedan individualizadas a partir del conocimiento de estos dos valores. Un ejercicio típico es dada las coordenadas de dos puntos hallar la recta que ellos determinan. De esta manera queda establecida la ecuación de la recta al "calcular" la pendiente y la ordenada al origen. El problema es de naturaleza determinística.

El modelo de regresión lineal aparece en un contexto distinto, se utiliza para modelar fenómenos aleatorios. Con el modelo se intenta describir utilizando el lenguaje matemático el comportamiento de ese fenómeno aleatorio. Para ello se recurre a conceptos o expresiones matemáticas que adquieren significados distintos cuando se usan en el campo de la estadística. Se tiene un conjunto de puntos que expresan los valores que asumen dos características reales medibles y se quiere encontrar la relación que las vincula. Si bien el procedimiento que se utiliza para ello es de naturaleza determinística su interpretación a priori es de naturaleza aleatoria ya que distintos conjuntos de puntos, que representan a las mismas características, podrían originar distintas rectas. Cuestiones tales como parámetros, estimadores, variables aleatorias, independencia y distintas interpretaciones de la linealidad hacen su aparición en este tipo de problemas.

## 2. MARCO REFERENCIAL DIDACTICO

Consideramos que el uso de métodos determinísticos en la resolución de problemas matemáticos opera como un obstáculo para el abordaje de problemas de naturaleza aleatoria. Por lo que la noción de obstáculo se constituye en la herramienta didáctica que nos permite el estudio de esta problemática. Por ello sintetizaremos la posición que Brosseau sostiene al respecto.

Entre las características que Brosseau atribuye a lo que el llama un obstáculo destacamos la siguiente: un obstáculo es un conocimiento no una falta de conocimiento. Dicho conocimiento es adecuado para producir respuestas correctas en un determinado contexto; pero resulta insuficiente para abordar situaciones de otro contexto.

En sus investigaciones identifica tres tipos de obstáculos según su origen:

1. De origen ontogenético
2. De origen epistemológico
3. De origen didáctico

En este trabajo, por las características del mismo, nos centraremos en los obstáculos de origen didáctico, los cuales proceden de "modos de enseñanza" y tienen un fuerte impacto en los aprendizajes de los alumnos. Son obstáculos que deben evitarse, lo que implica no solo identificarlos sino analizar las consecuencias que las distintas formas de enseñanza producen.

Desde este marco consideramos que la comprensión del **modelado estocástico** en la enseñanza media en particular el de la regresión lineal, debe ser abordado mostrando los límites de aplicación de la función lineal y de los procedimientos determinísticos para el abordaje de los problemas de naturaleza aleatoria. Por ello, sostenemos que el primer paso para intentar evitar el obstáculo didáctico planteado es ***clarificar el uso operacional que tiene la función lineal en el marco de la estadística y poner en evidencia el rol que cumplen conceptos esencialmente determinísticos cuando son usados para modelar fenómenos aleatorios.***

## 3. MARCO REFERENCIAL DISCIPLINAR

### El modelado estadístico y la función lineal

Los modelos son representaciones abstractas y simplificadas de la realidad y frecuentemente son usados tanto en ciencia como en tecnología. Si bien en rigor no se puede pensar que un modelo sea verdadero, en el sentido que sea una fiel representación de un fenómeno o sistema real, en estadística las inferencias están basadas en esta suposición.

Los modelos pueden ser determinísticos o probabilísticos. En los primeros los resultados están precisamente definidos, mientras que en los últimos los resultados muestran variabilidad debido a la presencia de factores desconocidos.

En el modelado estadístico se está interesado en descubrir patrones de comportamiento sistemático para datos obtenidos de experimentos u observaciones que contienen una componente aleatoria. Situaciones muy generales en las que se aplica el modelado estadístico están generadas a partir del interés que existe en explicar la variabilidad de una *característica* observada sobre unidades observacionales independientes y homogéneas para todos los aspectos de interés. Esta característica en el lenguaje de la estadística se llama *variable respuesta*. Se asume que existe algún mecanismo aleatorio desconocido que origina esa variabilidad y a partir de la información que entregan las observaciones realizadas, *los datos*, se construye un modelo simple que permita echar luz sobre esta variabilidad. En estas condiciones el modelo primario es una distribución de probabilidad. Pero en muchas situaciones, en las que no hay homogeneidad, se requiere algo más que una distribución de probabilidad para explicar la variación presente en los datos, la respuesta estaría cambiando además cuando cambia algunas de estas condiciones medibles, que operan sobre las unidades experimentales. Estas condiciones se llaman *variables explicativas* (explicativas de las respuestas). Así podría plantearse un modelo que tenga una parte sistemática además de una componente aleatoria. Uno de los modelos más simples es el llamado modelo de regresión lineal.

