



Aprendiendo idiomas IX

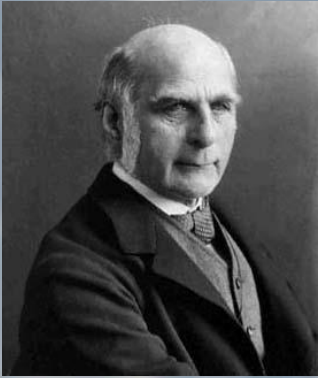
María Jesús Vázquez Gallo

Jean François Champollion, estudió la **pedra Rosetta** en la que hay una inscripción en tres sistemas de escritura. Uno es el griego, los otros dos son el jeroglífico y el demótico.

Durante 23 años, Champollion trabajó para descifrar el contenido de aquella piedra; finalmente, en 1.822 publicó la traducción completa de la inscripción trilingüe.



1. **Introducción.**
2. **Regresión.**
3. **Regresión simple.**
4. **Continuará...**



En *Aprendiendo Idiomas I*, mencionamos a Francis Galton, primo de Darwin, observador del fenómeno llamado **regresión** a la media, a finales del S. XIX.

¿Recuerdas?

En 1869, Galton publicó el libro "Hereditary Genius", y a través del estudio de problemas de la herencia, llegó al concepto de **correlación**, asignando a un conjunto de variables un número que daba una medida del grado de relación existente entre ellas.

Galton comparó la estatura de los padres con la de sus hijos y observó que personas excepcionalmente altas solían tener hijos de estatura menor, mientras que personas muy bajas solían tener hijos más altos.



1. Introducción

Esta observación de retorno a la media, le llevó a enunciar el llamado "*principio de la mediocridad*", según el cual los hijos tendían a tener estaturas más cercanas a la media que sus progenitores.

Aunque las observaciones de Galton están en el origen del actual análisis de regresión y sus observaciones fueron reales, la teoría de regresión a la mediocridad se considera actualmente como una **falacia**.

¿por qué?

Galton extendió la observación realizada sobre un conjunto de padres con talla excepcional a *cualquier tipo de padres*, excepcionalmente altos o no, excepcionalmente bajos o no.

Los valores extremos se deben en gran parte al azar, de ahí que los factores genéticos que producen una talla excepcional por exceso o por defecto no pasan necesariamente a los hijos.

1. Introducción

Actualmente, algunos economistas son partidarios del “*modelo de convergencia*”, según el cual los países pobres están alcanzando a los países más ricos.

Es cierto que las economías de países como Brasil, China, India y Rusia, han crecido significativamente en los últimos 15 años y que la economía de Estados Unidos ha entrado en crisis.

¿esto significa que el crecimiento económico tiende a la mediocridad?

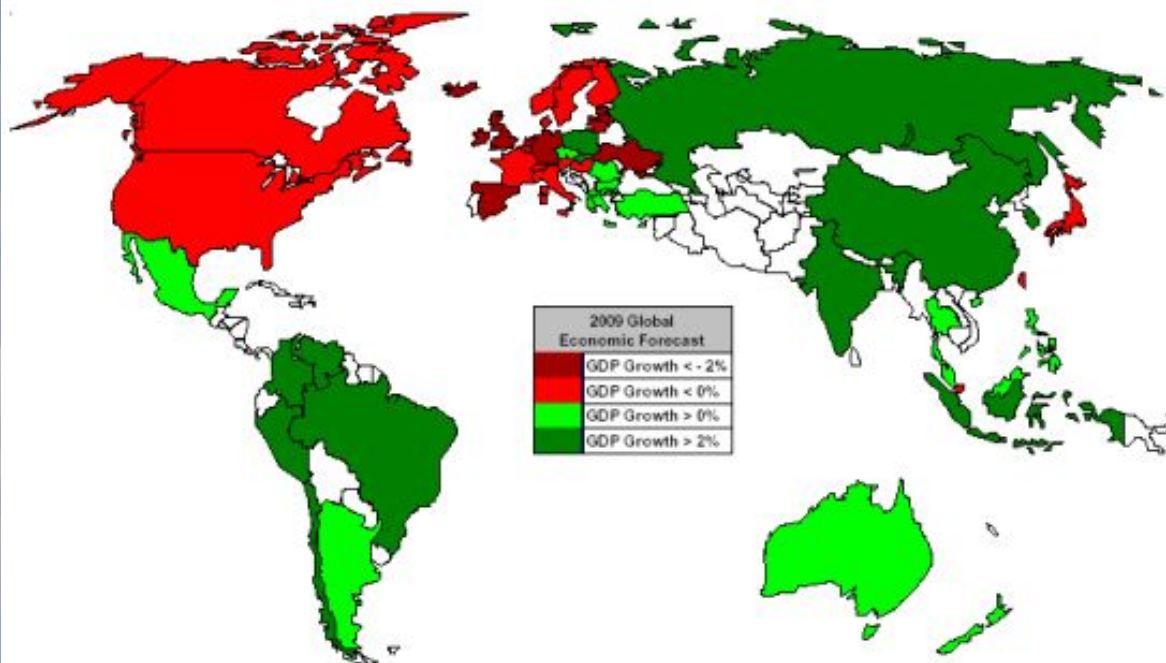
Estamos comprobando que la crisis originada en Estados Unidos está afectando al resto de economías en el mundo, en contra del modelo de convergencia.

Quizá los que postulan el modelo de convergencia hayan caído en la **falacia de Galton**...

¿qué opinas tú?



2009 World GDP Forecast Map



2009 Global Economic Forecast	
■	GDP Growth < -2%
■	GDP Growth < 0%
■	GDP Growth > 0%
■	GDP Growth > 2%

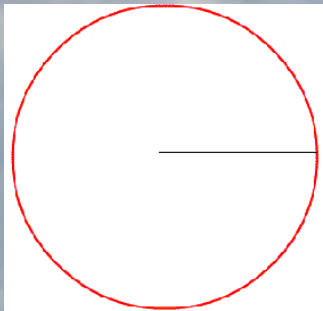
Estas son las previsiones de crecimiento que se hicieron para el 2009...

GDP Growth > 2%		GDP Growth > 0%		GDP Growth < 0%		GDP Growth < -2%	
China	7.70%	Thailand	1.90%	Belgium	-1.90%	Iceland	-9.30%
Indonesia	5.30	Bulgaria	1.80	Taiwan	-1.90	Latvia	-6.90
Peru	5.30	Romania	1.80	France	-1.80	Ireland	-5.00
India	5.20	Australia	1.70	Hungary	-1.60	Estonia	-4.70
Brazil	2.90	Czech. Rep.	1.70	USA	-1.45	Lithuania	-4.00
Russia	2.90	Malaysia	1.50	Japan	-1.40	Ukraine	-2.90
Chile	2.85	Cyprus	1.10	Sweden	-1.40	U.K.	-2.80
Slovakia	2.70	Hong Kong	1.00	Austria	-1.20	Germany	-2.30
Colombia	2.50	New Zealand	1.00	Finland	-1.20	Italy	-2.00
South Korea	2.45	Mexico	0.80	Denmark	-1.00	Netherlands	-2.00
Poland	2.00	Turkey	0.80	Singapore	-0.70	Spain	-2.00
Venezuela	2.00	Argentina	0.50	Switzerland	-0.70		
		Philippines	0.50	Canada	-0.50		
		Greece	0.20	Norway	-0.50		

Parece interesante saber algo más sobre todo esto:

Tras los estudios de Galton, los modelos estadísticos que explican la dependencia de una variable respecto de otra u otras variables cuantitativas, se llaman **modelos de regresión**:

- La primera se llama **variable respuesta**, digamos y .
- Las demás se llaman **variables explicativas**, digamos x .



Ejemplo 1: Si queremos estudiar la longitud de una circunferencia en términos de su radio, entonces la variable respuesta es y : longitud de la circunferencia; y la variable explicativa es x : su radio.

Ejemplo 2: Si queremos estudiar la cantidad de leche consumida al año en las mujeres con respecto a su edad, entonces la variable respuesta es y : nº de litros consumidos por las mujeres al año; y la variable explicativa es x : la edad de las mujeres.

2. Regresión

Ejemplo 3: Si queremos estudiar el sueldo de un directivo en función del número de lunares que posee en su cuerpo, entonces la variable respuesta es y : sueldo anual en euros del directivo; y la variable explicativa es x : n° de lunares que posee.

- a) Cuando el conocimiento de x determina completamente el valor de y , se dice que existe una **relación funcional** entre las variables.
- b) Cuando el conocimiento de x no da información sobre el valor de y , se dice que las **variables** son **independientes**.
- c) Cuando el conocimiento de x influye en mayor o menor grado en el valor de y , se dice que existe una **relación estadística o estocástica** entre las variables.

¿Te atreves a emparejar cada uno de los ejemplos con una de estas tres situaciones?

2. Regresión

Los **modelos de regresión** intentan explicar este tipo de relaciones. Es decir, tratan de construir modelos para **relaciones** entre la variable respuesta y las variables explicativas.

a) Cuando se considera una única variable explicativa se tiene un modelo de **regresión simple**.