La idea central para la construcción del modelo de regresión lineal es que **la respuesta media de la variable Y** cambia con las condiciones de manera lineal, esto es

$$E(Y_i/x_i) = \alpha + \beta x_i \tag{1}$$

donde $x_i$ es el valor que asume la variable explicativa $x$ al medirla sobre el individuo *i.* Tanto $\alpha$ como $\beta$ son desconocidos y se llaman los *parámetros* del modelo, ellos se conocerán o en el lenguaje estadístico se *estimaran* a partir de los datos.

El modelo (1) expresa como se espera que sea el comportamiento promedio de la variable respuesta $Y$ al observarla repetidas veces bajo las mismas condiciones $x_i$. En rigor desde un punto de vista práctico, generalmente, se observa solo una vez y este caso se escribe

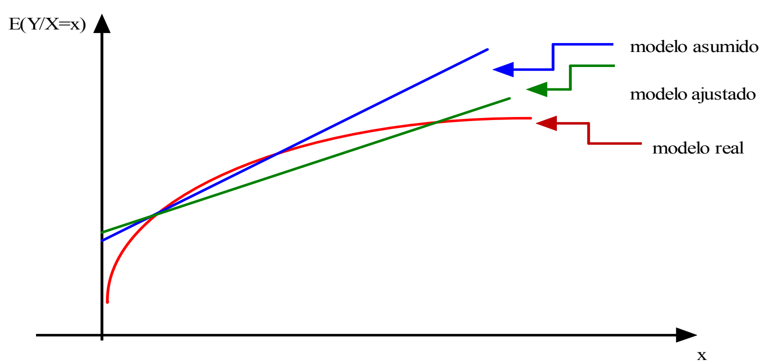
$$Y_i = \alpha + \beta x_i + \epsilon_i \tag{2}$$

Siendo $\varepsilon$ un error de naturaleza aleatoria que origina la variación, en consecuencia también aleatoria, de los $Y$ alrededor de su valor medio. Bajo estas condiciones los datos serán $n$ pares de puntos $(y_i\ ;\ x_i)$ a partir de los cuales se debe determinar la recta (1).
Al observar la estructura de la expresión (2) surge que puede ser escrita como

$$Y_i = f(x_i) + \varepsilon_i$$

Donde ahora aparece la forma funcional habitual de la llamada función lineal pero como la parte sistemática del modelo

Cuando se trabaja con el modelado estadístico y en particular con el modelo de regresión lineal están presentes tres modelos: el verdadero y desconocido, el supuesto y el ajustado, como se muestra en la figura, de allí que en la estadística los parámetros del modelo resulten *estimados* por los valores que se *calculan* a partir de los datos que proporciona una muestra.




Por otra parte, cabe mencionar que, cuando en el contexto de la estadística se habla de un modelo lineal se hace referencia a la linealidad en los parámetros. A manera de ejemplos

$$E(Y_i/x_i) = \alpha + \beta x_i^2 \ \text{ y } \ E(\log Y_i/x_i) = \alpha + \beta \frac{1}{x_i} \ \text{son lineales en } \alpha \text{ y } \beta$$
$$E(Y_i/x_i) = \alpha + \beta^2 x_i \ \text{, no es lineal en } \beta.$$

De allí que algunos modelos lineales especifiquen que la relación gráfica entre la variable respuesta y la variable explicativa sea una recta, y otros expresen vínculos gráficos diferentes.

Con lo expuesto podemos establecer las siguientes diferencias entre la función lineal y la regresión lineal.

| Función lineal | Regresión lineal |
|---|---|
| La variación de x e y no es aleatoria | La variación de una o ambas, x e y es aleatoria |
| Relación funcional | vínculo estadístico que extiende la noción de dependencia funcional. |
| El cálculo de los parámetros es independiente de los puntos seleccionados sobre la recta. | La estimación de los parámetros depende de los puntos seleccionados. |
| Dada una tabla de valores de *x* e *y* se obtiene una recta y dada esta es posible reproducir la tabla. | Dada una tabla de valores de *x* e *y* se obtiene una recta pero ella no permite reproducir la tabla. |
| El rol de variable independiente y dependiente es indistinto. $y = a_1x + b_1$ y $x = a_2y + b_2$ representan el mismo conjunto de puntos | Los roles de las variables explicativa y respuesta no son intercambiables. $y = \alpha_1x + \beta_1$ y $x = \alpha_2 y + \beta_2$ no representan el mismo conjunto de puntos |

4. ANALISIS DE UNA SITUACION

A los fines de poner de manifiesto el rol que juegan los procedimientos y conceptos determinísticos cuando son usados en la modelización de fenómenos aleatorios, presentaremos, a modo de ejemplo, una situación problemática extraída de un libro de texto de circulación actual en nuestro país. La que analizaremos en términos del uso que puede hacerse de la función lineal en un estudio de regresión lineal.