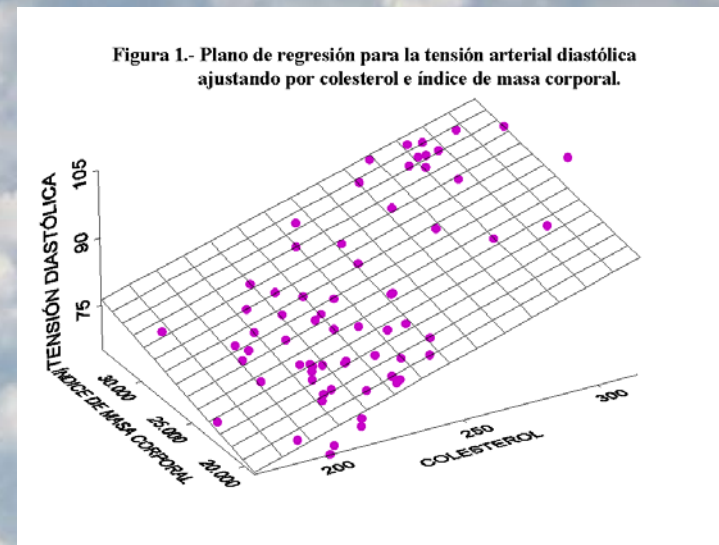
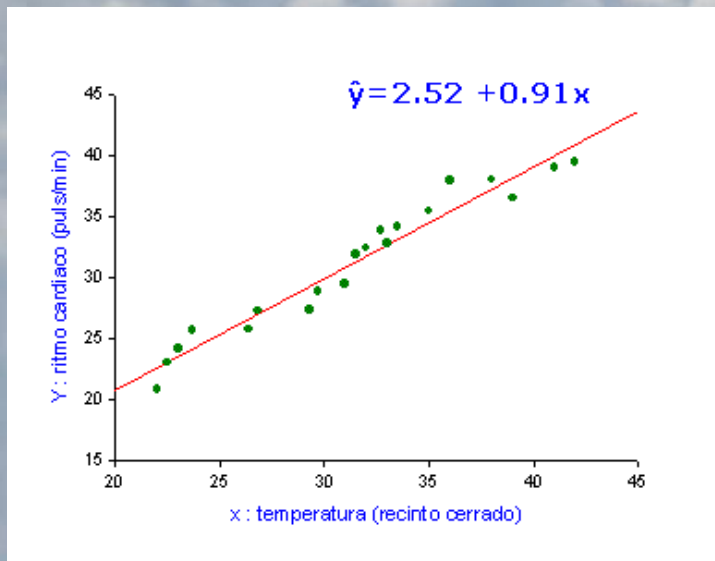
b) Cuando se consideran varias variables explicativas se tiene un modelo de **regresión múltiple o general**.



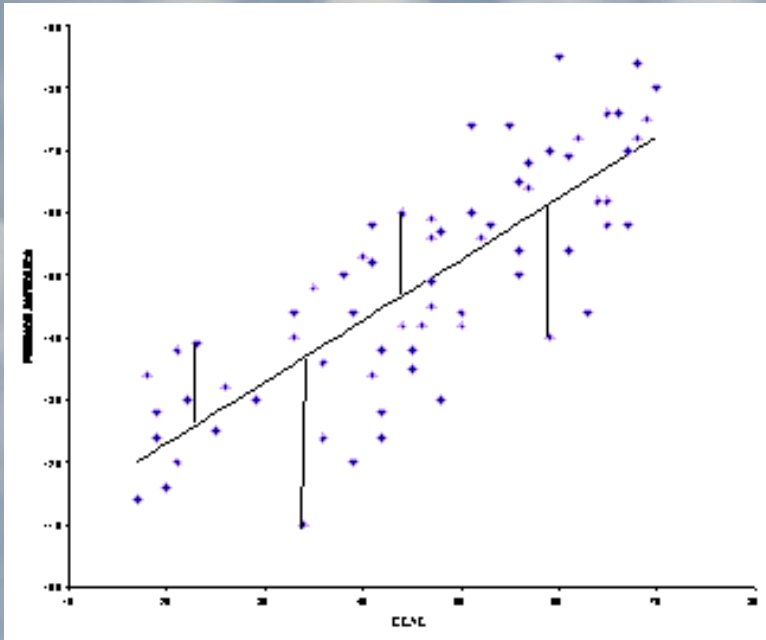
Ejemplos: una regresión simple podría explicar el consumo de combustible en un vehículo en términos de su velocidad media y una regresión múltiple podría utilizarse para explicar el precio de una vivienda en función de su superficie y de su antigüedad.

Las relaciones entre las variables que suponen los modelos de regresión son lineales. Construida una nube de puntos a partir de los valores de las variables, se busca un objeto lineal que se ajuste “lo mejor posible” a dicha nube:

- a) En la regresión simple, se busca una recta, la recta de regresión.
- b) En la regresión múltiple o general, se busca un plano, el plano de regresión.



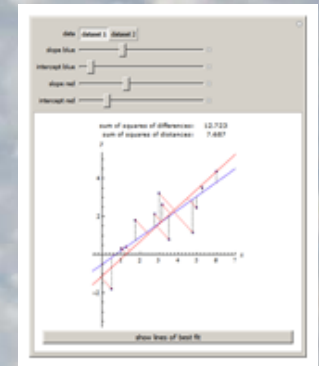
¿Cómo decidimos que el objeto lineal se ajusta lo mejor posible a la nube de puntos?



En la **regresión simple**, eligiendo como recta de regresión la que minimiza las distancias verticales de los puntos a la recta.

Esto se conoce como **estimación por el método de mínimos cuadrados**.

Otra opción, menos frecuente en la práctica, es minimizar las distancias de los puntos a la recta en dirección ortogonal a ella.



3. Regresión simple

Las hipótesis básicas en un modelo de **regresión simple** son las siguientes:

- 1) Los factores que influyen en la variable respuesta **y** son de dos tipos:
 - Una **variable explicativa x** conocida al observar **y**; de la que **y** depende **linealmente**.
 - Un serie de factores que, individualmente, influyen poco en la variable respuesta y que, conjuntamente constituyen una **perturbación aleatoria**.

En una fórmula:

$$y = \beta_0 + \beta_1 x + u$$

con **u** la perturbación aleatoria y β_0 , β_1 los parámetros que queremos estimar.

3. Regresión simple

Ejemplos: Suponemos que:

El consumo de combustible de un vehículo, y , depende linealmente de su velocidad media, x , y en la perturbación u englobamos a: las condiciones climáticas, el tipo de conducción realizada, el tipo de carretera, etc.

El ritmo cardíaco y depende linealmente de la temperatura x en el recinto (cerrado) en el que nos encontremos, pero se perturba por un conjunto u de factores como las emociones experimentadas, la cantidad de luz, el tamaño del recinto, etc.

El presupuesto de un hospital, y , depende linealmente de su nº de pacientes, x , y en la perturbación u podemos englobar a: la superficie de las instalaciones, el nº de médicos y enfermeros, el nº de personal administrativo, etc.

3. Regresión simple

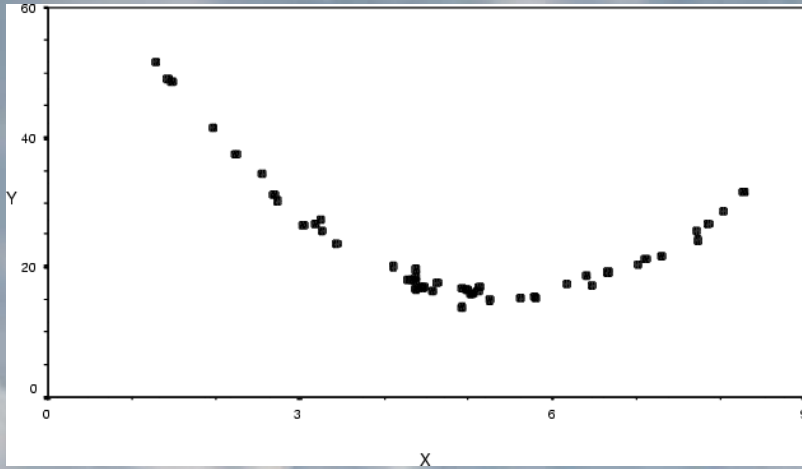
Otras hipótesis del modelo:

2) La esperanza de la perturbación u es nula. Esto viene a decir que el parámetro β_0 representa el valor medio de y cuando x es cero y el parámetro β_1 representa el incremento que experimenta la media de y cuando x aumenta en 1.

¿Recuerdas? En Aprendiendo Idiomas II vimos que el concepto de esperanza lo introdujo Huygens en el S. XVII en relación con el valor esperado de un juego.

Las hipótesis 1) y 2) dicen conjuntamente que la media de la distribución de la variable y , para x fija, crece linealmente con x . De esta suposición depende todo el modelo y conviene comprobarla.

¿Y qué ocurre si no se tiene esa relación lineal?



Si nuestra nube de puntos fuera ésta, claramente no parece que la relación entre las variables sea lineal...

Pero, afortunadamente:

Algunas relaciones no lineales pueden convertirse en relaciones lineales, transformando las variables apropiadamente, como veremos.

¿Se acabaron las hipótesis? Pues no, también se supone que las perturbaciones son independientes entre sí y alguna otra cosa.

Y con todo esto, en la siguiente entrega, seguiremos aprendiendo sobre regresión.