**Tarea**

La tabla recoge, de varias personas, el número de horas semanales que dedican a hacer deporte y el número de pulsaciones por minuto que tienen en reposo.

| Hs. deporte | 0 | 0 | 0 | 1 | 1 | 3 | 3 | 4 | 5 | 7 |
|---|---|---|---|---|---|---|---|---|---|---|
| pulsaciones | 66 | 62 | 73 | 72 | 65 | 60 | 66 | 58 | 57 | 54 |

**a)** Dibuja la nube de puntos
**b)** ¿Hay dependencia funcional, estadística o casual?
**c)** La ecuación de la recta de regresión es: $y = -1{,}5x + 68$

Estima el número de pulsaciones que tendrá una persona que dedica 2 horas semanales a hacer deporte.
Estima el número de pulsaciones de un ciclista profesional que entrena 4 horas diarias. ¿Te parece razonable esta estimación?

Observación: Cabe aclarar que la tarea precedente se presenta en el texto bajo el título Distribuciones Bidimensionales.

**Síntesis del análisis de la tarea**

El contexto planteado para la situación no especifica que rol cumplen cada una de las características que allí intervienen (horas semanales dedicadas al deporte y número de pulsaciones por minuto). Tampoco se explicita el método de muestreo.

En primer lugar se solicita dibujar la nube de puntos, para lo cual se provee una tabla del "tipo" de las que se presentan para graficar una función.

En el inciso b) donde se presenta el interrogante crucial en esta temática, los datos aportados por la tabla "evidencian" no sólo una respuesta parcial, sino claramente de un tipo: no funcional. No se aborda, por lo tanto, el hecho que la vinculación estadística entre dos características es de naturaleza aleatoria y no surge de la no existencia de relación funcional.

En el inciso c) se presenta "la forma típica" de una función lineal sin identificar cuál es la característica denotada con $x$ y cual la denotada con $y$, produciéndose así las siguientes contradicciones:

- La dependencia no es funcional y se presenta por medio de una función.
- No se puede reproducir la tabla a partir de la expresión analítica, procedimiento usual en la obtención de la gráfica de una función lineal.
- Bajo el supuesto que el alumno infiera que la variable independiente $x$ representa las horas deporte y la dependiente $y$ las pulsaciones, ¿qué interpretación le asignará a los valores de $y$ que obtiene para cada uno de los valores de $x$ de la tabla?

Por estas cuestiones es que la respuesta al pedido de "estimar" el número de pulsaciones será claramente de naturaleza determinística. Produciéndose en esta etapa un **afianzamiento** de la superposición conceptual función lineal-regresión lineal, calcular-estimar. Constituyéndose esto en una evidencia del tipo de obstáculo didáctico que caracterizamos.

## 5. CONCLUSIONES

En los lineamientos curriculares de la escuela media, en distintos países, se incluyen contenidos de estadística desde edades tempranas, sin embargo consideramos que si no se toman recaudos, se corre el riesgo de subordinar el abordaje de problemas referidos a fenómenos aleatorios al trabajo con los conceptos matemáticos utilizados para el modelado estadístico. De esta manera la metodología esencialmente determinística de la matemática podría constituirse en un verdadero obstáculo para la comprensión de la naturaleza aleatoria de las situaciones que la estadística permite resolver.

## BIBLIOGRAFIA


Batanero, Godino y otros, *Errores y dificultades en la comprensión de conceptos estadísticos elementales.* Recuperable en: [http://www.ugr.es/local/jgodino]

Brosseau (1983), *Les obstacles epistemologiques et les problemes en mathematiques*. Rechearches en Didactique des mathematiques.

Godino, J. (2003)*Teoría de las funciones semióticas. Un enfoque ontológico – semiótico de la cognición e instrucción matemática.* Recuperable en: [http://www.ugr.es/local/jgodino]

Rawling, J. (1988). *Applied Regression Analysis.* Wadsworth & Brooks

Hamilton, L. (1992). *Regression with Graphics.* Duxbury Press